

SCIENTIFIC MODELS OF HUMAN HEALTH RISK ANALYSIS IN LEGAL AND POLICY DECISIONS

DOUGLAS CRAWFORD-BROWN*

I

INTRODUCTION

Scientists working in the fields of law or policy faced with risk-based decisions find themselves in very different positions epistemologically than those in the routine practice of science. First, risk assessment is an interdisciplinary activity, requiring the piecing together of expertise from many fields. Only a community of scientists could possess so much expertise. It is problematic when a single scientist, or a small group of them, is called upon to speak on behalf of the entire community. Second, science is by nature forward-looking. It is a process by which truth is progressively refined and approached, rather than an end point. In court, or in the policy arena, however, scientists are called on to “freeze” this process in a snapshot that will be regarded as final. Scientists may have little or no experience in this activity. The quality of scientific predictions of risk in the courtroom and policy arena rests in large measure on how these two differences between the normal practice and the legal/policy practice of science are reconciled. This article considers a variety of issues that arise in reconciling these differences, and the problems that remain with scientific estimates of risk when these are used in decisions.

II

DEFINITION OF RISK

Before turning to the issues arising from risk predictions, a definition of risk is needed. Three schools of thought are prevalent in the field of risk assessment.¹

(1) Objective risk. This school treats risk as an objective property of the world, much like any other scientific concept. Risk is identified with a combination of the probability and the severity of an outcome.² Often, the combination

Copyright © 2001 by Douglas Crawford-Brown

This article is also available at <http://www.law.duke.edu/journals/64LCPCrawford-Brown>.

* Professor, Departments of Environmental Sciences and Engineering, and Public Policy Analysis; Chairman, Environmental Science and Studies, University of North Carolina, Chapel Hill.

1. See DOUGLAS CRAWFORD-BROWN, *RISK-BASED ENVIRONMENTAL DECISIONS: METHODS AND CULTURE* 5-12 (2000).

2. See *id.* at 5-9.

is the mathematical product of the probability and the severity, but this operation is not scientifically or logically necessary, as is demonstrated by the fact that there are competing operations in the scientific community such as Quality Adjusted Life Years ("QALY").³

(2) Subjective risk. This school treats risk as a subjective property of individuals in a society. Risk is identified as the degree of concern, or dread, expressed by an individual over some situation. It is measured using polls based on expressed or revealed opinions.⁴

(3) Psychologicistic risk. This school treats risk as a property of epistemological reflection, or analysis of existing evidence. Risk is identified as the degree of evidentiary support for the belief that an adverse effect will occur under certain circumstances. This school shares features with the objective school, such as the recognition that risks are related to probabilities. It is closer, however, to the Bayesian conception of probability because the objective school employs a long-term frequency concept of probability.⁵

While some social scientists adhere to the subjective concept of risk, it generally is not the concept employed in the practice of natural science or in legal or policy decisionmaking. The difference between the objective and psychologicistic conceptions of risk, however, is not so apparent in practical terms as might otherwise appear. For example, adherents of both would agree with the following statement by the National Academy of Science's National Research Council:

Human-health risk assessment entails the evaluation of scientific information on the hazardous properties of environmental agents and on the extent of human exposure to those agents. The product of the evaluation is a statement regarding the probability that populations so exposed will be harmed, and to what degree. The probability may be expressed quantitatively or in relatively qualitative ways.

Furthermore, both agree that risk involves probability and severity. Both agree that probability and severity are objective properties of the world. The objective school, however, equates risk solely with probability and severity, while the psychologicistic school adds a third component to risk: rational confidence in the estimates of probability and severity. The key difference between the two schools is that the objective school believes risk arises from this probability, while the psychologicistic school believes risk arises from a state of evidence that rationally supports this probability. To see this distinction, imagine a case in which a scientist from the objective school and a scientist from the psychologicistic school each agree that the probability of cancer following exposure to an environmental pollutant is either 0.1 or 0.2. They also agree that they are sixty

3. *See id.* at 29-30. QALY equals the effective number of years of life after subtracting years of life lost through fatal disease, and multiplying remaining years of life by a quality factor to account for the presence of non-fatal diseases that make life of lower quality.

4. *See id.* at 9-10.

5. *See id.* at 10-12. The Bayesian conception of probability is that all probabilities refer to the degree of subjective confidence in a belief, not to objective properties of phenomena.

6. COMMITTEE ON RISK ASSESSMENT OF HAZARDOUS AIR POLLUTANTS, NATIONAL RESEARCH COUNCIL, SCIENCE AND JUDGMENT IN RISK ASSESSMENT 25-26 (1994).

percent confident that the probability is 0.1 and forty percent confident that the probability is 0.2. They further agree that a full scientific picture of the risk should include both estimates of probability and their associated degrees of confidence.

Where do they disagree? Imagine that they both appear in court and are asked for their best estimate of the risk. The adherent of the objective school believes there is a *single* risk in the world, that it is *either* 0.1 or 0.2, and that she is *uncertain* which is the true risk. Since she is more confident that the probability of effect is 0.1 than that it is 0.2, her best estimate of the risk is 0.1. This is her testimony when faced with the challenge from the court.

The adherent of the psychologistic school, however, does not believe there is a single risk in the world. He agrees that there is a single probability of effect, but this probability is not conflated with risk. Each probability, 0.1 and 0.2, has some rational support in the evidence, which is why each has a degree of confidence above zero. As a result, one risks each state of affairs being true—that is, one risks living in a world in which each probability exists. When asked for the *best* estimate of the risk, this individual will reply that he is sixty percent confident that the probability is 0.1 and forty percent confident that it is 0.2. According to the psychologistic scientist, it is not correct to select a single probability and equate this with the risk. Although there may be a single best estimate of the probability of effect, risk is a more complex concept involving the ensemble of estimates of probability and their associated degrees of confidence.

In many ways, then, there are no practical implications when a court or policy-maker receives advice on risk but is unsure whether a scientist is of the objective or psychologistic school. Both scientists will provide estimates of the probability of effect, the severity of effect, and the confidence with which specific estimates of the probability may be asserted rationally. The sole difference lies at the moment a scientist is asked to reflect on this information, digest it, abstract it, and provide a summary statement representing the best estimate of the risk. At that moment, the scientist in the objective school will say she is uncertain about the risk, but is most confident it is 0.1. The scientist in the psychologistic school will say this uncertainty is part of the ontology of risk and will not equate any single probability of effect with the risk. Both scientists would agree that the best estimate of the probability of the effect is 0.1. Because they work from different definitions of risk, however, they disagree as to whether this estimate constitutes a statement of *risk*.

III

RISK ASSESSMENT AS A COLLECTIVE ACTIVITY

Human health risk arises from a confluence of three aspects of the world: (1) a risk *agent*—for example, the radiation released from Three Mile Island; (2) a potentially affected *individual* or group—for example, the people irradiated by the radiation from Three Mile Island; and (3) a *condition of exposure* in

which the risk agent and the potentially affected individuals are brought into contact—for example, the inhalation of radionuclides in the air after the Three Mile Island releases. All three of these aspects of the world must be in place for the risk to exist, and the risk is a property of all three of these. It is not proper to speak scientifically of *the* risk from a risk agent—for example, the risk from radiation—because the answer also depends on the potentially affected individual and the condition of exposure.

As a result, risk assessment includes a series of steps that both produce and integrate information on all three of these aspects of the world. As defined by the National Academy of Science's National Research Council, the components of a scientific risk assessment are Hazard Identification, Exposure-Response Assessment, and Risk Characterization.⁷ In Hazard Identification, evidence is assembled and assessed to determine whether the risk agent, under any conditions of exposure, produces adverse health effects in humans. If it does, the severity of those effects are assessed as well as whether they are restricted to any particular sub-group, for example the elderly.⁸ Exposure Assessment involves determining to what degree, and by which routes—for example, inhalation or ingestion—potentially affected individuals are exposed to this risk agent under the precise conditions of exposure of interest.⁹ Exposure-Response Assessment considers evidence to assess the relationship between the degree of exposure and the probability/severity of the health effects shown in Hazard Identification.¹⁰ Risk Characterization integrates the results of the first three components of risk assessment. This final step takes into account the actual exposures found in Exposure Assessment and the relationship between exposure and response found in the Exposure-Response Assessment to calculate the probability and severity of effects under the conditions of exposure and in the affected population of interest in this decision. Risk Characterization also determines the variability of this risk between different individuals and the uncertainty in these estimates of risk.

Hazard Identification involves assessment of scientific evidence from fields such as medicine, biology, biostatistics, epidemiology, and toxicology. Exposure Assessment analyzes scientific evidence from environmental science, atmospheric science, hydrology, demographics, and other similar areas. Exposure-Response Assessment involves assessment of scientific evidence from fields like toxicology, biology, biomathematics, and epidemiology. Finally, risk characterization integrates evidence using uncertainty analysis, logic tree analysis, and fault tree analysis. These four components of a risk assessment require methods and judgments from widely disparate fields of science.

Judgments of risk, therefore, result from a community of scientists whose expertise spans the range noted above. The scientific quality of the final as-

7. *See id.* at 26-27.

8. *See id.* at 57-60.

9. *See id.* at 43-45.

10. *See id.* at 60-66.

assessment results not merely from the formal methodologies employed, but from the scientific judgments made in selecting the most defensible and reliable bodies of data and models to use in those methodologies. While an individual might be able to master the formal methodologies, that individual will likely not have the expertise required to speak on behalf of the scientific community in all areas in which data and models must be selected for use in those methodologies. As an assessment is constructed from the base upwards—from Hazard Identification to Risk Characterization—the reasoning developed by an individual occupying one level of judgments rests on judgments made by those who came before him. The result is a line of reasoning in which the judgments of most of the experts are dependent upon the prior reasoning of experts in other fields. No one expert can warrant the judgments required in all the necessary steps of reasoning. A guarantee that the final judgment of risk represents “good” or “reliable” science results from the cumulative effects of individual judgments by an organized community, and not the judgment of any one individual, not even the individual integrating the scientific knowledge in Risk Characterization.

IV

SCIENTIFIC RATIONALITY AS A COLLECTIVE ACTIVITY

Part III introduced the idea that the complexity of a risk assessment prevents any one assessor from warranting the results of the entire assessment.¹¹ Science-based risk assessment is also a collective activity in a deeper sense. Some claim that scientific judgments themselves, even in a single area of expertise, are the result of a community effort.¹² This claim means that each individual judgment integrated into a risk assessment is a collective activity of a community of scientists. Therefore, the rationality of scientific judgments is the rationality of a community rather than an individual.

People who participate in risk assessment consider their own judgments rational, and may at times consider those of their competitors irrational, without being able to define the term “rational.” There are hundreds of definitions of rationality in the philosophical literature, but they share characteristics captured in the following definition: Rationality is the selection of apt means for the highest ends.¹³ Rationality focuses on systematically comparing alternative means to reach a set of ends and selecting the most appropriate means. “Apt” in this sense does not require complete certainty, but the given means must reasonably be expected to reach the given ends. If risk estimates justify the means, rationality requires an examination of the quality of the evidence underlying

11. See discussion *supra* Part III.

12. See BARRY BARNES ET AL., *SCIENTIFIC KNOWLEDGE: A SOCIOLOGICAL ANALYSIS* 16 (1996).

13. See Douglas Crawford-Brown & Jeffrey Arnold, *The Role of Evidential Reason and Epistemic Discourse in Establishing the Risk of Environmental Carcinogens*, in *COMPARATIVE ENVIRONMENTAL RISK ASSESSMENT* 261, 264 (C. Richard Cothorn ed., 1992).

those estimates to demonstrate that the evidence is sufficient to warrant adoption of the risk estimates.

What are the criteria by which a scientific expert may determine whether the evidentiary basis for a judgment is sufficient to allow a claim to rationality? Mario Bunge has identified seven desiderata that should be considered before a judgment is deemed rational: (1) conceptual clarity; (2) logical consistency; (3) ontological realism; (4) epistemological reflection; (5) methodological rigor; (6) practicality; and (7) valuational selection.¹⁴

Conceptual clarity requires that, to the extent possible, the argument leading to judgment should employ well-defined terms. In risk assessment, this desideratum would apply to the key terms of the field such as probability, severity, and exposure.

Logical consistency requires that supporting arguments be logically consistent. Risk estimates should follow deductively from the assumptions and the data, to the extent possible. Where default or science policy assumptions are used to fill gaps in scientific understanding, their status should be noted. The line of argument used in obtaining risk estimates should be clear.

Ontological realism requires that the analysis make scientifically supportable ontological commitments. All processes and quantities in the analysis should conform to the best scientific understanding of the phenomenon in question. For example, in cancer risk estimates, the models employed should reflect, to the degree possible, the best scientific understanding of the process of carcinogenesis. The models should not contain incorrect scientific assumptions and should employ the full suite of scientifically valid assumptions, to the extent feasible.

Epistemological reflection requires that all assumptions be scrutinized to determine the overall quality of the risk estimate.¹⁵ This quality is a function of the judgment's *epistemic status*, or degree of evidentiary support. Epistemological reflection is an uncertainty analysis in which competing assumptions are assessed for their relative degree of evidentiary support. Assumptions are introduced into an assessment only if they meet the *minimal epistemic status*, meaning the quality of the evidence must be above the level required for their rational adoption as a scientific belief.

Where appropriate, clearly defined methods proven reliable in past applications should be employed in selecting from alternative models, developing parameter values, and specifying confidence intervals on risk estimates.

The process of risk estimation should be selected for its practicality. It should use methods that can be completed in a reasonable length of time and that require only a reasonable amount of resources. A more scientifically valid

14. See Mario Bunge, *Seven Desiderata for Rationality*, in RATIONALITY: THE CRITICAL VIEW 5-16 (Joseph Agassi & Ian Charles Jarvie eds., 1987). Bunge's discussion does not consider the science of risk assessment, so the examples employed are those of the present author.

15. See DOUGLAS CRAWFORD-BROWN, RISK-BASED ENVIRONMENTAL DECISIONS: METHODS AND CULTURE 14-15 (2000).

method may, under this desideratum, prove irrational if unreasonable resources must be expended, or if the method will affect the ability to reach decisions in a timely manner.¹⁶

Finally, the assessment employs valuational selection. This factor ensures that the decisionmaker's highest values are being protected. For example, if the decisionmaker values environmental justice, the assessment might be required to produce an estimate of risk *homogeneity* in a population, not simply a point estimate of mean risk.¹⁷

Bunge recognizes that a single judgment often cannot satisfy all seven desiderata simultaneously in a single belief.¹⁸ Some judgments may compare favorably to other judgments under one desideratum but not another. Therefore, the risk assessor must appropriately balance these seven desiderata.

While all rational individuals seek this balance, logically, two rational scientific experts do not have to reach the same balance. One expert, for example, might prefer a model based on strong statistical methodology with little biological meaning, which satisfies the methodological desideratum but violates the desideratum of ontological realism. Another expert, however, might prefer a more biologically based model to which statistical methodologies for parameter estimation cannot be applied. Rationality requires only that the seven desiderata be considered in selecting from alternative risk estimates, but there is room for human values to guide the weighing of these desiderata because the science practice does not provide a "standard" weight. Deborah Mayo argues that meta-science may guide this choice, and it need not be simply a sociological one.¹⁹ Still, meta-science provides only rough guidelines for weighing and does not fully resolve the issue. Disagreement will remain among scientists as to how the seven desiderata should be weighed in analyzing competing scientific beliefs. These disagreements are likely traced in part to differences, such as social affiliation or personal styles of weighing evidence, that are not dealt with directly in metascientific discussions. This does not mean that such considerations cannot be accommodated in metascientific principles, only that I have found no evidence of their having been accommodated to date.

Scientists from the Classical School and the Dialogical School disagree as to whether there are potentially clear rules for determining the rationality of a particular scientific risk assessment under these seven desiderata.²⁰ To the Clas-

16. See CARL K. CRANOR, *REGULATING TOXIC SUBSTANCES: A PHILOSOPHY OF SCIENCE AND THE LAW* 151 (1993).

17. Homogeneity of risk is a measure of the degree to which risk is the same across all individuals in a population. A point estimate of risk is the average risk, ignoring this variability. Environmental justice might call for consideration of this homogeneity, seeking a policy that results in greater homogeneity. A point estimate of risk does not provide the information needed to judge homogeneity.

18. See Bunge, *supra* note 14, at 5-16.

19. See Deborah Mayo, *Sociological Versus Metascientific Views of Risk Assessment*, in *ACCEPTABLE EVIDENCE: SCIENCE AND VALUES IN RISK MANAGEMENT* 249-50 (Deborah Mayo & Rachele Hollander eds., 1991).

20. See HAROLD BROWN, *RATIONALITY* 5-14 (1988).

sical School, rationality applies formal rules universally.²¹ These rules of reasoning are clearly defined and agreed upon by all members of a rational community. Consistent application of the rules ensures that individuals do not select rules strategically by using a rule only when it leads to a desired conclusion. A belief must be logically necessary, following deductively from the evidence, with no room for ambiguity. Finally, a formal algorithm—for example, a mathematical or statistical procedure—must be applied to combine evidence in reaching a belief. In the example of judging whether a substance is a carcinogen, the process of assigning weight to the evidence must be formal. Therefore, under the classical theory of rationality, all properly trained rational scientists who are presented with the same evidence will reach the same conclusions.

The Dialogical School agrees that the rules must be formal rules, but disagrees that the scientific community must uniformly adopt these rules.²² Scientists will rationally disagree over the appropriate rules and when they should be applied. Two rational individuals may reach differing conclusions. The Dialogical School believes that rationality lies in the discussion between the individuals within the scientific community. Although disagreement may exist within this community, the individuals agree to compare their methods of argument, consider the possibility that they are incorrect, and continuously adjust their beliefs in light of different reasoning offered by others. Rationality, then, is established not only by the quality of individual arguments, but also by the character of the dialogue between individuals.

An individual scientist's claim that his judgments are scientific depends on whether he formed his judgments after reflection on the dialogue in the community. That dialogue, in turn, refers to at least the following principles culled from the literature on the philosophy of science for the scientific assessment of beliefs: empirical validity, conceptual success, axiomatic realism, and falsifiability.²³ Under the empirical validity principle, beliefs should be tested against data, and those beliefs that conform best to the data are considered strongest. For example, exposure-response models with a better fit to exposure-response data would receive greater weight in developing risk estimates. It is not enough, however, that the models simply fit the data; the models must also make accurate predictions. The conceptual success principle holds that beliefs should be tested by considering their ability to explain satisfactorily key phenomena identified by the scientific community. For example, models capable of explaining phenomena such as synergism, antagonism, or adaptive response would be assigned greater weight in forming judgments of risk than would models incapable of providing such explanations. Axiomatic realism requires that beliefs be tested by the degree to which they follow from models employing

21. *See id.*

22. *See generally* RICHARD BERNSTEIN, *BEYOND OBJECTIVISM AND RELATIVISM* (1983).

23. *See* DOUGLAS CRAWFORD-BROWN & KENNETH G. BROWN, *HAZARD IDENTIFICATION IN CARCINOGEN RISK ANALYSIS: AN INTEGRATIVE APPROACH*, REPORT TO THE UNITED STATES ENVIRONMENTAL PROTECTION AGENCY (1992).

terms with biological, chemical, or physical meaning. For example, physiologically and biologically based models would receive greater weight than models that are simple curve fits to data. Under the falsifiability principle, beliefs are rational only if they follow from data analysis that might potentially prove the beliefs wrong. For example, a complex exposure-response model with many unconstrained parameters could be fit to essentially any set of data. In that case no set of data would be construed as proving the model wrong. Therefore, if the model is too flexible, there is no possibility of showing it to be incorrect, and the model should not be assigned strong weight.²⁴

These four principles of epistemic reflection are applied with the seven desiderata to form specific scientific judgments. This process occurs in a community guided by two further criteria: (1) neutrality and objectivity, and (2) organized skepticism. Neutrality and objectivity suggest that beliefs should arise from a scientific community in which members apply specific principles, theories, models, and data based on their internal merits, and not whether they would lead to a predetermined policy position. Kristin Schrader-Frechette refers to this view as “scientific proceduralism.”²⁵ No individual can be completely neutral and objective, but rationality still exists because the community carries on a debate in which it is agreed that objectivity and neutrality are ideals towards which they should strive, however difficult their attainment might be. Under organized skepticism, all beliefs should be subject to continuous scrutiny and open to rejection on adoption. Scientists should understand that any belief might be false, and a wide range of scientific beliefs should always remain candidates for the truth.

In summary, scientific rationality in risk assessment is first a process by which the scientific community systematically applies the various previously discussed desiderata to assess alternative risk predictions, assigns rational measures of confidence to each, and ranks those beliefs by their reliability. Second, it is a process by which the scientific community discusses the criteria for judging the epistemic status of beliefs about risks’ magnitude and severity, debates the relative merits of these desiderata as guides to assigning rational confidence, and establishes processes to ensure the desiderata are applied reasonably. This discussion occurs in a well-defined community of qualified scientists and is guided by a principle of organized skepticism. Third, scientific rationality is the selection and transmission of well-founded beliefs generated by the above processes. If a belief is representative of the judgment of the scientific community and satisfies standards of minimal epistemic status, it not only is scientifically rational, but also represents “good science.” Beliefs that do not meet minimal epistemic status may still be scientifically rational and may serve

24. See generally KARL POPPER, *THE LOGIC OF SCIENTIFIC DISCOVERY* (1959). Popper goes so far as to say that the model is non-empirical. Since he considers science to be rooted in empiricism, this leads to the suggestion that such models also are non-scientific.

25. See Kristin Schrader-Frechette, *Reductionist Approaches to Risk*, in *ACCEPTABLE EVIDENCE*, *supra* note 19, at 218, 239.

as an adequate basis for scientific research, but will not represent “good science” in the sense intended here. “Good science” in this regard means science that is recognized not only as representing the current state of science, but as having satisfied standards of epistemic status established by the scientific community. Good science, therefore, is roughly equivalent to “scientific rationality,” with the former being used more frequently in the policy arena and the latter in the scientific literature.²⁶

This conception of scientific rationality also involves the systematic comparison of alternative risk estimates. This comparison distinguishes between *rationalizing* and *rationality*. Rationalization rhetorically presents evidence to show that a particular risk estimate is reasonable; such a presentation is likely to contain only supporting evidence and is unlikely to contain counterevidence. Rationality recognizes this by requiring that competing risk estimates be rationalized in the same manner and that the strengths and weaknesses of each rationalization be compared to each other. Rationality then involves selecting from competing beliefs *after* this process of rationalization has taken place.²⁷ Thus, “risk assessment provides a highly organized profile of the current state of knowledge of particular issues and systematically elucidates scientific uncertainties.”²⁸

V

RATIONALITY OF BELIEF IN RISK ASSESSMENT

Scientific reasoning refers to specific bodies of data and reasons from those data through models and theories. The quality of that reasoning depends on the scientist’s ability to satisfy two primary criteria: (1) agreement between any beliefs and the data; and (2) a satisfactory explanation of why the data exist in the form they do, that is, the mechanism by which the phenomenon operates. Stephen Toulmin has characterized these two criteria, respectively, as the properties of *foresight* and *understanding* in well-developed sciences.²⁹ The criteria are met by various combinations of five categories of evidence used routinely in risk assessment. These are described below.

A. Direct Empirical Evidence

This category of evidence includes direct measurements of the phenomenon of interest under the exposure conditions of interest. For example, the phenomenon of interest might be excess lifetime probability of cancer in humans under conditions of low exposure to a pollutant in water. Direct empirical evidence of risk, typically generated by either epidemiological or clinical studies, is

26. See, e.g., Peter Bush, *Uneasy Partners: A History and Analysis of the EPA’s Science Advisory Board 46* (1990) (unpublished manuscript, on file with the author).

27. See NATIONAL RESEARCH COUNCIL, *supra* note 6, at 27.

28. *Id.*

29. See STEPHEN TOULMIN, *FORESIGHT AND UNDERSTANDING* 18-42 (1961).

strongest when it is consistent with the criteria of science discussed above and constitutes the purest example of “good science.”

This evidence results from a study with the following nine features:

- (1) clearly articulated objectives or hypotheses and an objective-appropriate method of analysis;
- (2) proper selection and characterization of exposed and control groups;
- (3) adequate characterization of exposure;
- (4) follow-up of sufficient length to allow for the appearance of disease before the study ended;
- (5) valid ascertainment of causes of morbidity or mortality;
- (6) proper consideration of bias or confounding factors that might cloud the interpretation of results or skew the risk estimate;
- (7) adequate sample size to detect any possible effect;
- (8) adequate response rate and handling of missing data to ensure that the data stand up to a hard look; and
- (9) complete and clear documentation of results to ensure that discourse concerning the conclusions can take place within the scientific community.³⁰

These criteria determine whether a particular body of data provides an adequate basis to begin the reasoning process; they examine the *foundational* properties of the data. The goal of such an epistemological assessment is to ensure that only bodies of data that provide a sound foundation for reasoning are introduced.

Having established a firm foundation, the scientist must make the following five inquiries. The scientist must first determine whether the exposure and response were properly related in time. This inquiry asks whether the response follows the effect and whether the time elapsed between exposure and response is consistent with reasonable expectations. Second, the scientist asks whether the available studies are consistent. No single study may prove sufficient to establish either causality or the risk estimate. The conclusions are strengthened, therefore, if multiple studies obtain similar results. This consistency is essentially the criterion of *coherence* in the philosophy of science. The third step is to determine the magnitude of an association between exposure and effect. The stronger the association, or the higher the risk coefficient, the greater the likelihood that any response is due to the exposure rather than some confounding factor or bias. Fourth, the scientist inquires whether the associations are highly specific. If an effect occurs only under the exposure conditions of interest, there is a reduced possibility that the effect is due to either unknown factors or misclassifying exposures to known confounding factors. The scientist’s final inquiry is whether the effect is a plausible result of the exposure, given the state of

30. See ENVIRONMENTAL PROTECTION AGENCY, GUIDELINES FOR CARCINOGEN RISK ASSESSMENT (2000).

knowledge in biology. This question expresses the criterion of understanding in science.³¹ It focuses on the degree to which the scientific community is able to explain the occurrence of an effect through an understood mechanism that could be expected to yield the outcome. A scientific judgment of risk based on direct empirical evidence rests ultimately on the foundational quality of those data and the answers to the above five questions.

B. Semi-empirical Evidence

Risk assessments for a particular substance and health effect often are based on semi-empirical evidence. Such evidence is obtained from studies on the population of interest but differ from the situation of interest in some important aspects. The most common example is where the level of exposure is significantly above the level of interest to the decisionmaker. The scientist must extrapolate findings from the situation studied to the one of interest because the studied exposure level differs from the actual level. Extrapolation requires background assumptions, or *extrapolation premises*.³² The quality of these resulting risk estimates depends on two factors. The first factor is the quality of the judgments with respect to estimating the risk under the conditions of exposure found in the study. The criteria for assessing these judgments are those discussed in the previous section.³³ The second factor is the quality of any extrapolation premises required to estimate risk under the actual conditions of exposure.

The scientific criteria for judging these extrapolation premises are those used to judge models in general within science, because models are collections of assumptions. In particular, the model should:

- (1) agree with a well-defined set of relevant data at least under the conditions of the study, as measured by a goodness of fit test;
- (2) explain key phenomena identified by the scientific community;
- (3) have premises that correspond to scientific understanding of the process by which the effect is produced;
- (4) meet the criterion of falsifiability;
- (5) allow an understanding of the conditions under which the model's strengths and weaknesses will appear;
- (6) represent epistemic consensus in that an appropriate scientific community has judged the model acceptable;
- (7) to the extent possible, result from a neutral and objective assessment relative to other models, rather than being selected strategically; and
- (8) be applied under conditions of skepticism and conditional adoption, recognizing that it may eventually prove incorrect.

31. See Crawford-Brown, *supra* note 13, at 268.

32. See CRAWFORD-BROWN, *supra* note 15, at 54.

33. See discussion *supra* Part VA.

Risk estimates based on semi-empirical evidence typically do not have an epistemic status as high as those based on well-established direct empirical evidence. They can still, however, represent “good science” so long as the foundational quality of the data on which they are based is high, and the extrapolation premises are reasonably well established.

C. Empirical Correlations

At times, data in which the effect of interest has been generated will not exist, but there may be data that encompass other effects correlated with the effect of interest. For example, data on carcinogens do not exist, but data are available on transformation of cells *in vitro*. In that case, one might argue that substances causing such transformation also tend to be carcinogens. By using data with correlated effects, one draws inferences based on the empirical correlation between the two types of events, instead of drawing inferences from observations on the effect of interest. This process of reasoning and establishing the reliability of the inferences of risk based on empirical correlation requires two judgments: (1) the quality of the observations and (2) the quality of the association between the measured and target effects. The criteria for judging the quality of the observations are precisely those outlined in the discussion of direct empirical evidence.³⁴ One must remember, however, that such data justify only the claim that the measured effect occurs, not the claim that the effect of interest occurs.³⁵ To justify the occurrence of the effect of interest, one must also judge the quality of the association between the measured and target effect. The epistemic status of the risk estimate increases with the strength and specificity of the association. Strength and specificity, in turn, increase when the measured effect is characterized by the appearance of the target effect and the effect of interest usually appears only if the measured effect appears.

The usefulness of associations is controversial, since an individual’s view on their role in science may range from the view that all science rests ultimately on associations to the view that associations provide only the base of experience from which the process of scientific explanation may begin. Still, there is a significant tradition of using associations in scientific risk assessments.

D. Theory-based Inference

Under well-developed scientific theories, it is possible to calculate risk based on observations of phenomena intermediate between exposure and response.³⁶ This calculation is based on the theory or model in which the causal relationships leading from exposure to the effect of interest are well understood.³⁷ This calculation is not a case of empirical evidence because the effect of interest has

34. *See id.*

35. *See id.*

36. For example, this might include observations of DNA breakage or adduct formation.

37. For example, it might be claimed that DNA breakage is the first step in initiation within the progress of a cell to a full tumor.

not been observed directly. Nor is it empirical correlation because the measured phenomena are not simply correlated with the target effect. Instead, the measured phenomena are part of the causal chain leading to the effect, and the mechanisms linking the elements of this chain are understood well enough to calculate the final effect. The epistemic status of predictions from such evidence can be no stronger than the combined epistemic status of the models on which they are based and the data employed. As a result, the epistemic status of risk estimates based on theory-based inference are judged by the same criteria of data and model judgment specified in previous sections.³⁸

E. Existential Insight

Finally, risk estimates, or at least qualitative judgments of risk significance, may be based on expert opinion. Presumably, such opinions are the result of experience and training and, as such, may have a rational basis. To the extent the individual offering the opinion has experience and training in relevant areas of science, these opinions may be instances of scientific rationality. Robert L. Winkler claims that evidence of such relevant experience and training includes: (1) certifications such as degrees or professional training; (2) reputation of the individual as a reliable guide; (3) impartiality; and (4) multiplicity of viewpoints—for example, the judgment was reached only after considering multiple bodies of data and methods of interpretation.³⁹ The use of existential insight raises the issue of how these judgments should be elicited to reflect truthfully the expert's scientific opinion. The relevant issues in such elicitations include: (1) asking only questions whose answers would be formed on the basis of data (although if data are available, existential insight might not be needed); (2) forming teams of experts to reflect differences of opinion; (3) calibrating expert judgments so the scale in which judgments are expressed is understood and applied consistently; (4) providing a range of summaries of risk judgments, including best estimates, means, central tendency, and confidence intervals; and (5) creating an environment in which the scientific merits of an argument, rather than the legal or policy implications, are the focus of attention.⁴⁰

When using existential insight, one must consider that scientists with different institutional commitments, such as industry or academia, often differ systematically in those judgments.⁴¹ Again, understanding the communal nature of scientific judgments is essential for understanding the rational basis for judgments by individuals.

38. See discussions *supra* Parts VA, VC.

39. See Robert L. Winkler et al., *The Quality of Expert Judgment Elicitations*, at 2-2, Center for Nuclear Waste Regulatory Analyses, San Antonio, Texas (1992); *supra* Part III (discussion of the principle of neutrality and objectivity).

40. See *id.*

41. See generally Frances M. Lynn, *The Interplay of Science and Values in Assessing and Regulating Environmental Risks*, 11 SCI. TECH. & HUM. VALUES 40 (1986) (providing a range of examples of risk assessment questions, such as whether risks have been overstated for Three Mile Island, that are answered differently by scientists in industry, government, and academia).

F. Combining the Five Categories of Evidence

Typically more than one category of evidence will be available for use in a risk assessment. It is difficult to determine how these bodies of evidence and their lines of reasoning should be combined, especially when they lead to different conclusions. The practice of science does not provide anything but general guidelines. The final judgment of a scientific expert will depend on both the epistemic status of his beliefs formed on the basis of each of the five categories of evidence and on the weight assigned by that expert to each of these general categories. This weight may be viewed as an *intellectual obligation* assigned to each category. The belief formed on the basis of a given category will weigh more in the final judgment as both the evidence on which that belief is formed and the intellectual obligation for that category increase. Still, this general observation does not completely prescribe the process of forming judgments by the expert when often conflicting conclusions can be drawn from the five categories of evidence. The scientific community has an evident role in forming these judgments and community discourse provides the stage on which to play that role.

VI

MARGINS OF SAFETY, UNCERTAINTY, AND DEFAULT ASSUMPTIONS IN REGULATORY SCIENCE

One of the desiderata of rationality discussed in Part IV is epistemological reflection, which requires that the quality of scientific claims be analyzed systematically, strengths and weaknesses noted, and uncertainty characterized. Consideration of uncertainty does not indicate a weakness in science, needed only when the science fails to meet some standard of reliability. Instead, it is a necessary component of scientific analysis.⁴² Without such consideration, risk assessment does not fully incorporate rationality.

Science is not monolithic. Competing data, theories, models, and methods on which to base scientific estimates of risk always exist. As a result, there is a range of answers supported by the available scientific evidence and judgments. This fact does not imply that all answers are equally valid, or deserve the same status as “good science.” Any summary of the state of science, however, must reflect the existing uncertainty. Therefore, all competing risk estimates should be considered, the strengths and weakness of each estimate assessed, some measure of rational confidence assigned to each, and the results summarized by an expression of the uncertainty with which the expert is faced in making a judgment.

42. For example, the National Academy of Sciences calls explicitly for risk assessments to include a formal uncertainty analysis as part of scientific practice within the EPA. See COMMITTEE ON RISK ASSESSMENT OF HAZARDOUS AIR POLLUTANTS, *supra* note 6, at 12.

This feature of science in general and of risk assessment in particular has been formalized in regulatory risk assessment.⁴³ Faced with uncertainty, a strategy akin to the precautionary principle has been adopted, in which a “margin of safety” is built into the assessment by selecting a reasonable upper bound to the risk estimate from within the range of possible estimates.⁴⁴ The theory of this strategy is that the goal of regulatory decisions is to protect human health within a margin of safety, and that doing so requires the use of risk estimates that are likely to be too high if incorrect. Satisfying this goal has been met historically in two ways. First, when faced with uncertainty about a key assumption in a line of reasoning, such as the appropriate exposure-response model, a default assumption is adopted that is likely to be the most protective of health. This default assumption cannot be totally without scientific merit, but it should be one that produces risk estimates at the “upper end” of those made possible by scientific analysis.⁴⁵ Second, when faced with uncertainty about what parameter values to use in models, safety factors or uncertainty factors are adopted. Again, the resulting risk estimate is likely to lie at the “upper end” of risk estimates generated by scientific analysis.⁴⁶ The use of default assumptions and uncertainty factors thus represents one way to deal with the uncertainty inherent in scientific analysis.

Such an approach does not, however, fully meet the requirement of epistemological reflection, nor does it fully embody scientific rationality as previously described. It may also be inconsistent with the goals of the legal setting, where standards of proof are not necessarily guided by the precautionary principle. It replaces the characterization of uncertainty with a single risk estimate that is likely to lie at the “upper end” of the range of reasonable estimates. Those who use results from these default approaches, therefore, are not confronted with a full picture of uncertainty, and so cannot determine whether the science is “good.” Essential information about the state of that science—information needed to judge whether the science is capable of providing an apt basis for selecting means—is missing when only a single value is provided based on default assumptions. More important, one cannot assess the degree of conservatism built into the resulting risk estimate, and so cannot judge whether the confidence associated with the risk estimate is appropriate. This latter weakness occurs because one typically knows only that the default assumptions and uncertainty factors are likely to produce risk estimates at the “upper end” of the confidence interval. There is no further specification as to whether this “upper end” represents the 50th, 99th, or 99.99999th percentile.

To give the risk estimates a more scientific basis, a full uncertainty analysis is required. The result is a probability density function, or distribution, of the predictions of risk, showing the confidence with which any particular statement

43. *See id.* at 618-19.

44. *See id.*

45. *See id.*

46. *See id.*

of risk can be made in the face of existing evidence. Such an analysis provides a complete picture of the epistemic state of that science, showing the confidence that may be assigned to each of the competing risk estimates. If, in the end, a decisionmaker wishes to use a particular risk estimate from this characterization of uncertainty, the choice may be justified based on the goals and consequences of that particular decision. For this process to follow scientific rationality, however, it is necessary to understand the percentile represented by this single risk estimate, at least qualitatively, and to show that this percentile is a reasonable requirement for protective action or standards of proof. The simple use of default assumptions and uncertainty factors containing an unknown degree of conservatism will not satisfy this requirement, and so the use of such assumptions and factors does not generally constitute “good science” or even scientific rationality—although their use might still be rational in other senses, such as prudence or expediency. As risk estimates move between arenas of decision, such as from the regulatory setting to the court, it is essential to understand how uncertainty is being treated in the assumptions underlying these estimates, and whether that treatment is an apt means to reach the ends of that arena. This point becomes more difficult considering that the scientific community does not possess clearly defined rules to specify the point in an uncertainty distribution that should be transmitted to decisionmakers.

VII

CONCLUSION

Considerations of scientific rationality, and the link between science, policy decisions, and the law, led to the creation of the Science Advisory Board (the “SAB”) and the Science Advisory Panel (the “SAP”) at the Environmental Protection Agency (the “EPA” or “Agency”).⁴⁷ The original formation of the SAB resulted in part from the breakdown in scientific consensus in the late 1960s and early 1970s.⁴⁸ Faced with this disagreement, decisionmakers recognized uncertainty in the science and sought either to reduce that uncertainty or better characterize it, while still proceeding toward decisions.⁴⁹ The belief was that strong scientific review would in part reduce some of the uncertainty, as well as provide a justification for selection of particular scientific risk estimates from within the uncertainty distributions generated by risk assessments.⁵⁰ The result of these considerations is a role for the SAB in bringing together diverse groups of scientists so that competing views can be aired, and ensuring a larger role for science in policy decisions. This role of the SAB was championed by Congress.⁵¹

47. See Bush, *supra* note 26.

48. See *id.* at 5.

49. See *id.*

50. See *id.*

51. See *id.* at 38.

Peter Bush raised the issue of how uncertainty, a natural product of scientific analysis, may be used strategically to facilitate or block legal or policy decisions.⁵² During the years when Anne Gorsuch was Administrator, she repeatedly called for the use of “good science” in driving regulatory decisions.⁵³ Presumably, her goal was to raise the standard of minimal epistemic status required before a risk estimate could form the basis of a regulatory decision.⁵⁴ She was concerned that the use of default assumptions and uncertainty factors had become separated from the available science and no longer reflected standards of “good science.”⁵⁵ As a result, she argued for much greater scientific review by the SAB and other groups, such as the NAS, of risk estimates, and for use of assumptions and uncertainty factors that were more consistent with the best available science.⁵⁶

Industry welcomed this call for stronger use of science.⁵⁷ The environmental community and many within the Agency, however, argued that her call for “good science” was simply a strategic ploy to slow regulation.⁵⁸ It was possible, they argued, to raise the standard of minimal epistemic status so high that a burden of proof could never be satisfied, preventing all regulation.⁵⁹ Gorsuch and others countered by arguing that it also was possible to set the bar so low as to allow essentially all risk estimates, however unsupported, to enter the policy arena and drive regulatory decisions.⁶⁰ In this way, she argued, science could be manipulated by the legal and policy community at the Agency to support predetermined positions.

In the end, it is not clear which side was correct, since science provides no firm rules specifying how scientific rationality is to yield risk estimates in the face of different evidence. This historical episode does indicate the need to consider the role of science in regulatory decisionmaking, and how science’s uncertainty may be used by one side or the other to support particular approaches to regulation. Several lessons are germane here. First, science provides the least contentious, and most informative, basis for decisions when it gives a full characterization of uncertainty to the decisionmaker, rather than ignoring uncertainty or treating it by default science policy options. Second, bodies such as the SAB are essential to provide scientific rationality to regulatory decisions, especially if they systematically compare alternative bases for risk estimates. Third, complete certainty is not required for scientific rationality, or even for “good science.” At least the epistemic status of risk estimates, however, must be assessed and a standard of minimal epistemic status met before using that

52. *See id.* at 47.

53. *See id.* at 46.

54. *See id.* at 66.

55. *See id.* at 46.

56. *See id.* at 49-50.

57. *See id.* at 46.

58. *See id.* at 49.

59. *See id.* at 47.

60. *See id.*

science in decisions. Finally, science and law will remain “uneasy partners,” due to a fundamental conflict between how science works and “the public’s thirst for certitude that is written into the EPA’s laws.”⁶¹ Procedures that hide uncertainties in science and fail to recognize the essential role of discourse within the scientific community significantly weaken the rationality of subsequent decisions.

61. William D. Ruckelshaus, *Science, Risk and Public Policy*, 221 SCIENCE 1026 (1983).