HOW MUCH EVIDENCE IS ENOUGH? CONVENTIONS OF CAUSAL INFERENCE

DAVID KRIEBEL*

I

INTRODUCTION

One of the most important issues for science in the courtroom is the determination of causality. Like science in the courtroom, science in the regulatory arena can also bring a clash of cultures, misunderstanding, and controversy—especially when decisions must be made with some urgency with interested parties watching closely. In this article I will discuss some conventions in the conduct of science and in the ways that scientific information is communicated to nonscientists that can make it difficult for judges, lawyers, regulators, and politicians to do their jobs making decisions about complex environmental and health issues.

Of particular concern are the methods and conventions of causal inference as they are applied to controversies over whether chemicals and technologies are harming human health and the environment. There are far too many examples of environmental hazards that were permitted to be produced long after the evidence for harm was substantial.[1] Political and economic forces are partly to blame for many of these cases, but scientific methods—particularly methods of causal inference—could be improved to make it easier for society to assess when evidence of a hazard is sufficient to take action. This article briefly describes some of the different conventions of causal inference in different scientific fields. As opposed to that in "pure" sciences, causal inference in environmental health sciences must necessarily include consideration of the social responsibility to act in the face of uncertainty, which has implications for the scientific process. A key component of environmental health sciences, distinct from many other science disciplines, is the management and communication of uncertainty. This article presents some examples of how this might improve the contributions of science to environmental health problems.

This Article is also available at http://law.duke.edu/journals.

* Professor, Department Chair and Co-Director, Lowell Center for Sustainable Production, University of Massachusetts Lowell.

1. *See, e.g.*, EUROPEAN ENVIRONMENT AGENCY, LATE LESSONS FROM EARLY WARNINGS: THE PRECAUTIONARY PRINCIPLE 1896–2000 52, 64, 110 (2001) (documenting cases in which regulations lagged behind scientific evidence of harm posed by asbestos, lead, and polychlorinated biphenyls).

II

SETTING THE CONTEXT—WHAT'S SPECIAL ABOUT ENVIRONMENTAL HEALTH SCIENCE?

Most of the important environmental health crises share a fundamental characteristic: they appear to arise from disruptions of natural systems or cycles, whose behaviors are only partially understood.[2] Global warming, endocrine disruption, ecologic and health risks from genetically modified organisms, environmental breast-cancer risks—these are all hazards about which a great deal of uncertainty remains. Martin Krayer von Krauss, a scholar of policy applications of science, characterizes the problems this way:

> There is not one problem, but a tangled web of related problems[:] . . . The dynamics of the systems studied are not necessarily regular, but are characterized by synergistic [or] antagonistic relationships [or both], indirect relationships, long delay periods between cause and effect, [and] thresholds, or non-linear behaviours[.] The issue lies across or at the intersection of many disciplines[.] . . . [I]t has economic, environmental, socio-cultural, and political dimensions . . . .[3]

Because the human body, an ecosystem, a human society, or an economy are all complex dynamic systems, their behaviors are subject to fundamental uncertainties that will not be reduced no matter how long they are studied. This characteristic of complex systems sets them apart from many of the problems that western science has so successfully conquered. This inherent complexity is one defining feature of the terrain in which the environmental health sciences operate.

A second defining feature is the urgency to act to prevent harm; the environmental health sciences look more like medicine in this regard than, say, astronomy or geology. Public-health scientists often do not have the luxury of waiting for further study to tie up the loose ends before the need to act. In other words, science conducted to inform policy on disease prevention and environmental protection is different from that of most conventional science in fundamental ways: facts are uncertain, social values are in dispute, the stakes are high, and decisions are urgent.[4]

The central question of causality (Do greenhouse gases warm the planet? Does urban air-pollution kill people?) is often a stumbling block when "traditional" scientific perspectives are applied to problems like these because of the mistaken belief that there is a single truth that science can, with enough resources, identify. Although there may in principle be such a truth, in public health we recognize a duty to act when we are far from that truth.

---

2. *See* BARRY COMMONER, THE CLOSING CIRCLE: NATURE, MAN, AND TECHNOLOGY 293–300 (1971) (describing the complexity of relationships between social and environmental problems).

3. Martin Krayer von Krauss, Uncertainty in Policy Relevant Sciences 5 (Dec. 2005) (unpublished Ph.D. dissertation, Danish Technical University), *available at* http://orbit.dtu.dk/getResource?recordId=186518&objectId=1&versionId=1 (last visited Sept. 28, 2008) (bullets omitted).

4. *See* SILVIO O. FUNTOWICZ & JEROME R. RAVETZ, UNCERTAINTY AND QUALITY IN SCIENCE FOR POLICY 7–16 (1990) (explaining the importance of quality assurance in policy-related research).

III

LATE LESSONS FROM EARLY WARNINGS

There are too many examples of environmental hazards that have been allowed to be produced and used long after there was evidence of their hazardousness. Sadly, many of the most toxic chemicals introduced into widespread use in the early twentieth century are still causing illness and death despite overwhelming evidence of their effects: benzene, asbestos, and lead are three prime examples. It has been well documented that economic interests played a major role in blocking actions to control these hazards, and some effective voices have spoken out about this perversion of public health.[5] David Michaels and his colleagues have documented repeatedly how the manufacturing of uncertainty has been one of the principal tools of industries wanting to slow or prevent regulation of hazardous products.[6]

But these tragic histories of failure to act to prevent disease and pollution teach another important lesson: neither scientists nor policymakers are clear about how and when to determine that scientific evidence of risk is sufficient to provoke preventive actions. When there is uncertainty about the evidence for harm (and in complex systems, there always is), it creates the hesitation and confusion into which economic forces find it rather easy to insert their own interests. Thus, the two problems are linked and both are essential to resolve: corruption of science and poor understanding of how to proceed from scientific evidence to causal conclusion.

One can see the intertwining of these phenomena again and again in an outstanding report published by the European Environment Agency.[7] The report presents a series of case studies of environmental hazards documenting how evidence of hazardousness accumulated and at what points along the way causal judgments were made by relevant government and scientific organizations.[8] The report describes how lack of "sufficient" evidence of harm was misinterpreted as evidence of safety in fourteen different cases—including asbestos, lead, and polychlorinated biphenyls.[9]

How should uncertainty be managed? How can uncertain evidence be communicated to the public and policymakers in a useful way? How can less-

5. *See, e.g.*, SHELDON KRIMSKY, SCIENCE IN THE PRIVATE INTEREST: HAS THE LURE OF PROFITS CORRUPTED BIOMEDICAL RESEARCH? (2003) (examining the influence of corporate interests on public-health research); David Michaels, *Scientific Evidence and Public Policy*, 95 AM. J. PUB. HEALTH (SUPPLEMENT 1) S5 (2005) (explaining how judicial standards for measuring scientific validity can delay protective measures for public health); David Michaels et al., *Advice Without Dissent*, SCIENCE, Oct. 25, 2002, at 703 (arguing that special interests unduly influenced the selection of scientific advisory committees under President George W. Bush).

6. *See* Michaels, *supra* note 5, at S6 (explaining how the tobacco industry resisted regulation by challenging the validity of scientific data); Michaels et al., *supra* note 5, at 703 (predicting that ideologically unbalanced advisory committees would "emphasize the uncertainties of health and environmental risks, supporting the administration's antiregulatory views.").

7. EUROPEAN ENVIRONMENT AGENCY, *supra* note 1.

8. *Id.* at 192.

9. *Id.* at 52, 64, 110.

than-perfect evidence lead to appropriate policy responses? These are the topics this article will address.

IV

THE RULES OF CAUSAL INFERENCE WERE NOT HANDED DOWN ON STONE TABLETS

Scientists do not follow a single overarching "scientific method."[10] There are a small number of defining principles of science, but these are hardly a prescription for daily work. Nearly all scientists will agree that their work consists of formulating and testing hypotheses by applying logical reasoning and gathering evidence in a repeatable manner. But beyond these essential steps, a tremendous diversity in method and convention occurs. Indeed, different communities of scientists are defined by their methods—the tools they use to understand the world—as much as by the particular piece of the world they choose to study.[11] This is understandable given the lengthy training and practice that is generally needed to become adept at a sophisticated method like chemical analysis, gene sequencing, or climate modeling. Often, developing new methods or even combining existing ones requires a willingness to reach across disciplinary boundaries, learning the language and perspectives of other researchers. These cross-border excursions are notoriously difficult to carry out, and while interdisciplinarity is often proclaimed to be a laudable goal, there are many structural reasons why it remains exceptional.

One particular aspect of the scientific methods merits focus: the ways in which conclusions are drawn from observations, or how one decides if a hypothesized causal pattern is sufficiently supported by data to be accepted as real. For some scientists, the ability to make accurate predictions from a theory is seen as the critical confirmatory evidence (in physics, for example). But in observational fields (for example, epidemiology), this step is impractical or useless, and instead much weight is given to the accumulation of observations that appear to be coherent. When studying chronic disease, or climate change, it may be impractical or unethical to wait around to see if predictions are borne out as a way to confirm a theory. In some disciplines (for example, molecular biology), controlled experiments are essential, while this is rarely so in environmental sciences, in which the most informative experiments cannot be conducted for logistical or ethical reasons (there is only one atmosphere to

10. *See* Jerome R. Ravetz, Essay, *Conventions in Science and in the Courts: Images and Realities*, 72 LAW & CONTEMP. PROBS. 25 (Winter 2009).

11. For example, the scientists in the fields of surgical oncology, cancer biology, and cancer epidemiology all are concerned with the same disease, but in their daily work, the design of their research strategies and even much of their technical language, these researchers work in very different fields. The surgeon treats an individual human being, the biologist often works with cells in culture, and the epidemiologist uses statistics to detect patterns in population data. These differences are like cultural barriers which can be crossed, but with difficulty.

study; potential reproductive toxins cannot be administered to humans in double-blind trials).

The types of evidence used in arriving at a judgment that a particular chemical is a human carcinogen illustrate the complexity of the causal-inference process in environmental health science. Data on the geographic distributions of cancers may suggest a natural component of soil (as in the case of some asbestiform minerals that cause mesothelioma) or the emissions from a factory. Time trends in cancer frequency may be correlated with trends in consumption of a product (as in tobacco and lung cancer). The occurrence of cancers in highly exposed working populations is often viewed as carrying considerable inferential weight. Animal experiments have an essential role, but one that is limited by the problem of cross-species extrapolation. Finally, experimental knowledge of chemical pathways of cancer induction and predictions about chemical structures and their likely biologic activity can also be considered. Any one of these lines of argument is imperfect, and none is either necessary or sufficient. It is the accumulated preponderance of evidence that is assessed to reach a causal judgment.

Within the accepted practices of science is a very wide range of "rules of evidence" and, in particular, beliefs about how much and what kinds of evidence are necessary to reach a causal judgment. Two brief detours into very different fields of science may help to make the point.

A. Cause in Physical Anthropology and Paleobiology

In the fields of physical anthropology and paleobiology, it seems that Darwin's methods, and not just his conclusions, continue to loom large in the study of ancient life. The evaluation of causal hypotheses in paleobiology, following Darwin, gives considerable weight to "consili[e]nce of induction" and "identification of unique and quirky artifacts of history."[12] More recently, investigators also invoke the importance of the consistency of arguments and successful predictions from mathematical models.[13]

Consilience of induction was the method used by Darwin in marshalling the evidence for the theory of evolution.[14] The term means simply the converging of diverse lines of evidence, all of which are best explained by a single "cause." So, in Darwin's case, he argued for evolution using comparative anatomy, the fossil

12. Paul J. Morris, *Testing Patterns and Causes of Faunal Stability in the Fossil Record, with an Example from the Pliocene Lusso Beds of Zaire*, PALAEOGEOGRAPHY, PALAEOCLIMATOLOGY, PALAEOECOLOGY, Dec. 20, 1996, at 313, 319 (citing CHARLES DARWIN, ON THE ORIGIN OF SPECIES BY MEANS OF NATURAL SELECTION (Harvard Univ. Press 1964) (1859); CHARLES DARWIN, THE VARIOUS CONTRIVANCES BY WHICH ORCHIDS ARE FERTILISED BY INSECTS (Chicago Univ. Press 1984) (1877)).

13. Morris, *supra* note 12, at 319.

14. The method was first proposed by William Whewell. *See* 2 WILLIAM WHEWELL, THE PHILOSOPHY OF THE INDUCTIVE SCIENCES, FOUNDED UPON THEIR HISTORY 74 (1847) ("The Consiliences of our Inductions give rise to a constant Convergence of our Theory towards Simplicity and Unity.").

record, evidence from animal- and plant-breeding experiments, and other observations.

When nonexperts read of developments in physical anthropology, it often seems to them presumptuous when someone concludes, for example, that a new hominid species has been identified—seemingly on the basis of examination of a single, often quite incomplete skeleton and comparison of it to the bones of a handful of other individual skeletons in museums. From inside these debates, it may be clear that quite a number of different types of evidence have been marshaled, and that the weight given to the single skeleton is quite appropriate. Of course, peers may disagree with the conclusions, and often question the adequacy of the evidence. But there is no chance that these critics will demand experimental evidence or statistical power calculations when judging the quality of the research!

B. Cause in Molecular Biology

The second example is another life science, but one that is about as different as possible in the ways that causal inference is conducted. Molecular biology has achieved profound insights into the chemical and cellular machinery of life. It is largely a laboratory science that benefits from the power of experimental methods. Perhaps because of this ability to experiment while unraveling life's mysteries, molecular biology has been held up as a paragon of the power of rigorous and systematic investigation.[15]

The title of John Platt's 1964 article said it all: *Strong Inference—Certain Systematic Methods of Scientific Thinking May Produce Much More Rapid Progress than Others*.[16] Molecular biology was Platt's star example of rapid progress. This widely cited work argued for the importance in all scientific work of an approach that is essentially Francis Bacon's inductive inference, "familiar to every college student."[17] According to Platt, the essential elements are "(1) [d]evising alternative hypotheses; (2) [d]evising a crucial experiment . . . with alternative possible outcomes, each of which will, as nearly as possible, exclude one or more of the hypotheses; [and] (3) [c]arrying out the experiment so as to get a clean result."[18] In subsequent steps, new "subhypotheses" refine the conclusions from the first experiment. Platt argued that molecular biology was already (in 1964) showing such rapid progress because its practitioners followed this method, while other fields did not.[19]

Although Platt's *Strong Inference* has limited application in observational sciences like much of public health, the article remains valuable for its insistence on careful formulation of hypotheses and a preference for logical

15. *See* John R. Platt, *Strong Inference*, SCIENCE, Oct. 16, 1964, at 347, 348 (explaining that the structure and reporting of experiments in molecular biology is systematically inferential).

16. *Id.* at 347.

17. *Id.*

18. *Id.*

19. *Id.* at 348.

thought over mindless data gathering. The fact that it is frequently cited, particularly in molecular biology, indicates that researchers in this field do aspire to Platt's pure inductive inference—an approach that is very different from most environmental health sciences and different as well from paleobiology. Of course no one field of science is superior in its causal judgments to another; what should be clear, though, is that these rules of evidence are shaped by the tools and evidence that are available—by pragmatism and tradition.

Considering how diverse the accepted procedures are for weighing evidence and drawing conclusions among the sciences, scientists in environmental health science have license to consider which procedures are most appropriate in this particular field. This is especially important for applied sciences whose purpose is not simply to understand the world better (to puzzle out the human family tree, for example), but to inform policy decisions.

V

BRADFORD HILL AND THE NECESSITY OF ACTING UNDER UNCERTAINTY

At nearly the same time that John Platt was promoting strong inference, Sir Austin Bradford Hill was laying out a very different set of principles for causal inference in epidemiology.[20] Bradford Hill and his colleague Sir Richard Doll are probably the best known epidemiologists of the twentieth century. Bradford Hill proposed a set of considerations, sometimes incorrectly called criteria, by which one could arrive at a causal judgment about an environmental risk, such as smoking and lung cancer.

According to Bradford Hill, these nine considerations should be weighed when evaluating the causality of a particular exposure–disease association.[21] Five of these are most frequently cited:

1. Strength of association—the risk of disease should be substantially higher in an exposed group than an unexposed one;

2. There should be a "biological gradient"—evidence that those with higher exposure are at even higher risk than those at low or moderate exposure;

3. Consistency—repeated studies of the same association should find the same result;

4. Plausibility—there should be experimental evidence or a mechanistic understanding of how the exposure could conceivably cause disease; and

5. Temporality—there should be evidence that the cause preceded the effect.[22]

Bradford Hill was adamant that these considerations were not a checklist, nor were some more or less important than others: "What I do not believe . . . is

20. Austin Bradford Hill, *The Environment and Disease: Association or Causation?*, 58 PROC. ROYAL SOC'Y MED. 295 (1965).

21. *Id.* at 295–99.

22. *Id.*

considerations. "But this is not science!" some will protest; and they are correct. There may or may not be a single, true answer to the question, but in the real world, there will always be many uncertainties in the science. And in the gray area generated by the substantial uncertainties, it is realistic to expect other "nonscientific" issues to enter the debate. In other words, whereas the scientific research can tell us something about the costs, risks, and benefits of a proposed action, in the end there will have to be value judgments embedded in political decisions. Some would argue that the science should be kept entirely separate from these value judgments, but this is unrealistic because the scientific data used for making policy will nearly always be limited by uncertainty. Managing and communicating this uncertainty should be a responsibility of the scientists, but it cannot be separated entirely from the policy debates that it is meant to inform.

The precautionary principle has been a useful way to focus attention on the particular challenges of taking actions in the face of scientific uncertainty.[26] The definition of the precautionary principle developed for the Rio Declaration of 1989[27] is often cited, and the 1998 Wingspread Statement[28] contains similar language: "When an activity raises threats of harm to human health or the environment, precautionary measures should be taken even if some cause and effect relationships are not fully established scientifically."[29]

The term "precautionary principle" was introduced into English as a translation of the German word *Vorsorgeprinzip*.[30] An alternative translation might have been "foresight principle," which connotes more positive, anticipatory action, rather than precaution, which to many sounds negative.[31] The Wingspread Statement involves four central components of the principle:

1. taking preventive action in the face of uncertainty,
2. shifting the burden of proof to the proponents of an activity,
3. exploring a wide range of alternatives to possibly harmful actions, and
4. increasing public participation in decisionmaking.[32]

---

26. *See, e.g.*, David Kriebel & Joel Tickner, *Re-energizing Public Health Through Precaution*, 91 AM. J. PUB. HEALTH 1351 (2001) ("[T]he precautionary principle . . . promotes the search for safer technologies, encourages greater democracy and openness in public health policy, and stimulates reevaluation of the methods of public health science."); David Kriebel et al., *The Precautionary Principle in Environmental Science*, 109 ENVTL. HEALTH PERSP. 871 (2001) (explaining how environmental science can best protect human health and the environment by applying the precautionary principle).

27. *See* Conference on Environment and Development, June 3–14, 1992, *Rio Declaration on Environment and Development*, ¶ 15, U.N. Doc. A/CONF.151/26 (Aug. 12, 1992).

28. Science & Environmental Health Network, Wingspread Conference on the Precautionary Principle, Jan. 23–25, 1998, *Wingspread Statement on the Precautionary Principle*, para. 5, *available at* http://www.sehn.org/state.html#w [hereinafter *Wingspread Statement*].

29. *See also* PROTECTING PUBLIC HEALTH AND THE ENVIRONMENT: IMPLEMENTING THE PRECAUTIONARY PRINCIPLE 8 (Carolyn Raffensperger & Joel Tickner eds., 1999).

30. Kriebel & Tickner, *supra* note 26, at 1351.

31. *Id.*

32. *See* Raffensperger & Tickner, *supra* note 29, at 8–9 (condensing the text of the statement into

The precautionary principle can be useful in debates at the science–policy interface because it encourages consideration of the public (and broader environmental) good when decisions must be made under scientific uncertainty. When there is uncertainty about the risks and benefits of a proposed activity, the precautionary view is that decisions should be made in a way that errs on the side of caution with respect to the environment and the health of the public. But precaution is only a useful insight—it does not resolve the central problem of how to determine how much evidence is needed to take a certain action in a certain context.

Some examples of specific problems may help to illustrate the point:

1. The U.K. National Radiological Protection Board has advised "a precautionary approach to the use of mobile phone technologies," especially by children.[33] "There is no hard evidence at present that the health of the public, in general, is being affected adversely by the use of mobile phone technologies," the Health Protection Agency said in a press release accompanying the report, "but uncertainties remain and a continued precautionary approach to their use is recommended until the situation is further clarified."[34] In other words, they said there was enough evidence of risk only to recommend a cautious policy for a presumably vulnerable group, but not to make a flat assertion that there is a risk. This runs counter to the "risk-based approach" that is commonly taken in U.S. policy, whereby the first step in any policy decision is to assess the risk and decide whether or not it is "acceptable."[35]

2. The Federal Aviation Administration (FAA) does not permit the use of cellular telephones during "critical phases" of commercial airline flights.[36] This is based on concerns that the use of radio frequency transmitting devices may interfere with navigation or communication equipment.[37] This evidence is almost entirely anecdotal,[38] but despite this lack of experimental evidence, the

---

these four points).

33. NAT'L RADIOLOGICAL PROT. BD., MOBILE PHONES AND HEALTH 2004: REPORT BY THE BOARD OF NRPB 7 (2004) (U.K.), *available at* http://www.hpa.org.uk/web/HPAwebFile/HPAweb_C/1194947333240.

34. Press Release, Health Protection Agency, Mobile Phones and Health (Jan. 11, 2005) (U.K.), *available at* http://www.hpa.org.uk/webw/HPAweb&HPAwebStandard/HPAweb_C/1195733730486?p=1158945066117.

35. *See, e.g.*, EUROPEAN ENVIRONMENT AGENCY, *supra* note 1, at 181 (noting the danger of "paralysis by analysis," when risk assessment procedures delay action to protect public health); John C. Bailar III & A. John Bailer, *Risk Assessment—The Mother of All Uncertainties: Disciplinary Perspectives on Uncertainty in Risk Assessment*, 895 ANNALS N.Y. ACAD. SCI. 273, 285 (explaining that "[r]isk assessment is a process of analysis, not a specific kind of research and not a result, and it must be viewed as a process that is subject to much uncertainty").

36. Advisory Circular No. 91-21.1B, Federal Aviation Administration, Use of Portable Electronic Devices Aboard Aircraft 1 (Aug. 25, 2006), *available at* http://rgl.faa.gov/Regulatory_and_Guidance_Library/rgAdvisoryCircular.nsf/0/b5c85ae8e38df676862571da00529da4/$FILE/AC%2091.21-1B.pdf.

37. *Id* at 3–4.

38. *Portable Electronic Devices: Hearing Before the Subcommittee on Aviation of the Committee on Transportation and Infrastructure*, 106th Cong. 28 (2000) (statement of Paul McCarthy, Executive Air Safety Chairman, Air Line Pilots Association, explaining that "[a]lthough such reports might not be scientifically repeatable they are proof enough for the pilots that these devices should not be used

FAA holds that the catastrophic nature of the risk justifies a precautionary stance and that the onus is on the air carriers to show that use is *not* hazardous.[39] This example illustrates how the weight of evidence needed to make a policy decision depends on the consequences of being wrong.

3. Certain phthalate plasticizers have been removed from children's toys because of the possibility they may be harmful if ingested.[40] These chemicals are still permitted in many other products, including medical devices, because the manufacturers argue that risks are small and no good alternatives exist.[41] Thus, the lack of a clear social benefit for using these plasticizers in children's toys shifts the amount of evidence needed for its ban to less than it might be if its use were clearly indispensable.

In each of these cases, the chosen action would be easier to understand, question, and even revise if there were better methods of summarizing the weight of evidence and the amounts and types of uncertainties that researchers identify.

VI

MANAGING AND COMMUNICATING UNCERTAINTY

Managing and communicating uncertainty are not topics covered by John Platt in *Strong Inference*, while Austin Bradford Hill did address them (albeit in different terms) in his article on causal inference.[42] We are slowly learning that, in fields like environmental health science, it can be just as important to say what we don't know as what we do (or think we do) know. The eminent Australian scientist Ian Lowe said it well: "We have to start by recognizing that our understanding of nature–society interactions is still limited. . . . [M]odern science could still be described as islands of understanding in oceans of ignorance. We are constantly engaged in land reclamation, but there is no chance of filling in the oceans."[43]

At present, there is only a limited form of evaluation of error or uncertainty in the large majority of environmental health science research articles. Random error resulting from sampling is always evaluated using the familiar—but widely misunderstood—concept of statistical significance. Beyond this, there may be a

---

during critical phases of flight").

39. Matthew Brelis, *Despite Study, Ban Holds on In-Flight Cell Phone Use*, BOSTON GLOBE, Oct. 6, 1999, at D2 (quoting FAA spokeswoman Kathryn Creedy's explanation that "[r]egulations require airlines to prove to us that anything brought on board is not harmful to the system. If data is submitted, we will take a look at it and then make a judgment.").

40. *See* Katherine M. Shea, *Pediatric Exposure and Potential Toxicity of Phthalate Plasticizers*, 111 PEDIATRICS 1467, 1468 (2003) (noting "the removal of all phthalates from infant bottle nipples, pacifiers, teethers, and infant toys").

41. *Id*. at 1468, 1472.

42. *Compare* Platt, *supra* note 15, *with* Hill, *supra* note 20, at 300.

43. Interview by Robyn Williams with Ian Lowe, on *Ockham's Razor: Sustainability Science* (Radio National broadcast June 24, 2001), *available at* http://www.abc.net.au/rn/science/ockham/stories/s317194.htm.

qualitative examination of limitations of the findings, which is usually left to the Discussion section at the end of the article. Statistical significance, evaluated using p-values and confidence intervals, addresses only the magnitude of potential error in the statistical parameter estimates due strictly to sampling variability. But in observational studies of complex, poorly understood systems, this may be the least important source of uncertainty. Potentially more important are errors in the independent variables, errors arising from choice of the wrong form for the model(s) used to analyze and interpret the data, and biases from problems in the conduct of the study.

For example, a study of the effects of an environmental contaminant on sexual differentiation in fish would typically report the amount of sampling error around the final estimate of the degree of association found between the contaminant and several different measures of sexual development. But this would typically not take into consideration the error in measuring the levels of the contaminant in the fish or in the environment, and would not investigate the sensitivity of the findings to the choice of statistical models used to link exposure with reproductive outcome.

Explicitly addressing uncertainty is an important contribution of research because it clarifies what is known and unknown and thus stimulates further investigation. But there is also a strong desire on the part of scientists to be precise. This may result from a confusion of uncertainty of information with quality of information, but the two concepts are distinct.[44] Quality has been described as "the totality of characteristics of an object that bear on its ability to satisfy an established need."[45] Uncertainty (or lack of it) contributes to quality of information, but while uncertainty is a feature of the information itself, the information quality depends on its use and context. One can find many examples of highly uncertain information that is still of adequate quality.

In order to feel more comfortable with uncertain evidence, policymakers must accept that absolute certainty is not obtainable. But when they look to scientists for answers, both policymakers and the courts may still retain a deep hope or faith that there is a single truth or certainty that science can provide. The myth of certainty is convenient for scientists, too. Environmental research involves many assumptions, choices, and inferences based on professional judgment and standard practices. These should be made explicitly (and a priori), but in practice this is very hard to do completely. The many hidden assumptions make results appear more certain and less value-laden than they really are.

Scientists can do more to emphasize how little we know about environmental risks instead of only emphasizing what we do know. This leads to two desirable outcomes: recognition of the need for more and smarter research programs on environmental hazards, and removal of a major prop from the

44. FUNTOWICZ & RAVETZ, *supra* note 4, at 29 (noting that "information of low certainty may yet be of high quality for its function").

45. Krayer von Krauss, *supra* note 3, at 18.

myth that science can give clear and certain answers about environmental hazards.

VII
SOME WAYS FORWARD

> Old ideas give way slowly; for they are more than abstract logical forms and categories. They are habits, predispositions, deeply engrained attitudes of aversion and preference. Moreover, the conviction persists—though history shows it to be a hallucination—that all the questions that the human mind has asked are questions that can be answered in terms of the alternatives that the questions themselves present. But in fact intellectual progress usually occurs through sheer abandonment of questions . . . We do not solve them: we get over them. Old questions are solved by disappearing, evaporating, while new questions corresponding to the changed attitude of endeavor and preference take their place.
>
> John Dewey (1859–1952)[46]

"Standard practice" in science is not fixed but responds to new measurement tools, new ways of framing old problems, outside pressures, fads, and fashions—all within the bounds of what the scientific community will largely agree are acceptable methods. Environmental health scientists can still learn better ways to manage and communicate uncertainty and more effectively participate in deciding when the evidence is sufficient to act to protect health or environment.

For example, at least two systems are available for formally assessing and presenting the full(er) range of uncertainties (and therefore the level of confidence) in a scientific datum. These have not yet been used in many specific applications, but several research groups are currently promoting them. Funtowicz and Ravetz have proposed one of these systems, called NUSAP (after the first letters of each of the five components), for communicating uncertainty in quantitative information through standard scales or descriptors.[47] The idea is to provide the public and policymakers with a compact set of codes that are attached to a policy-relevant bit of data, which indicate how the scientists who have developed the datum think it should be weighed when making decisions. One of the leading proponents of this system describes NUSAP as a "Patient Information Leaflet" containing essential warnings, limitations, and potential pitfalls if the information is to be used in setting policy.[48] The five components of NUSAP are as follows:

> 1. Numeral: the actual bit of data.
>
> 2. Unit: the unit of measure, but also containing contextual information like the period of time or type of people [or] place to which the information applies.

46. JOHN DEWEY, *The Influence of Darwin on Philosophy*, *in* THE INFLUENCE OF DARWIN ON PHILOSOPHY, AND OTHER ESSAYS IN CONTEMPORARY THOUGHT 1, 19 (1910).

47. FUNTOWICZ & RAVETZ, *supra* note 4, at 28.

48. Jeroen P. van der Sluijs, *Uncertainty and Precaution in Environmental Management: Insights from the UPEM Conference*, 22 ENVTL. MODELLING & SOFTWARE 590, 596 (2007).

3. Spread: a range of plausible or expected results—confidence or credibility limits, for example.

4. Assessment: a qualitative judgment of the researcher about the importance of the number—such things as its "significance" or if it represents an "optimistic" or "pessimistic" estimate.

5. Pedigree: a descriptor of how the information was obtained—was this an educated guess, the result of a single study, or the output of an exhaustive systematic review of many studies?[49]

The first three of these are not very different from the way that scientific data are typically presented, whereas the last two introduce important and difficult new concepts designed to help users understand how much weight to give to a piece of scientific information.

The fourth component, assessment, is what many people think statistical significance does. The limitations of statistical significance and the chronic misuse of the concept have been well covered.[50] But, even though statistical significance is widely misunderstood and misused, one reason it has been so hard to dissuade researchers from using it is that it seems to fulfill an important function of "quality" assessment—to distinguish between what is real and what is not. As the literal meaning suggests, it would be useful to know if a datum is "significant" or not. The p-value does not actually provide this information, but the desire to have it is not wrong. Thus, one can understand the NUSAP assessment component as a separate bit of data attached to the main datum and attesting to the quality of the latter.

Another way to organize and present information on the uncertainty in scientific information used for policymaking identifies three essential dimensions of uncertainty: location, level, and nature.[51] A matrix can be constructed for each policy-relevant piece of information that presents a systematic summary of the associated uncertainties, indicating where in the methods or models the uncertainty occurs (location); how severe the uncertainty is—perhaps using a verbal descriptor or semiquantitative scales (level); and whether the uncertainty is due to imperfections in our knowledge and therefore theoretically reducible, or on the contrary due to inherent variability in natural phenomena and not reducible (nature).[52]

49. FUNTOWICZ & RAVETZ, *supra* note 4, at 28–29.

50. *See, e.g.*, Douglas G. Altman & J. Martin Bland, *Absence of Evidence Is Not Evidence of Absence*, 311 BRIT. MED. J. 485, 485 (1995) (observing that "non-equivalence of statistical significance and clinical importance has long been recognised, but this error of interpretation remains common"); Stephen N. Goodman, *Toward Evidence-Based Medical Statistics. 1: The P Value Fallacy*, 130 ANNALS INTERNAL MED. 995, 995 (1999) (noting that "the methods of statistical inference in current use are not 'evidence-based' and thus have contributed to a widespread misperception"); Kenneth J. Rothman, *Significance Questing*, 105 ANNALS INTERNAL MED. 445, 445 (1986) (explaining that "[w]ith the focus on statistical significance, if chance seems to be a plausible explanation, then other theories are too readily discarded, regardless of how tenable they may be").

51. Walker et al., *Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support*, INTEGRATED ASSESSMENT, Mar. 1, 2003, at 5, 8.

52. *Id.* at 9–14.

These systems are as yet in only limited use, and much more work needs to be done before scientists will feel comfortable with any formal codification of uncertainty. Yet we must move in this direction.

In the meantime, there are a number of less comprehensive yet better-accepted ways for researchers to explore and communicate aspects of uncertainty. Bayesian statistics provides one such partial solution. The Bayesian view of statistical inference starts from accepting that we have beliefs about the phenomena under study, and it seeks to formalize the role these play in the way we view information.[53] Bayesian statistical methods correspond more closely to everyday understanding of logical inference and have the potential to allow more-transparent communication of the confidence that an investigator has in data. Methods-development work is still needed in most fields, however, before Bayesian statistics can be routinely applied. Another practice that is slowly gaining ground is sensitivity analysis, in which investigators assess the degree to which the main results are changed by using different assumptions or analytic methods.[54] Sensitivity analyses should probably become standard in every policy-relevant research article.

VIII

CONCLUSION

Whether cause exists is a judgment—the result of an inferential process. The important policy question is when do we have enough information to act as if it is a cause, while being prepared to be wrong? Framing causal inference as a purely scientific problem is inappropriate in fields like public health, where there is a duty to act. Instead, we have to ask, When do we have enough information to act as if something is a cause? And the answer is very context dependent. It depends on what the alternatives are and on the risks of being wrong. These are not fundamentally scientific considerations, although science can assist.

This is one of the crucial ways in which scientific research and policymaking are linked. Scientists, while maintaining objectivity as much as possible and remaining mindful not to undermine the legitimacy that society grants them as purveyors of valuable information, must nevertheless engage with the public and policymakers if they are to uphold their social responsibility. We are currently witnessing an extremely important and unprecedented example of this responsible social engagement from the scientists of the Intergovernmental Panel on Climate Change. They have managed to communicate an important yet complex message: Much more needs to be learned about climate, *and* we

---

53. *See* BRADLEY P. CARLIN & THOMAS A. LOUIS, BAYES AND EMPIRICAL BAYES METHODS FOR DATA ANALYSIS 6 (1996) ("Bayesian formalism encourages documentation of assumptions.").

54. *See* Sander Greenland, *Basic Methods for Sensitivity Analysis of Biases*, 25 INT'L J. EPIDEMIOLOGY 1107, 1115 (1996) (explaining the need to address systematic bias and reviewing the basic methods of bias assessment).

know enough that we cannot remain silent about the need for action.[55] These scientists have stepped out of their roles as data gatherers and analysts and have spoken publicly about the need for action. Researchers in other environmental health fields should follow their lead.

---

55. *See, e.g.*, INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE, CLIMATE CHANGE 2007: SYNTHESIS REPORT 58 (2007), *available at* http://www.ipcc.ch/pdf/assessment-report/ar4/syr/ar4_syr.pdf ("Both bottom-up and top-down studies indicate that there is *high agreement* and *much evidence* of substantial economic potential for the mitigation of global GHG emissions over the coming decades that could offset the projected growth of global emissions or reduce emissions below current levels.").