Journal of Law and Policy

Volume 15
Issue 1
SCIENCE FOR JUDGES VII:
Evaluating Evidence of Causation & Forensic
Laboratories: Current Issues and Standards

Article 6

2007

The Nature and Necessity of Scientific Judgement

Douglas L. Weed

THE NATURE AND NECESSITY OF SCIENTIFIC JUDGMENT

Douglas L. Weed, M.D., Ph.D.[*]

"Judgment is a utensil proper for all subjects and will have an oar in everything."[1]

INTRODUCTION

Judgment sits at the center of the intersection where science, law, and policy meet. It is regularly invoked when scientific evidence is used to make a claim about disease causation in the courts[2] or in regulatory risk assessment.[3] Judgment also features

---

[*] Chief, Office of Preventive Oncology, and Director, Cancer Prevention Fellowship Program, at the National Cancer Institute; M.D. Ohio State University, Ph.D. University of North Carolina at Chapel Hill. For many helpful comments and suggestions, the author is grateful to Dr. Joe Cecil at the Federal Judicial Center, Dr. Graca Dores at the National Cancer Institute, and Dr. Robert McKeown at the University of South Carolina.

[1] 8 MICHEL DE MONTAIGNE, *Of Democritus and Heraclitus, in* THE COMPLETE WORKS OF MICHEL DE MONTAIGNE 266, 266-68 (Alfred P Knopf ed. Donald M. Frame Trans. 2003).

[2] *See* CARL F. CRANOR, TOXIC TORTS: SCIENCE, LAW, AND THE POSSIBILITY OF JUSTICE 142 (2006). "Implicit in scientific inference is the role of professional judgment." Cranor notes that causal inferences made by experts in toxic tort litigation involve judgment at several steps: reviewing, selecting and weighing the relevant scientific data (evidence), incorporating their background understanding of the underlying biology (and/or toxicology), evaluating the different possible explanations (e.g. noncausal explanations) in light of all the evidence, and finally, making a statement about the most likely explanation for the available evidence. *See also* Jerome P. Kassirer & Joe S. Cecil, *Inconsistency in Evidentiary Standards for Medical Testimony: Disorder in the Courts*, 288 JAMA, 1382, 1382-87 (2002). "In the final analysis, assessment of evidence and causal inferences (in the courts) depend

prominently when scientific evidence is used to support the need for a public health intervention[4] or a medical treatment.[5] From

---

on accumulating all potentially relevant evidence and making a subjective judgment about the strength of the evidence." *See also* Michael D. Green, Michal Freedman & Leon Gordis, Reference Guide on Epidemiology *in* Reference Manual on Scientific Evidence 375, Federal Judicial Center (2000) ("Drawing causal inferences after finding an association and considering these factors (i.e. causal guidelines based on Hill's so-called causal criteria) requires judgment and searching analysis. . . .").

[3] *See, e.g., Guidelines for Carcinogen Risk Assessment*, United States Environmental Protection Agency, 2-49-51 (2005). In the first stage of the process of regulatory risk assessment—often referred to as hazard identification—the final summary statement is called the "weight of evidence narrative" within which the committee summarizes its decision regarding the carcinogenicity of a chemical in terms of one of the following descriptors: "carcinogenic to humans," "likely to be carcinogenic to humans," "suggestive evidence of carcinogenic potential," "inadequate information to assess carcinogenic potential," or "not likely to be carcinogenic to humans." As the guidelines clearly state: "choosing a descriptor is a matter of *judgment* and cannot be reduced to a formula." (emphasis added). *See also* Preamble, *International Agency for Research on Cancer (IARC) Monograph on the Evaluation of Carcinogenic Risks to Humans,* World Health Organization International Agency for Research on Cancer, 3 (2006) ("The evaluations of IARC Working Groups are scientific, qualitative *judgments* on the evidence for and against carcinogenicity provided by the available data.") (emphasis added). *See also* H. OTWAY, H. & D. VON WINTERFELDT, EXPERT JUDGMENT IN RISK ANALYSIS AND MANAGEMENT: PROCESS, CONTEXT, AND PITFALLS, RISK ANALYSIS 92 (1992) ("The use of formal expert *judgment* in policy arenas (involving the selection and evaluation of new and alternative technologies) is likely to increase.") (emphasis added).

[4] Causal inferences in epidemiology and other public health disciplines have relied upon "aides" to judgment (*see* J.J. Schlesselman, *"Proof" of Cause and Effect in Epidemiologic Studies: Criteria for Judgment*, 16 PREVENTIVE MEDICINE 195, 195-210 (1987)), often referred to as "criteria," that is, considerations that assist judgments about disease causation (*see* M. SUSSER, CAUSAL THINKING IN THE HEALTH SCIENCES (1973); M. Susser, *Judgment and Causal Inferences: Criteria in Epidemiologic Studies*, 105 AM. J. OF EPIDEMIOLOGY 701, 701-15 (1977)), or that serve as the centerpiece of a mixed quantitative and qualitative methodology within which judgment plays a prominent role (*see* D.L. Weed & L.S. Gorelic, *The Practice of Causal Inference in Cancer Epidemiology*, 5 CANCER EPIDEMIOLOGY, BIOMARKERS & PREVENTION 303, 303-11 (1996)). As Weed and Gorelic note, "At its core, causal inference (in epidemiology) is a complex matter of

every corner of the complex crossroads of science and law and policy, the judgment of scientific experts is now hailed; it is "essential and irreducible,"[6] "required,"[7] a "virtue."[8]

reasoned judgment . . ." *Id.* at 308. These so-called criteria typically include: consistency, strength, biologic gradient (dose-response), temporality, experimental evidence, biological plausibility, coherence, specificity, and analogy. S*ee* Austin Bradford Hill, *The Environment and Disease: Association or Causation?* 58 PROC. ROY. SOC. MED. 295, 295-300 (1965). For a discussion of these criteria in the context of general causation, see D.L. Weed, *Causation: An Epidemiological Perspective*, 12 J. L. & POL'Y 43, 43-53 (2003). For a historical overview of causal inference in epidemiology and its application to the prevention of cancer, see D.L. Weed, *Causal and Preventive Inference in* CANCER PREVENTION AND CONTROL 285, 285-302 (Greenwald, Kramer & Weed, eds. 1995). Causal inferences are important components of the process of making public health recommendations regarding preventive interventions.

[5] For therapeutic decision making, clinical judgment typically teams up with scientific evidence. *E.g.,* Cynthia D. Mulrow & Kathleen N. Lohr, *Proof and Policy from Medical Research Evidence*, 26 J. HEALTH POL. POL'Y & L. 249, 249-66 (2001) ("When judging the benefits and harms of health care and predicting patient prognosis, clinicians, researchers, and others must consider many types of evidence."). However, Edmund D. Pellegrino points out that there are also other components of clinical judgment: language, rhetoric, critical thinking and logic, as well as the psychosocial, cultural, and ethical dimensions of clinical care decision making. Edmund D. Pellegrino, *Commentary: Clinical Judgment, Scientific Data, and Ethics: Antidepressant Therapy in Adolescents and Children*, 184 J. NERV. MENT. DIS. 106, 106-08 (1996). These non-scientific components of decision-making must be recognized and rigorously addressed: Pellegrino writes that "Any model of clinical thinking that uses only the scientific paradigms is insufficient to encompass all things involved in clinical judgments." *Id.* at 107.

[6] CRANOR, *supra* note 2, at 143. "The role of judgment is essential for the individual scientist (the expert) and (for the) consensus scientific committees (i.e. groups of experts)." He adds, ". . . there is an (almost) irreducible scientific judgment that enters into assessing scientific evidence and inferring that a substance causes a disease." *Id.* at 152.

[7] R.A. Carpenter has described the necessary role for judgment in environmental risk assessment: "Judgment on both technical and nontechnical issues and on their interaction is (thus) required." R.A. Carpenter, *Scientific Information, Expert Judgment, and Political Decision Making*, 18 J. OCCUP. MED. 295 (1976).

[8] ". . . [A]ll empirical investigation demands the same epistemic virtues:

Traditionally, however, many scientists and laypersons alike have held that, ideally, science should be free of judgment and its subjective influences, free to discover the causal laws that explain the world.[9]

Nonetheless, no one has ever found a way to circumvent the need for judgment, perhaps because the systems studied by biomedical sciences offer no universal laws, or because no science can offer the proof (nor disproof) upon which mathematics thrives,[10] much less a definitive algorithm for establishing causal relationships.[11] Perhaps we cannot do without

---

respect for evidence, care and persistence in seeking it out, good judgment in assessing its worth . . . ." SUSAN HAACK, DEFENDING SCIENCE WITHIN REASON: BETWEEN SCIENTISM AND CYNICISM 167 (2003).

[9] Philip Kitcher calls this "Legend," the idea that science is free of anything but rationality. He writes that there have been those who believe "science is a clearing of rationality in a jungle of muddle, prejudice, and superstition." PHILIP KITCHER, THE ADVANCEMENT OF SCIENCE: SCIENCE WITHOUT LEGEND, OBJECTIVITY WITHOUT ILLUSION 3-4 (1995).

[10] The failure of scientific evidence to provide definitive proof (or disproof) is called "underdetermination." Within many circles in the philosophy of science, it is the "most obvious" and "most troublesome" problem for science in the area of theory appraisal, i.e. deciding which theory gives the best explanation for the available evidence. E. McMullin, *Underdetermination*, 20 J. MED. PHILOS. 233, 241 (1995). At a more practical level, such as the testing of hypotheses about disease causation in population-based studies, underdetermination is a constant companion of the epidemiologist. "Typically, scientific evidence underdetermines the hypotheses tested in (epidemiologic) research studies, providing neither proof nor disproof." D.L. Weed, *Underdetermination and Incommensurability in Contemporary Epidemiology*, 7 KENNEDY INST. ETHICS J. 107 (1997).

[11] Most commentators agree that there is no decisive algorithm—i.e. no methodology—for establishing causation that avoids the need for judgment. *E.g.*, Susser, CASUAL THINKING, *supra* note 4, at 140. ". . . [T]here are no absolute rules (for causal inference). . . ." *See also* J. Doull, K. K. Rozman, & M. C. Lowe, *Hazard Evaluation in Risk Assessment: Whatever Happened to Sound Scientific Judgment and Weight of Evidence?* 28 DRUG METAB. REV 285, 291 (1996) ("It has long been recognized that there are relatively few absolutes in biology, and that any interpretation of observed phenomena must be tempered by sound scientific judgment.") Wandall notes that the process of accepting a (causal) hypothesis on the basis of evidence has no accompanying algorithm nor decision rule and therefore must rely upon value

judgment because it provides us with a way to finish off the essential scientific puzzle:[12] laying claim to the causal relationships that connect the physical "real" world with our perceptions of that world.[13] Neither carefully crafted observations nor the accuracy of causal hypotheses are sufficient for this purpose. We need something else to help us decide what those causal relationships are. We need judgment.

Expert judgment is inescapable and is as important as all that which surrounds science, feeds it, and feeds upon it: the funding and politics, the professional societies and journals, the universities, corporations, and the research institutions. It is as much a part of science as the theories, hypotheses, methodologies, and evidence. Judgment is an integral part of the history of science, its ethic, and its philosophy. In short, science would not be science without judgment.

---

judgment, including epistemic (instrumental-goal-oriented) values. Birgitte Wandall, *Values in Science and Risk Assessment*, 152 TOXICOL. LETT. 265, 267 (2004). *See also* Green, Freedman & Gordis, *supra* note 2, at 375 ("There is no formula or algorithm that can be used to assess whether a causal inference is appropriate based on these (causal) guidelines.") Susan Haack expresses this idea as follows: "When something requires judgment, it can't be formalized into rules or some kind of decision-procedure that can be followed mechanically." E-mail to author from Susan Haack, July 5, 2006.

[12] Thomas Kuhn links scientific problem-solving with puzzles in an immensely influential book. THOMAS KUHN, THE STRUCTURE OF SCIENTIFIC REVOLUTIONS 36-39 (3rd ed. 1996).

[13] The idea that science involves the interconnections of distinct "worlds" is attributed to the philosopher Karl Popper, whose theory of scientific method is laced with falsifications. Interestingly enough, the Supreme Court cited his theory in their landmark *Daubert* decision. Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579, 593 (1993). *See* D. Goodstein, *How Science Works in* REFERENCE MANUAL ON SCIENTIFIC EVIDENCE, Federal Judicial Center 67-82 (2000); Susan Haack, *Trial and Error: The Supreme Court's Philosophy of Science*, 95 AM. J. PUB. HEALTH S66, S66-S73 (2005). Popper proposed that the practice of science is comprised of three different worlds: that of physical states, another of our perceptions and observations of that physical world, and a third world of ideas. In Popper's third world lie theories, hypotheses, and causal explanations. KARL POPPER, OBJECTIVE KNOWLEDGE: AN EVOLUTIONARY APPROACH 154-56 (rev. ed., 1981).

That said, scientific judgment is not easy to define, although we are fairly clear about the kind of judgment we prefer: good, sound, and unbiased.[14] We are much better, in fact, at describing the biases that affect judgment than we are at defining judgment itself. Overconfidence, anchoring, availability bias, conflicts of interest, mindsets and ideologies, and "wish" bias are commonly cited,[15] any of which may shift a particular expert's judgment from "unbiased" to "biased." Although this shift does not necessarily signal an accompanying shift from

---

[14] *See* SUSAN HAACK, DEFENDING SCIENCE WITHIN REASON: BETWEEN SCIENTISM AND CYNICISM 167 (2003) for a definition of "good judgment." Haack has also described "shrewd judgment." Susan Haack, *An Epistemologist Among the Epidemiologists*, 15 EPIDEMIOL. 522 (2004). *See also* Doull et al., *supra* note 11, at 291. "It has long been recognized that there are relatively few absolutes in biology, and that any interpretation of observed phenomena must be tempered by *sound* judgment." (emphasis added). For a discussion of the biases that can affect scientific judgment, see *infra* note 15. *See also* OTWAY & VON WINTERFELDT, *supra* note 3.

[15] *See id*. They describe a wide range of biases that can affect the process of expert judgment: *mindsets and ideologies* that spring from shared experiences and the conventional wisdom of a scientific discipline, *cognitive biases* such as overconfidence, anchoring, and availability bias, *motivational biases* that track closely with conflicts of interest, and mindsets and ideologies reflecting their respective scientific communities. *Overconfidence* is the tendency for experts to be more certain about probability estimates than their knowledge can justify. *Anchoring* occurs when an expert holds to his original estimate and fails to adjust it sufficiently as new or disconfirming evidence accumulates. *Availability* is a cognitive bias in which there is a tendency to base judgments on readily recalled information rather than the full body of available information. E.L. Wynder, I.T. Higgins & R.E. Harris, *The Wish Bias*, 43 J. CLIN. EPIDEMIOL. 619, 619-21 (1990) is an interesting example of a bias that can directly affect a scientist's claims about disease causation: it occurs in those situations wherein the scientist has published a study on a particular exposure-disease relationship and then writes a review of the literature within which a causal claim is made about that same relationship. Such a claim can be biased if the investigator is influenced in their overall conclusion regarding causation by the direction of the results of their own study, in either direction, positive or null. The opportunity for "wish" bias is particularly prominent because typically those scientists reviewing the literature have previously published studies on the topic. *See* Weed, *supra* note 10, at 107.

"sound" to "unsound" or from "good" to "bad," we often regard it as suspect, and rightly so. The fact that biases can affect one's judgment suggests that judgment can be improved upon by learning to recognize and avoid the many sources of bias.

Some argue that groups of experts, who may better recognize individual biases, should make decisions about causation, thereby circumventing biased conclusions.[16] However, expert groups are not in fact protected from bias; mindsets and ideologies can be important sources of bias that arise from within a scientific community, representing the common and conventional wisdom of the discipline shared by its many members.[17] For example, epidemiologists may discount the need for fully understanding the biological mechanism of a purported causal relationship because they follow a popular disciplinary maxim (or, rule-of-thumb) that the biological plausibility of a causal association—and so the extent to which its mechanism *can* be known—depends upon the biological knowledge of the day.[18]

---

[16] Daniel Gilbert opines that it is often easier to see bias in the decisions of others than in one's own decisions. He writes: "Research suggests that decision-makers don't realize how easily and often their objectivity is compromised . . . Much of what happens in the brain is not evident to the brain itself." Daniel Gilbert, *I'm O.K., You're Biased*, N.Y. TIMES, Apr. 16, 2006, at A12. On the other hand, Gilbert also notes that we tend to overestimate and exaggerate the effects of bias on others' judgments. "Research shows that while people underestimate the influence of self-interest on their own judgments and decisions, they overestimate its influence on others." *Id.*

[17] *See* OTWAY & VON WINTERFELDT, *supra* note 3, at 91, for a discussion of mindsets and ideologies as sources of bias.

[18] Weed and Hursting have examined the rules of evidence aligned with the so-called criterion of "biological plausibility" in the practice of causal inference in epidemiology. D.L. Weed & S.D. Hursting, *Biologic Plausibility in Causal Inference: Current Method and Practice*, 147 AM. J. EPIDEMIOL. 415, 415-25 (1998). A popular and widely cited rule (of thumb) is that biological plausibility—one of Austin Bradford Hill's original 1965 causal considerations—is dependent upon the knowledge of the day and therefore cannot be required. The growth of molecular biology and, in particular, molecular epidemiology, may change this widely accepted "mindset" in the future. Indeed, on the other side of this coin, there are others who place an

Such a low threshold for the mechanistic understanding of disease causation could bias groups (and, for that matter, individuals) toward making premature causal judgments.[19]

One potential solution to this problem is to select scientists from many different disciplines to sit on panels formulating causal conclusions, as is commonly done by regulatory agencies and less commonly by the courts in toxic tort litigation.[20] Unfortunately, another bias may be introduced if members of one discipline discount the results of another because they believe those results to be generally unreliable. Epidemiology, with its emphasis on non-experimental, observational studies, is sometimes singled out in this way, as if it were the only science with "limits."[21]

Expert judgment, although not easily defined and subject to bias, appears to perform many different functions. It produces (or, more precisely, is required in the production of) conclusions,[22] decisions,[23] evaluations,[24] inferences,[25]

---

exceedingly high threshold on evidence in support of biological plausibility; for them, causation cannot be claimed until the mechanism involved has been well established. *Id*.

[19] It can also be argued that lowering the threshold for making causal claims is a precautionary measure rather than a premature one. *See* D.L. Weed, *Precaution, Prevention, and Public Health Ethics*, 29 J. MED. PHILOS. 313, 313-32 (2004).

[20] *See* W.W. Schwarzer & J.S. Cecil, *Management of Expert Evidence, in* REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 59-63 (Federal Judicial Center, 2000), for a discussion of the legal authority and practical considerations, and for examples of court-appointed scientific experts.

[21] Epidemiologists are sometimes their own worst enemy on this particular issue. In a widely publicized news article in *Science* magazine, entitled "Epidemiology Faces its Limits," prominent epidemiologists (who later claimed they were misrepresented) noted that "people don't take us [epidemiologists] seriously," that "we're pushing the edge of what can be done with epidemiology" and that "epidemiology is stretched to its limits or beyond." Gary Taubes, *Epidemiology Faces its Limits*, 269 SCIENCE 164, 164-69 (1995). Furthermore, epidemiologists sometimes regard toxicology as subject to bias through vulnerable extrapolations across dosage levels and species. Cecil, personal communication, August 18, 2006.

[22] Carpenter, *supra* note 7, at 295, notes under the heading "Expert Judgment" that an advisory committee (is) convened to translate technical

uncertainties,[26] and last, but not least—circularly and ambiguously enough—judgments.[27] These "judgments" concern all kinds of causes and risks: estimating the risk of cancer from exposure to chloroform or radon in drinking water,[28] characterizing exposure to benzene,[29] assessing the safety of a new bicycle crossing design,[30] calculating the likelihood that

information into conclusions for public policy.

[23] "It is a widely accepted claim that scientific practice contains value judgments, i.e. decisions made on the basis of values." Wandall, *supra* note 11, at 265.

[24] "The evaluations of IARC Working Groups are scientific, qualitative judgments on the evidence for and against carcinogenicity provided by the available data." World Health Organization, International Agency for Research on Cancer, IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Preamble, Lyon, 2006. *See also* Wandall, personal communication, June 13, 2006. "On a very general level, I would say that a judgment is a kind of evaluation . . . ." *See also* OTWAY & VON WINTERFELDT, *supra* note 3, at 84 ("Judgments are inferences or evaluations that go beyond obvious statements of fact, data, or the conventions of a discipline.").

[25] *See* OTWAY & VON WINTERFELDT, *supra* note 3, at 84.

[26] "In some cases, scientific judgment is used by a Bayesian analysis to estimate (the) uncertainties." D.J. Crawford-Brown & C.R. Cothern, *A Bayesian Analysis or Scientific Judgment of Uncertainties in Estimating Risk Due to ^{222}Rn in U.S. Public Drinking Water Supplies*, 53 HEALTH PHYSICS 11 (1987).

[27] MORRIS R. COHEN & ERNEST NAGEL, AN INTRODUCTION TO LOGIC AND SCIENTIFIC METHOD 28 (1934) ("A judgment is an ambiguous term, sometimes denoting the mental act of judging and sometimes referring to that which is judged.").

[28] *See* John S. Evans, et al., *Use of Probabilistic Expert Judgment in Uncertainty Analysis of Carcinogenic Potency*, 20 REG. TOX. AND PHARMACOL., 15, 15-36 (1994) (asking experts in cancer biology/toxicology, pharmacokinetics, and dose-response modeling to make judgments about carcinogenic potential). *See also* Crawford-Brown & Cothern, *supra* note 26.

[29] Katherine D. Walker, John S. Evans & David MacIntosh, *Use of Expert Judgment in Exposure Assessment. Part I. Characterization of Personal Exposure to Benzene*, 11 J. EXP. ANAL. ENVIRON. EPIDEMIOL. NO. 4, 308-22 (2001).

[30] *See* L. Leden, P. Garder & U. Pulkkinen, *An Expert Judgment Model Applied to Estimating the Safety Effect of a Bicycle Facility*, 32 ACCIDENT

medications cause adverse drug reactions,[31] and attributing lung cancer mortality to individual risk factors.[32] The International Agency for Research on Cancer has also employed its judgment to assess the carcinogenic potential of a list of chemicals now several hundred compounds long.[33]

We should not be surprised by the pervasiveness, utility, and flexibility of judgment. It is, after all, a necessary component of scientific experts' causal claims and should therefore be useful in all sorts of specific situations. Judgment is not unique to science and it finds its way into many other aspects of life. As the French philosopher, Michel de Montaigne, reminds us from many centuries ago: "Judgment is a utensil proper for all subjects and will have an oar in everything."[34] Value judgments,[35] moral judgments,[36] and the legal decisions—the

---

ANAL. PREV. 589, 589-99 (2000).

[31] *See* Yannick Arimone et al., *Agreement of Expert Judgment in Causality Assessment of Adverse Drug Reactions,* 3 EUR. J. CLIN. PHARMACOL. 61, 169-73 (2005).

[32] *See* Elizabeth A. Casman & M. Granger Morgan, *Use of Expert Judgment to Bound Lung Cancer Risks*, 39 ENVIRON. SCI. & TECH. 5911, 5911-20 (2005).

[33] At present, the IARC has evaluated over 900 different chemical compounds (and other agents, including complex mixtures such as tobacco smoke) for carcinogenicity. *See* International Agency for Research on Cancer, http://www.iarc.fr/IARCPress/index.php (last visited Oct. 11, 2006).

[34] 8 MICHEL DE MONTAIGNE, *Of Democritus and Heraclitus, IN* THE COMPLETE WORKS OF MICHEL DE MONTAIGNE 266, 266-68 (Donald M. Frame Trans.) (2003) New York: Alfred P Knopf.

[35] "It is a widely accepted claim that scientific practice contains value judgments." Wandall, *supra* note 11, at 265. But value judgments are also made in moral philosophy and in applied science areas such as risk assessment and risk management. ". . .[I]n a specific situation (of risk assessment) one must make very specific value judgments about what is 'doing good'. . . ." Kristin Shrader-Frechette, *Reductionist Approaches to Risk*, *in* ACCEPTABLE EVIDENCE: SCIENCE AND VALUES IN RISK MANAGEMENT 243 (D.G. Mayo & R.D. Hollander, eds., 1991).

[36] For a discussion of the relationships between causal judgments and moral judgments (about wrongful behaviors), see J. KNOBE & B. FRASER, CAUSAL JUDGMENT AND MORAL JUDGMENT: TWO EXPERIMENTS (forthcoming).

judgments—of judges are obvious and not necessarily mutually exclusive examples.[37]

By recognizing the universal applicability of judgment are we really any closer to understanding its nature? What exactly is judgment? Is scientific judgment uniquely scientific? Or, is there something called judgment that can be applied equally well to science, to law, and to policy decisions? Finally, is judgment always subjective and so devoid of any claim to objectivity? To shed some light on answers to these questions, it will prove helpful to turn to two very different sources: one theoretical, the other empirical. I begin with judgment in philosophy, followed by empirical studies of expert (scientific) judgment.

I. JUDGMENT IN PHILOSOPHY

Judgment can be found in many philosophical contexts, but those most pertinent to this discussion are: its role in scientific explanation, its role in medical decision making (seen through the reflective lens of Edmund Pellegrino, physician and philosopher), the characterization of judgment as a virtue, and finally, the relationship of judgment to values. This section will offer a briefly describe some of the philosophical contexts which evoke and define the use and purpose of judgment and its connection to different disciplines.

A. Judgment and Scientific Explanation

The aim of science is explanation, typically a causal explanation that arises from the scientific method, or more

---

[37] For a discussion of the biases that can affect the judiciary, see Guthrie, Rachlinski & Wistrich, *Judging by Heuristic: Cognitive Illusions in Judicial Decision Making,* 86 JUDICATURE 44 (2002), who note that "the institutional legitimacy of the judiciary depends upon the quality of the judgments that judges make." Jasanoff points out that in a post-*Daubert* world, judges must not only make legal judgments but also scientific judgments about the relevance and reliability of scientific evidence. Sheila Jasanoff, *Law's Knowledge: Science for Justice in Legal Settings*, 95 AM. J. PUB. HEALTH S49-S58 (2005).

generally, from scientific inquiry.[38] To be useful, judgment must serve this aim. Haack nicely captures the link between judgment and explanation in the context of scientific inquiry:

> They (the scientists) need imagination to think up plausible potential *explanations* of problematic phenomena, to devise ways to get the evidence they need, and to figure out potential sources of error. They need care, skill, and persistence, and intellectual honesty, the moral fibre to resist the temptation to stay out of the way of evidence that might undermine their conjectures; and good *judgment* in assessing the weight of the evidence, unclouded by wishes or fears or hopes of getting tenure or resolving a case quickly or pleasing a patron or mentor or becoming rich and famous.[39]

At first glance, judgment appears to be a distinct mental facility utilized near the end of the inquiry, from which the best scientific explanation of a phenomenon emerges. Science has no monopoly on this approach to inquiry and to the arrangement between judgment and explanation. Anyone who considers herself an "empirical inquirer," including journalists and historians as well as scientists of all stripes, has a similar strategy. A general model of empirical inquiry involves informed conjectures, contrasts of conjecture with the relevant evidence, and the application of judgment to help decide whether any given conjecture should be kept as the best explanation, or

---

[38] The general aim of biomedical science is explanation in terms of causal relationships. Indeed, Lipton argues that all scientific explanations are causal explanations: "To explain a phenomenon is to give an account of its causal history." PETER LIPTON, INFERENCE TO THE BEST EXPLANATION 30 (2nd ed., 2004). Karl Popper notes that ". . . it is the aim of science to find satisfactory explanations or whatever strikes us as being in need of explanation." He also aligns scientific explanation with causal explanation. KARL POPPER, OBJECTIVE KNOWLEDGE: AN EVOLUTIONARY APPROACH 191 (rev. ed., 1981).

[39] *See* SUSAN HAACK, DEFENDING SCIENCE WITHIN REASON: BETWEEN SCIENTISM AND CYNICISM 97 (2003) (emphasis added).

whether it should be dropped or modified.[40] Nevertheless, the scientific inquiry associated with disease causation (as carried out in the biomedical sciences) conforms well to this model, a model rich in conjectures, evidence, explanations, and the various forms of reasoning that can connect these together: inductive, deductive, retroductive, hypothetico-deductive, Bayesian, and "inference to the best explanation," to name a few.[41]

Scientific reasoning is also not enough to accomplish the work of science. Ethical reasoning (whether in the form of principles, rules, or the maxims of case-based 'casuistic' thinking) is also important, especially in studies involving human subjects, where requirements for informed consent, confidentiality, and avoiding harm determine how and even whether studies can be carried out.[42] All science lives in the shadow of scientific misconduct and its menacing instigator, the "fame and fortune viper," whose fangs can be avoided by promoting professional virtues (Haack's "moral fibre") and by recognizing that values can influence the practice of science in powerful ways (Haack's fears, hopes, and wishful thinking).[43]

---

[40] *Id.*

[41] Scientific reasoning is as much about the past as the present and future; within it can be found the background knowledge and experience that a scientist brings to the process of inquiry as well as the methods of research synthesis that summarize and aid in the interpretation of scientific evidence.

[42] For general treatments of the role of ethics in biomedical research—i.e. bioethics—see the classic text on the principles of biomedical ethics, BEAUCHAMP & CHILDRESS, PRINCIPLES OF BIOMEDICAL ETHICS (4th ed., 1994). See also the interesting and somewhat controversial discussion of case-based ethical reasoning—casuistry—in JONSEN & TOULMIN, THE ABUSE OF CASUISTRY: A HISTORY OF MORAL REASONING (1988). For a general discussion of the application of bioethics to a specific health science discipline, namely epidemiology, see COUGHLIN & BEAUCHAMP, ETHICS AND EPIDEMIOLOGY (1996).

[43] There is no quick fix for preventing scientific misconduct. D.L. Weed, *Preventing Scientific Misconduct*, 88 AM. J. PUB. HEALTH, 125, 125-29 (1998). It can be a huge burden on the prestige and public perception of any science, although it has a particularly devastating effect on biomedical science, where individual patients' lives and, more generally, the public's

What is not yet clear, however, is the extent to which judgment plays a role in all these components of scientific inquiry or, alternatively, whether it can be separated out, to be brought into the process near the end. On the one hand, it seems reasonable for judgment to be involved in formulating hypotheses and analyzing data from the start, long before a decision regarding causation is attempted.[44] For is it not the case that a new hypothesis might emerge from a careful consideration—a judgment—of the lack of evidence?[45] There are also the value judgments and moral judgments that affect the practice of science all along its journey from one explanation or paradigm to another.[46] On the other hand, it is also reasonable to place judgment near the end of a causal inquiry as a separate and distinct facility applied to the heavily metaphorical notion of the "weight of evidence."[47] For is it not also the case that once reasoning has been applied to the available evidence, values identified and biases avoided, the expert then puts it all together

---

health lies in the balance. *See* Jerome P. Kassirer, *The Frustrations of Scientific Misconduct*, 328 New Eng. J. Med., 1634, 1634-36 (1993).

[44] *See supra* note 1 for a further discussion.

[45] The philosopher Michael Polanyi is probably best known for his comprehensive look at the influential role of the individual scientist in the process of science. ". . . [I]nto every act of knowing there enters a tacit and passionate contribution of the person knowing what is known. This personal coefficient is no mere imperfection but a necessary component of all knowledge." MICHAEL POLANYI, PERSONAL KNOWLEDGE: TOWARDS A POST-CRITICAL PHILOSOPHY 312 (1958). *See also* Karori Mbugua, *Michael Polanyi and the Personal Element in Science,* 17 SOUTH AFR. J. PHIL. 152, 152-60 (1998).

[46] *See* Wandall, *supra* note 11, at 265.

[47] "Weight of evidence" is almost as popular a concept in the biomedical science literature as "judgment." Indeed, they often appear together, especially in the areas of causal inference and risk assessment. "Weight of evidence" is a particularly vague concept that typically appears in solely a metaphorical sense, without pointing to any specific methodology of interpreting evidence, much less a qualitative or quantitative weighting scheme. Douglas L. Weed, *Weight of Evidence: A Review of Concept and Methods*, 25 RISK ANALYSIS 1545, 1545-57 (2005); Sheldon Krimsky, *The Weight of Scientific Evidence in Policy and Law*, 95 AM. J. PUB. HEALTH, S129-36 (2005).

and makes her final judgment? The exact nature of judgment, then, remains elusive.

B. Judgment in Science-Based Clinical Medical Decisions

Pellegrino, the physician-philosopher, appreciates the complex nature of judgment as it is used in the practice of evidence-based medicine. A clinical judgment, in his view, is the decision regarding the best course of action for a patient, a decision that combines science with ethical norms, the virtues and the values of physician and patient alike.[48]

Virtues are particularly important and Pellegrino promotes both the intellectual virtues of excellence, objectivity, and practical wisdom with the moral virtues of benevolence, honesty, and self-effacement. He puts great stock in prudence, the "capstone" virtue, linking the intellectual, truth-bearing, virtues with the moral, good-seeking, virtues. Prudence connects medical science with medical practice and serves both the explanatory aim of science and the beneficent aim of medicine. Virtue, then, supports the professional practice of evidence-based medicine along with the familiar bioethical principles—non-malfeasance, beneficence, respect for persons, and justice—and all that science itself has to offer each clinical judgment.[49]

C. Judgment as Practical Wisdom

Judgment can be narrowly aligned with a specific virtue, one of the many character traits that assist us in doing our work—in this case, our scientific work—well. Some see judgment as practical wisdom, an Aristotelian virtue.[50] Good judgment, in

---

[48] *See* Pellegrino, *supra* note 5, at 106-08.

[49] *See* EDMUND PELLEGRINO & DAVID THOMASMA, THE VIRTUES IN MEDICAL PRACTICE (Oxford University Press 1993).

[50] ". . . [J]udgment (is) essentially the faculty Aristotle described as practical wisdom, which reveals itself over time in individual decision rather than in the enunciation of general principles." THOMAS NAGEL, *The Fragmentation of Value, in* PHILOSOPHICAL ETHICS 50 (T. Beauchamp, ed., McGraw-Hill 1982).

this view, reveals itself in individual decisions in the same way that any virtue—excellence, objectivity, honesty, or prudence—reveals itself in the actions and decisions of those in possession of a good, virtuous, character. Scientific judgment, then, can be considered an epistemic virtue, serving the explanatory aim of science.[51]

A virtue, it should be noted, is more an attribute of character than an attribute of action. Put another way, a virtue is a disposition rather than a skill or a technique. It disposes us—motivates us—to do our work well.[52] As the virtue of practical wisdom, therefore, judgment is an inherent tendency that motivates us to make a good decision about causation given the available evidence. It does not act independently of the other components of such a decision, including other virtues, but in concert with them.

D. Judgment and Values

Up to this point, my characterization of judgment has run along two parallel yet shared tracks. Judgment can be a virtue, plain and not-so-simple, but it can also be considered a multidimensional mental capacity—a process—within which reasoning, virtues, and values operate on scientific evidence. Before sorting out this overlap, and to complete this brief philosophical snapshot of the nature of judgment, I turn to the role of values in scientific inquiry.

Values and value judgments in science emerged from the postmodern, largely historical, view of scientific progress in the

---

[51] SUSAN HAACK, DEFENDING SCIENCE WITH REASON: BETWEEN SCIENTISM AND CYNICISM 167 (Prometheus Books 2003).

[52] *See* ALASDAIR MACINTYRE, AFTER VIRTUE (2d ed. 1984) for a general account of virtues in ethical theory. *See* Pellegrino and Thomasma, *supra* note 5, for the role of virtues in medical practice. In general, "virtues are to be distinguished from skills and techniques. They are dispositions or tendencies and are valued differently. Errors in technique are expected (i.e. the honest mistake); greater reservations arise when judgments suggest a lack of personal integrity." Douglas Weed and Robert McKeown, *Epidemiology and Virtue Ethics*, 27 INT'L J. EPIDEMIOLOGY 344 (1998).

mid-twentieth century,[53] with intellectual roots deep in American pragmatism.[54] They come in two basic varieties: epistemic (or constitutive) values with origins in an objective tradition, and another, sometimes called contextual values, with origins in a subjective or humanistic tradition.[55] The constitutive values are those criteria, for example, that characterize a good scientific theory. These assist scientists in selecting one theory or hypothesis over another. Thomas Kuhn's list of epistemic values includes: accuracy, consistency, scope, fruitfulness, and simplicity.[56] From Popper's philosophy we could add predictability, testability (or refutability), and explanatory power.[57]

The epidemiologists, interestingly, have their own list of discipline-specific constitutive values—so-called causal "criteria" that assist them in distinguishing causal relationships from

---

[53] THOMAS KUHN, THE STRUCTURE OF SCIENTIFIC REVOLUTIONS (University of Chicago Press 3d ed. 1996) (1962), is generally regarded as the most important contemporary source of the notion that values play a critical role in scientific progress.

[54] Two giants in American pragmatism were Charles Peirce and John Dewey. Judgment featured prominently in their respective works. "A judgment is the mental act by which the judger seeks to impress upon himself the truth of a proposition." CHARLES S. PEIRCE, PHILOSOPHICAL WRITINGS OF PEIRCE 103 (Justus Buchler, ed., Dover Publications 1955). "Philosophy is inherently criticism . . . .a criticism of criticisms. Criticism is discriminating judgment . . ." INTELLIGENCE IN THE MODERN WORLD: JOHN DEWEY'S PHILOSOPHY 260 (Joseph Ratner, ed., Random House, Inc. 1939).

[55] *See* HELEN LONGINO, SCIENCE AS SOCIAL KNOWLEDGE: VALUES AND OBJECTIVITY IN SCIENTIFIC INQUIRY (Princeton University Press 1990).

[56] *See* THOMAS KUHN, *Objectivity, Value Judgment, and Theory Choice*, *in* THE ESSENTIAL TENSION: SELECTED STUDIES IN SCIENTIFIC TRADITION AND CHANGE 321 (1979).

[57] Popper's philosophy is often described in terms of his signature concepts of refutability (the methodological equivalent of falsifiability) and testability. He also emphasized how increasing the falsifiability of a hypothesis in turn increases its explanatory power. "A theory is the bolder the greater its content. It is also the riskier: it is the more probable to state that it will be false." POPPER, 53 THE TWO FACES OF COMMON SENSE 32-105 *in* OBJECTIVE KNOWLEDGE (rev. ed. Clarion Press, 1981).

statistical associations.[58] These include: consistency, strength, gradient (dose-response), experimentation, plausibility, coherence, temporality, specificity, and analogy.[59] Investigators using these criteria (or guidelines) in an assessment of evidence have many choices to make: which evidence to select, how to prioritize them, and what rule of inference they will assign to each so-called "criterion" to signify what it will take for the available evidence to satisfy that particular criterion.[60]

For example, in cancer epidemiology, investigators typically emphasize consistency (the extent to which published, peer-reviewed, studies have similar results), strength (the quantitative magnitude of the observed relationship between exposure and disease outcome as measured in a study population), and biologic plausibility (the extent to which the biological mechanism is known).[61] These, in turn, reflect general scientific principles (or maxims) that play an important role in scientific reasoning.[62] Consistency of findings in the face of repeated (and

---

[58] Causal inference has been discussed in epidemiology for decades and is most often described in terms of "criteria" which are also called "considerations" or "guidelines."

[59] The British medical statistician, Austin Bradford Hill, published a now-classic article setting forth nine causal considerations (now most often called "criteria"). Austin Bradford Hill, *The Environment and Disease: Association or Causation?* 58 PROC. ROY. SOC. MED. 295 (1965).

[60] For a discussion of the value-laden nature of Hill's criteria, see Weed, *supra* note 9.

[61] The use of causal criteria in epidemiology typically occurs in published reviews of the scientific literature, textbook chapters, and editorials commenting upon the causal significance of a particularly important (new) study. In a systematic review of a series of these published reviews in cancer epidemiology, several patterns of usage were revealed: the most common criteria used in practice were: consistency, strength, and biologic plausibility. Other criteria were also employed, depending upon the particular approach of the investigators. *See* Weed & Gorelic, *supra* note 4.

[62] The causal criteria of epidemiology used by many public health and medical disciplines were developed in the mid-20th Century without much explicit attention to their theoretical roots. More recently, these criteria have been linked to more general scientific principles, including testability and predictability. *See* CAUSAL INFERENCE 15-32 (Kenneth J. Rothman, ed., 1988); Douglas Weed, *Methods in Epidemiology and Public Health: Does*

more stringent) tests reflects the principle of testability. Strong associations typically mean that alternative explanations—confounders—are less likely. And the existence of a well-evidenced biological mechanism may permit better predictions of future effects.

Aligning constitutive (scientific) values with general scientific precepts and principles such as quantitation, testability, and predictability, suggests that it is incorrect to necessarily align values with subjectivity. Since "epistemic" values play such an important—indeed, key—role in scientific reasoning, it follows that scientific judgment, inasmuch as it is part reasoning and part values, has some claim to objectivity. In other words, calling expert judgment "subjective," without qualification, is misleading at best, inappropriate at worst.[63]

Contextual or subjective values also play an important role in scientific inquiry. These arise from many sources: psychology, social and cultural forces, politics, and ethics. They affect many parts of the scientific endeavor, including but not limited to the choice of research topic as well as the interpretation of evidence.[64] An interesting example can be found in epidemiological literature, in which an academic scientist (with strong but undisclosed anti-abortion views) published a review of the scientific evidence on the topic of the potential relationship between induced abortion and breast cancer, concluding that the causal association was responsible for thousands of breast cancer deaths each year and characterizing it as a public health tragedy.[65] This scientist later admitted in a published newspaper

Practice Match Theory?, 55 J. OF EPIDEMIOLOGY AND COMMUNITY HEALTH 104-10 (2001).

[63] *See* Y. Arimone, B. Begaud, G. Miremont-Salame, A. Fourrier-Reglat, M. Molimard, N. Moore & F. Haramburu, *A New Method for Assessing Drug Causation Provided Agreement with Experts' Judgment*, 59 J. OF CLINICAL EPIDEMIOLOGY 308-14 (2006).

[64] *See* Wandall, *supra* note 11, at 267.

[65] J. Brind, V.M. Chinchilli & W.B. Severs, et. al., *Induced Abortion as an Independent Risk Factor for Breast Cancer: A Comprehensive Review and Meta-Analysis*, 50 J. OF EPIDEMIOLOGY AND COMMUNITY HEALTH 481-96 (1996).

interview that his conclusions had been colored by his moral values.[66] It should be noted that other epidemiologists, examining the same evidence at the same time and using the same criteria of causation, concluded that not only was there no causal association, but more significantly, that there was no evidence of even a statistical association between induced abortion and breast cancer.[67]

Values are vital and powerful players in scientific inquiry, with influence at both theoretical and practical levels. Some say that in the actual practice of making an expert (scientific) judgment of the available (weight of the) evidence, these objective and subjective value-laden traditions cannot be separated.[68] In the end, an expert's judgment is always at least one part fact and two parts value.

E. Four Types of Judgment

Perhaps not surprisingly, this brief philosophical overview has both complicated matters and provided insights into the nature of judgment. A relatively simple model—that sound and unbiased judgment, applied to the scientific evidence, results in a causal assessment or claim—is probably insufficient. Either way it is vague, because it does not explicitly recognize that judgment can include scientific and ethical reasoning as well as values and virtues, or it is too narrow, because it assigns judgment to one of the virtues and ignores the role of reasoning, other virtues, and values in evidentiary judgments.

Several different senses of the concept of judgment have been identified. It can be an outcome, as in the final causal judgment (or decision or claim) made by an individual scientist or group of scientists. Similarly, a physician's decision to intervene on behalf of her patient is a clinical judgment. But judgment can also be the multidimensional mental capacity that

---

[66] R.A. Knox, *Bias in Abortion Study is Charged*, BOSTON GLOBE, Oct. 12, 1996, at A1.

[67] *See* Weed, *supra* note 10, at 111.

[68] *See* OTWAY & VON WINTERFELDT, *supra* note 3, at 84.

incorporates forms of reasoning, various values, and several virtues, and with which one produces an outcome—a judgment—when confronted with evidence. Thirdly, judgment can be more narrowly framed as just one of the dimensions of this broader capacity: given the evidence, the scientist reasons, considers (when she can) the values that may guide or influence her conclusion, and then employs her judgment—that reflects her character—along the way. This third sense of judgment tracks most closely with the virtue of practical wisdom.

There may also be room for one more type of judgment, which is X: as an irreducible, if not independent, component of the process not captured by reasoning, values nor virtues, and yet without which the final product—a decision regarding the existence (or not) of a well-evidenced causal explanation—cannot be completed. Perhaps this is just intuition or perhaps it is just common sense, although we should never forget Thoreau's warning about this particularly ill-defined human capacity: "The commonest sense," he wrote, "is the sense of men asleep, which they express by snoring."[69]

Separating out this irreducible and rather vague—"whatever remains"—version of judgment may be difficult. It may also be difficult to separate out any of the components of judgment in practice inasmuch as each intimately influences the other and likely operates simultaneously with all others. Practical wisdom, for example, helps us to reason well. Similarly, judgment (as the virtue of practical wisdom) is not only applied at or near the end of the process, but also during the initial phases of scientific inquiry when selecting which hypothesis to study, a decision that is also affected by scientific reasoning (which puzzles remain to be solved) and social values (which studies funding agencies will support given their respective research priorities). Practical wisdom will also come in handy, along with scientific reasoning, and the virtues of excellence, and objectivity, when deciding (i.e. judging) which studies are relevant, their validity, and their relative weights in an interpretative assessment of causation or risk. Finally, judgment—as a multidimensional capacity to make

[69] H.D. THOREAU, WALDEN 287 (Alfred P. Knopf 1992).

a reliable decision in the face of evidence—aids in selecting which causal criteria (or other epistemic values) to employ or, alternatively, how to prioritize them, given the available evidence, a very good example of a type of a judgment that is part fact, part value; part objective, part subjective; part quantitative, part qualitative.

We can conclude that a scientist's judgment is applied to the evidence, both after and in parallel with the complex firing of all the neurons and mental pathways needed to reason scientifically, think ethically, and act in accordance with one's character (i.e., with the virtues) and values, whether those values come from within or from outside the scientific community. Given all these factors, it is a complex mix indeed that characterizes an expert's scientific judgment.

II. EMPIRICAL STUDIES OF EXPERT JUDGMENT

Those who study expert judgment, recognizing the difficulty of pinning down a precise definition and appreciating its complex multidimensionality, have primarily examined the extent to which a carefully selected and small group of experts, when shown exactly the same information (i.e., the scientific evidence) and having been interviewed (and coached) by trained elicitators,[70] can arrive at similar conclusions.[71] These studies do

[70] "Elicitators" are those trained in the study methodologies of eliciting (guiding) judgments. Facilitators would be a reasonable synonym.

[71] A systematic review of the state of the science of studying expert judgment is beyond the scope of this paper. Walker et al. include a brief history of this topic in their introduction to a study that elicited what they call "subjective judgment" regarding personal exposure to benzene in the absence of adequate amounts of relevant data. Katherine Walker et al., *supra* note 29, at 308-22. Note that the outcome of interest in this study is the extent to which a group of seven scientific experts can agree on the concentration of benzene individuals in a target population (being study separately) were exposed to at home, indoors, and in their ambient environment. These experts "exhibited striking differences in the degree of uncertainty expressed" despite being reasonably close in agreement about the mean (average) benzene exposure concentrations. *Id.* at 308. In another study by Evans et al., the investigators elicited statements regarding causality (carcinogenicity)

not examine the independent role of judgment so much as the end result of a process that involves the scientific evidence, whatever reasoning process and values and biases each individual expert brings to the table, and the interventions of the investigators, all of which may have some impact on whatever final judgments are made. In some studies, the experts' individual judgments agree enough to be aggregated. In other studies, there are irresolvable differences. Both outcomes reflect what has been observed in another type of study—more descriptive—examining how experts in the everyday practice of causal inference have judged causation, given epidemiologic and toxicological evidence. In the practice of public health, for example, some reviewers of a body of epidemiologic and toxicologic evidence may claim causation, while others will not. Similarly, experts invited by the courts to examine scientific evidence can come to vastly different conclusions regarding general causation.

In *Soldo v. Sandoz Pharmaceuticals*,[72] for example, differing claims of causation regarding a medication's capacity to cause stroke were partially explained by the fact that different scientific disciplines were involved. A neurologist and an epidemiologist were unconvinced about general causation; the pharmacologist, on the other hand, disagreed and made the causal claim.[73]

---

of chloroform in drinking water, given the available evidence, plus a formal (protocol-driven) elicitation process. In the words of the study authors, "risk distributions varied considerably between experts," ranging from no risk to risks considered to be of regulatory significance. John Evans et al., *Use of Probabilistic Expert Judgment in Uncertainty Analysis of Carcinogenic Potency*, 20 REG. TOX. AND PHARMACOL 15, 15-36 (1994). In an elicitation study of the judgment of five lung cancer experts, on the other hand, all five agreed that cigarette smoking, residential radon, and environmental tobacco smoke were the major causes of lung cancer and that between 80 and 95 percent of lung cancer deaths were due to cigarette smoking. Casman & Morgan, *supra* note 32, at 5914.

[72] Soldo v. Sandoz Pharmaceuticals Corp., 244 F. Supp. 2d 434 (W.D. Pa., 2003).

[73] *See* Joe Cecil, *Ten Years of Judicial Gatekeeping Under* Daubert, 95 AM. J. PUB. HEALTH S74, S76-S78 (2005).

We can conclude, then, that in both concocted and real world scenarios, when experts are faced with exactly the same evidence, they may or may not disagree on its causal interpretation.

A. Disagreements about Causality

Disagreement may not always be a problem for science because it can pave the way for scientific progress.[74] Reasonable error-seeking resolution of scientific disagreements can stimulate progress at the "micro" level (when disagreements with the conclusions of a study give rise to a belief that a different, presumably better, study is warranted), and at the "meta" level (when disagreement with an assessment of the causal significance of a body of evidence gives rise to a belief that a new research program is needed to reveal the underlying mechanism). At the "macro" (paradigmatic) level, however, where the proponents of a new—even, revolutionary—theory do battle with those protecting the status quo, more research may not be the key to resolution. Paradigmatic shifts are more like religious conversions than rational exercises.[75] Nevertheless, at the micro and meta levels, the resolution of the disagreement involves the familiar refrain, "more research is needed." Scientific disagreement that is reasonable, in other words, can be resolved—or, more precisely, potentially resolved—with better studies, including improvements in the methods for interpreting those studies, the so-called methods of research synthesis.[76]

---

[74] Haack, personal communication, July 5, 2006. The term "disagreement" as used here means the reasonable kind, where both sides of an issue employ "sound" judgment (not yet well-characterized) yet still cannot agree on the interpretation of the available evidence. The term "progress" means the replacement of one scientific problem-solution with another, typically achieved with more research designed to eliminate errors in earlier studies. *See* POPPER, *supra* note 13.

[75] *See* KUHN, *supra* note 12.

[76] *See* D. Weed, *Evidence Synthesis and General Causation: Key Methods and an Assessment of Reliability*, 54 Drake L. Rev. 639, 639-50

More research, of course, does not necessarily lead to less controversy. If new studies firm up the evidence on both sides of a disagreement, the controversy continues. Indeed, this phenomenon has been frequently observed for decades-long runs of epidemiologic studies, and it has been cited by the media (and occasionally by scientists) as strong evidence of a certain weakness of this inherently observational yet essential biomedical science.[77]

B. Does More Research Mean Less Judgment?

It is commonly said that in the absence of sufficient scientific evidence, we must rely on the scientist's (subjective) judgment.[78] But does it follow that as more evidence accumulates, there is less judgment employed? The answer to this question depends on the type of judgment under consideration.

Recall the four senses of judgment described earlier,[79] to which subscripts have been added for ease of presentation: *judgment*$_0$, an outcome and the final result of *judgment*$_1$; *judgment*$_1$, the multidimensional mental capacity within which scientific and ethical reasoning operate along with the virtues and values; *judgment*$_2$, the virtue of practical wisdom (and so a component of *judgment*$_1$); and *judgment*$_3$, a sense of the concept not captured by the others; intuitive, almost mystical, and commonly applied to the evidence at the end of the overall process of inquiry.

As more evidence accumulates, the complexity of the final assessment regarding causation can only increase, requiring more *judgment*$_1$ (and so, *judgment*$_2$). Put another way, both these forms of judgment are needed at all steps along the journey undertaken in a systematic assessment of the available evidence

(2006).

[77] *See supra* note 21 for a further discussion.

[78] For a recent example, see Walker et al., *supra* note 29, at 308. "Frequently, however, data are neither abundant nor directly relevant, making it necessary to rely to varying degrees on subjective judgment." *Id.*

[79] For a further discussion s*ee supra* Section E.

and so must increase along with the increase in the amount, kinds, and varying characteristics of the accumulating evidence, regardless of the results of that evidence. Only *judgment*$_3$ could possibly shrink away as the evidence accumulates inasmuch as it is compatible with a rather naïve view of the relationship between evidence and judgment wherein a causal conclusion can, with enough strong and unequivocal evidence, be determined by that evidence, without the need for any "judgment" at all, as if to say that the evidence "proved" the hypothesis. Yet despite the popularity of the concept of "proof" among scientists, especially at conferences where randomized clinical trial results are presented, it is widely accepted that science has no such claim on determinism and that evidence always underdetermines the hypotheses being tested.[80] We can presume, therefore, that the mystical, intuitive sense of judgment (*judgment*$_3$) like its cousins, *judgment*$_1$ and *judgment*$_2$, will always be required in any causal interpretation of evidence. Perhaps it will fade as evidence accumulates, although there is no way to measure it.

C. Disagreements about Causality in the Courts and in the Regulatory Environment

Disagreement about causation may be good for science, but it is a serious concern in the courts, due to the adversarial nature of the legal process, with (ideally) equally "expert" experts representing defendants' and plaintiffs' claims about the reliability and relevance of the scientific evidence as well as causal interpretation. Precisely the same situation exists within the regulatory context, where typically there are also two contrary views: one favoring looser standards and the other favoring stricter standards on some potentially hazardous material.

There is no easy solution to this problem. There is no a priori reason why we should necessarily expect any two experts

---

[80] *See* E. McMullin, *Underdetermination*, 20 J. MED. PHILOS. 233, 233-52 (1995). *See also, supra* note 8.

to agree on the overall causal interpretation—having used their judgment—of the available evidence. The sheer complexity of causal assessments is one concern, reflected in the many theoretical, methodological, and practical choices available to causal decision makers. Another concern is the influence of virtues and values. Finally, we cannot forget the very real effect of various perspectives on causation that corresponds to membership in different scientific disciplines.

The problem of disagreement is such a great concern for judges (especially, but not exclusively, in toxic tort litigation) and for policymakers (including public health and medical professionals) because they are charged with, and are therefore responsible for, decisions about what should be done "today"—that is, without the luxury of waiting for additional scientific evidence. Something must be decided, at *this Daubert* hearing or at *this* clinic visit, at *this* regulatory hearing, or in the face of *this* outbreak of disease in *this* community.[81] When faced with scientific disagreement, the fact finders in court or decision makers in other contexts must either interject their own version of the scientific state of affairs—like a tie-breaker—or they must find a way to determine that one or the other side of the disagreement is incorrect, or more likely so. In any case, this situation must involve an assessment of the soundness of others'—the experts'—judgment.

III. SOUND JUDGMENT

If there were rules (or algorithms) for causal decisions, then an answer to the question, "What is sound judgment?" would be straightforward and involve following the rules. Unfortunately, no such algorithm exists and, to make matters more difficult, it

---

[81] A central message of Daubert v. Merrell Dow Pharmaceuticals, Inc., 516 U.S. 869 (1993) was the Supreme Court's designation of the trial judge as the "evidentiary gatekeeper," responsible for screening expert testimony to determine whether the relevancy and reliability requirements for scientific evidence are met. *See* NATIONAL RESEARCH COUNCIL, DISCUSSION OF THE COMMITTEE ON DAUBERT STANDARDS: SUMMARY OF MEETINGS (National Academies Press 2006).

can be inferred from our earlier discussion that judgment (whether *judgment*$_1$ or *judgment*$_2$ or both) is important even in those situations in which algorithms (say, for statistical analysis) produce quantitative results; conclusions about causality do not leap out from the numbers, judgment is always required.

So what constitutes "sound" judgment and can we distinguish it from "good" judgment and from "unbiased" judgment? As noted earlier, we generally prefer good, sound, and unbiased judgment in matters scientific.

To say that someone has "good" judgment suggests that in both the short and long run, their decisions have proven to be well-regarded or, more often "right" than not, or some other such evaluative assessment that focuses on the outcome more so than the process through which the decision or outcome was produced.[82] "Sound" judgment, in contrast, suggests that the process itself—the mental process that produced the outcome—was exemplary in some way. "Good" judgment, therefore, primarily refers to *judgment*$_0$, whereas "sound" judgment refers to *judgment*$_1$.

But it would probably be a mistake to suggest that "goodness" and "soundness" are not overlapping in meaning and importance. In these matters, outcome and process always go hand-in-hand. And so it seems reasonable to suggest that good outcomes could conceivably come from unsound judgment (perhaps because the process is infected with unsound reasoning, e.g., frank errors in scientific reasoning) and vice versa: that sound judgment could, in some instances, produce bad outcomes, given the complexity and unpredictability of real-life matters.

How bias relates to the soundness and goodness of judgment is also important. An unbiased judgment is not necessarily sound; there are many other components, such as scientific

---

[82] Section III is limited to judgment as outcome (*judgment*$_0$) and to the multidimensional mental capacity (*judgment*$_1$) within which the virtue of practical wisdom resides (i.e., *judgment*$_2$). Section III does not address the problem of trying to determine if common sense or intuition (i.e., *judgment*$_3$) can ever be unsound in the absence of a pathologic condition.

reasoning and ethical reasoning that could have gone astray. Likewise, a sound judgment may still be biased. There are many different forms of bias that cannot be easily identified, much less controlled. Bias, it seems, is unlikely to ever be completely extracted from (or missing in) the mental process of scientific judgment. After all, we are human beings, not reasoning machines.

We may conclude that both good judgment and sound judgment are revealed in outcomes and reasoning processes that are relatively unbiased and that are rarely in error (that is, outcomes that are rarely made with frank errors in reasoning and are acceptable to others in the long run). Both good and sound judgment involves the open admission of values which play a role in the ultimate outcome.

Who is qualified to make such an assessment; that is, who can judge the outcomes, and the judgment of experts, judges, and whoever else uses judgment in their professional practices? The answer is not immediately apparent. Nevertheless, it seems reasonable to think that some people's judgment is better than others.[83]

CONCLUSION

This brief excursion into the nature of judgment has reaffirmed its necessity in causal thinking and has generated more questions than answers. At the top of the list is the problem of assessment: assessing the soundness of scientific judgment as a mental capacity or process ($judgment_1$) as well as assessing the goodness of a judgment as outcome ($judgment_0$). The latter is well known in the biomedical sciences; it is the problem of causation typically described in terms of the quantity, types, and characteristics of evidence needed to make a claim of disease causation. Perhaps it would be wise to add the various dimensions of judgment as mental capacity or process, for the two forms of judgment are intimately linked. The problem of assessing the soundness of judgment—how we think

---

[83] Haack, personal communication, July 5, 2006.

about causation—also deserves much more attention from the philosophers, psychologists, and cognitive scientists who study such matters, and from those who, in their professional practices, are called upon to employ judgment in our courts and clinics, and in the area of public health.

The many dimensions and components of scientific judgment give us some guidance on how it can be improved: by practicing our reasoning skills, building professional character, and by dutifully identifying the values that influence, even bias, our decisions and actions. Avoiding bias is always a good idea and always difficult to achieve. We must keep careful track not only of our causal decisions but also of the processes that result in those decisions.

In the end, we find ourselves close to where we began: at the intersection of science, law, and policy, where so much complexity arises, where so many decisions of vital importance must be made, and where judgment matters most.