Nobel prizes would have flunked *Benzene*: judicial review of administrative evidence overlooks science's linguistic tradition

*By Jimmy J. Zhuang, Ph.D.**

The administrative law literature has frequently questioned the wisdom and efficacy of generalist courts reviewing specialist agency decisionmakers. This concern has been ballooning as more and more agencies regulate activities at the frontiers of science. The courts have generally followed the substantial evidence standard enunciated in the Benzene case when reviewing scientific matters within the administrative record. In this Essay, I argue that the Benzene court inappropriately emphasizes the language of the scientific evidence when determining whether the agency has met the requisite certainty. Post-Benzene courts mandate not only a high degree of scientific certainty for decisions, but also require a high-confidence linguistic style for composing the supporting evidence. This latter requirement contradicts the writing style of the scientific community. Scientific literature uses a restrained and tentative linguistic style that the Benzene Court is unaware of, but is deeply suspicious of. Therefore, this Essay argues that close-reading of scientific literature's language is an incorrect tool for determining scientific evidence's certainty because it overlooks science's linguistic tradition.

I. INTRODUCTION

We live in an administrative state. The technical complexities of modern governance require expertise that broad stroke legislatures do not

possess.[1] The constitutional safeguard against this ever-growing fourth branch is supposedly the expansive and presumptive availability of judicial review.[2] However, the administrative law literature has frequently questioned the wisdom and efficacy of generalist courts reviewing specialist agency decisionmakers.[3] This concern has been ballooning as more and more agencies regulate activities at the frontiers of science.[4] The courts have thus sought a unifying standard for reviewing scientific matters within the administrative record. The seminal case of *Industrial Union Department v. American Petroleum Institute*, known as the *Benzene* case,[5] established the canonical framework for the substantial evidence standard in reviewing decisions that rely on scientific evidence.

In this Essay, I outline the *Benzene* framework and focus on the Court's discussion of the scientific certainty that it requires. Using a particularly troubling style of textualism, the *Benzene* Court seems to

---

* J.D., 2016, Yale Law School. Ph.D. biochemistry, 2013, Harvard University. A.B. molecular biology & applied mathematics, 2008, Dartmouth College. I would like to thank Professor Jerry Mashaw for discussion and suggestions for this Essay.

[1] *See generally* Edward L. Rubin, *Law and Legislation in the Administrative State*, 89 COLUM. L. REV. 369 (1989) (arguing that today's legislatures are limited to duties of delegation and funding, with all other lawmaking handled by administrative agencies).

[2] Raoul Berger, *Administrative Arbitrariness and Judicial Review*, 65 COLUM. L. REV. 55, 93 n. 207 (1965) ("the availability of judicial review is by far the most significant safeguard against administrative excess . . . . [An] important source of the efficacy of judicial review is its preventive value."); *see also* Peter L. Strauss, *Revisiting* Overton Park*: Political and Judicial Controls over Administrative Actions Affecting the Community*, 39 UCLA L. REV. 1251, 1263 (1992) (discussing the presumptive availability of judicial review after *Overton Park*); Louis L. Jaffe, *Judicial Review: "Substantial Evidence on the Whole Record"*, 64 HARV. L. REV. 1233, 1238 (1951) (discussing the broad scope of judicial review for substantial evidence after *Universal Camera*).

[3] *See, e.g.*, Kenneth W. Starr, *Judicial Review in the Post*-Chevron *Era*, 3 YALE J. REG. 283, 309–10 (1986) ("Technical expertise . . . may greatly aid the interpretation of regulatory statutes, especially where an understanding of congressional intent requires familiarity with technical issues . . . . Although technical statutes are, of course, produced by generalist legislators[,] . . . the agency inevitably enjoys an edge in understanding technical concepts and terminology contained in the statute or its legislative history. The agency is also more familiar with the regulated industry. These advantages . . . are all the more evident during an era of burgeoning judicial caseloads . . . ."); Peter L. Strauss, *One Hundred Fifty Cases per Year: Some Implications of the Supreme Court's Limited Resources for Judicial Review of Agency Action*, 87 COLUM. L. REV. 1093, 1127 (1987) ("[In] an area generally characterized by statutes of substantially greater complexity and technical detail than those of an earlier generation[,] . . . a judge's limited resources, his only occasional opportunities to seek understanding, and the often distorting character of the litigation perspective . . . can lead him to fear that his decision will be more disruptive than helpful to the statutory scheme.").

[4] *See, e.g.*, Industrial Union Dep't, AFL-CIO v. Hodgson, 499 F.2d 467, 474 (D.C. Cir. 1976) ("[S]ome of the questions . . . are on the frontiers of scientific knowledge, and consequently . . . [d]ecision making must in that circumstance depend to a greater extent upon policy judgments and less upon purely factual analysis.").

[5] 448 U.S. 607 (1980) [hereinafter *Benzene*].

emphasize the language of the scientific evidence within the administrative record when determining whether the agency has met the requisite certainty. Therefore, the *Benzene* Court not only mandates a high degree of scientific certainty for decisions, but also requires a high-confidence linguistic style for composing the supporting evidence. This latter requirement is in direct contradiction with the writing style of the scientific community.

To show this important oversight, I excerpt capstone conclusions from papers that won the Nobel Prize in chemistry during the four years after the *Benzene* case as well as the most recent four years. They show that scientific literature uses a restrained and tentative linguistic style that the *Benzene* Court is unaware of. One historic reason for this subdued linguistic style is that science is thought to be continuously progressing. While today's scientific writing still follows this tradition, the language does not actually reflect a lack of certainty; in fact, scientists have readily-available and extremely powerful statistical tools to precisely determine certainty. Thus, although the theoretical reasoning for the *Benzene* Court's certainty requirement seems correct – legal determinations and their errors carry great risk, which means they mandate greater certainty – close-reading scientific literature remains an inappropriate tool for meeting this stringency because it overlooks science's linguistic tradition.

II. THE *BENZENE* FRAMEWORK

In reviewing agency determinations based on scientific evidence, the *Benzene* court laid down the now canonical precept that

> [although an agency] is not required to support its finding . . . with anything approaching scientific certainty[,] . . . the Agency's findings must be supported by substantial evidence . . . . [W]here its findings must be made on the frontiers of scientific knowledge . . . , so long as they are supported by a body of reputable scientific thought, the Agency is free to use conservative assumptions in interpreting the data.[6]

To operationalize this standard, the Court specifically explains that the substantial evidence must demonstrate "***at least more likely than not***" that the agency's position is correct.[7] On its face, this more-than-likely standard for evaluating scientific evidence seems perfectly reasonable, as it approximates the civil standard of preponderance of evidence. Moreover, in the *Benzene* case itself, this standard seems to have been correctly applied: OSHA failed to provide evidence that there was *any*

---

[6] *Id.* at 656.
[7] *Id.* at 653 (emphasis added).

significant workplace risk for benzene exposure.[8] Thus, to rule on the case, the Court did not need to reach the question of what specific benzene exposure levels are acceptable. However, the Court goes on to evaluate OSHA's evidence anyways, in order to provide guidance on what is considered sufficient substantial evidence. Before discussing these guidelines, it is worth noting that the Court emphasizes throughout the opinion that it does not require absolute scientific certainty to meet the substantial evidence standard.[9]

The positive guidance the *Benzene* Court provides for what meets the substantial evidence standard is highly textualist. In its conclusion, the Court states that "[t]he closest [OSHA] came to making a finding that benzene presented a significant risk of harm in the workplace was its statement that the benefits to be derived . . . were ***'likely'*** to be ***'appreciable."***[10] In the Court's opinion, the evidence written as able to "likely" provide "appreciable" support is insufficient for the more-than-likely standard. The quotation marks surrounding "likely" and "appreciable" indicate that the Court is attentive to the specific language used in the scientific evidence within the administrative record. The implication is that this evidence is insufficient because it was written tentatively, using qualifiers such as "likely." Thus, the Court is assuming that the scientists composing this evidence contemplated the plain-meaning scienter of every word like a lawyerly scrivener would. A subdued writing style, in the eyes of the Court, reflected a lower certainty of the underlying evidence.

While this linguistic focus may have been unintentional, many courts since *Benzene* have adopted this approach in analyzing the language of the evidence within the administrative record. For example, in *Massachusetts v. EPA*, the Court becomes satisfied with the presented evidence when the expert uses the phrase "strong consensus" to describe the views of climate scientists during his testimony.[11] In another example, a California appeals court refused to accept evidence of benzene's harms because experts could only describe its causal link to leukemia as "probable."[12] These and many other courts since *Benzene* have focally evaluated the language of the scientific evidence within the administrative record, with special attention to qualifiers, as a means to gauge the evidence's certainty. This is very troubling because scientists write tentatively by tradition.

---

[8] *Id.* at 614–15.

[9] *See, e.g.*, *id.* at 656 ("OSHA is not required to support its finding that a significant risk exists with anything approaching scientific certainty.").

[10] *Id.* at 653 (emphasis added).

[11] 549 U.S. 497, 521 (2007).

[12] Santiago v. Firestone Tire & Rubber Co., 224 Cal. App. 3d 1318, 1328 (1990).

III. THE CHEMISTRY BETWEEN SCIENTISTS AND COURTS

Winning the Nobel Prize is the ultimate recognition of scientific merit. To demonstrate the tentative nature of scientific writing, I have excerpted scientific papers that won the Nobel Prize in chemistry for the four years after *Benzene* was decided and for the four most recent years. It is evident that even these authoritatively rigorous findings are written in a restrained and understated style.

Professor Frederick Sanger won the 1980 chemistry Nobel Prize,[13] months after the *Benzene* case was decided. He described his key experiment as follows: "The sequences can be read with ***reasonable accuracy*** starting at 88 nucleotides . . . for about 80 nucleotides . . . . For the next 50 nucleotides there is ***some uncertainty*** in the number of nucleotides in 'runs' because bands are not actually resolved."[14]

Professor Roald Hoffman won the 1981 chemistry Nobel Prize.[15] He concluded his seminal paper by stating: "It is now clear that in the course of a chemical reaction the nuclei of the reaction components ***may*** undergo some extremely complicated motions in order to provide low-energy electronic paths."[16]

Professor Aaron Klug won the 1982 chemistry Nobel Prize.[17] He summarized his groundbreaking research as: "For units that can be assembled into an icosahedral shell, these 'bonds' must be arranged at . . . fixed angles . . . . The building unit is, therefore, ***likely*** to be a hexamer, which ***tends*** *to* aggregate in plane sheets, but this hexamer ***could also be*** bonded . . . at a vertex or edge of the icosahedron."[18]

Professor Henry Taube won the 1983 chemistry Nobel Prize.[19] He summarized his key findings as follows: "If the data on affinities indicate anything about the type of binding, they ***suggest*** that it is remarkably similar for the four cations considered, aluminum ion ***perhaps*** being the most 'ionic'."[20]

---

[13] *All Nobel Prizes in Chemistry*, NOBELPRIZE.ORG, http://www.nobelprize.org/nobel_prizes/chemistry/laureates (last accessed Jan. 7, 2016).

[14] Frederick Sanger, S. Nicklen & Alan R. Coulson, *DNA sequencing with Chain-Terminating Inhibitors*, 74 PROC. NAT'L ACAD. SCI. USA 5463, 5466 (1977) (emphasis added).

[15] *See supra* note 14.

[16] Roald Hoffman & Robert Burns Woodward, *Orbital Symmetry Control of Chemical Reactions*, 167 SCIENCE 825, 825 (1970) (emphasis added).

[17] *See supra* note 14.

[18] Donald L. D. Caspar & Aaron Klug, *Physical Principles in the Construction of Regular Viruses*, 27 COLD SPRING HARBOR SYMP. QUANT. BIOL. 1, 19 (1962) (emphasis added).

[19] *See supra* note 14.

[20] Henry Taube, *Rates and Mechanisms of Substitution in Inorganic Complexes in Solution*, 50 CHEM. REV. 69, 88 (1952) (emphasis added).

This style of scientific writing is as prevalent today as it was in the 1980s. Professor Robert Lefkowitz won the 2012 chemistry Nobel Prize.[21] He crystallized his core findings by stating: "Our data ***suggest*** that P-arrestin is required to effect inhibition of receptor function subsequent to phosphorylation by βARK."[22]

Professor Arieh Warshel won the 2013 chemistry Nobel Prize.[23] He concluded his seminal paper as follows: "[I]t ***seems*** that a ***promising*** way to simulate the surface crossing in rhodopsin ***might*** be achieved by a chemical approach of building a model compound with a Schiff base bridge between the ring to the end of the chain."[24]

Professor Eric Betzig won the 2014 chemistry Nobel Prize.[25] He concluded his seminal paper by remarking that "the good fit of the data to the model for each molecule . . . ***suggests*** that the apparent absorption cross section and quantum yield are quite uniform . . . . Finally, . . . it is ***possible*** to use these same methods . . . to determine the emission dipole orientations as well as the absorption dipole directions shown here."[26]

And most recently, Professor Paul Modrich won the 2015 chemistry Nobel Prize.[27] He concluded his seminal paper by stating: "[T]hese findings ***suggest*** that mutS-DNA interaction involves an asymmetry inherent to the four mispairs tested or, ***alternatively***, an asymmetry imposed by this set of mispairs on the DNA helix."[28]

All of these excerpts reflect a common theme in scientific writing: they are littered with qualifiers. The sentence structure "[our data] or [this] may suggest perhaps [that]" is probably the most commonly-used phrase in all of science. A Google Scholar search for the three-word string "may suggest perhaps" will confirm this. In fact, one of the greatest scientific discoveries ever was written as such an understatement: "[What] we have postulated . . . ***suggests*** a ***possible*** copying mechanism for . . . genetic

---

[21] *See supra* note 14.

[22] Martin J. Lohse, Jeffrey L. Benovic, Juan Codina, Marc G. Caron & Robert J. Lefkowitz, *β-Arrestin: a Protein That Regulates β-Adrenergic Receptor Function*, 248 SCIENCE 1547, 1549 (1990) (emphasis added).

[23] *See supra* note 14.

[24] Arieh Warshel, *Bicycle-pedal Model for the First Step in the Vision Process*, 260 NATURE 679, 683 (1976) (emphasis added).

[25] *See supra* note 14.

[26] Eric Betzig & Robert J. Chichester, *Single Molecules Observed by Near-Field Scanning Optical Microscopy*, 262 SCIENCE 1422, 1424 (1993) (emphasis added).

[27] *See supra* note 14.

[28] Shin-San & Paul Modrich, Escherichia coli mutS-*Encoded Protein Binds to Mismatched DNA Base Pairs*, 83 PROC. NAT'L ACAD. SCI. USA, 5057, 5060–61 (1986) (emphasis added).

material."[29] Therefore, the predominant writing style in science is to be tentative.

Science likely first developed this understated linguistic style because it is an endlessly progressive endeavor. There is always the possibility that new discoveries trump old facts. From this ancient mindset, scientists have passed on this written tradition of humility. However, over the years, this writing style has become more and more a pro forma tradition, and less and less a reflection of any actual scienter of hesitation about the reported findings. With the development of strong statistical tools, scientists can now describe exactly how certain they are. Today, in any reputable scientific paper, universal publication standards require a *p* value of less than 0.05 before a scientific finding can be reported; roughly, this means almost all published scientific findings are 20 times more likely to be true than not.[30] This is far greater precision and certainty than *Benzene*'s "at least more likely than not" standard. And yet, despite such statistical certainties, all published scientist writings still add on phrases such as "our findings may suggest perhaps that" to describe their results, in order to pay homage to the traditional scientific writing style. Therefore, within a modern scientific paper, authors often present precise statistical certainties in the raw data figures,[31] but still heed the revered tradition of writing with qualifiers in the paper's text.[32] In other words, scientific writing reads like it is hesitant about its conclusions when in fact it is not. It is indeed a misleading linguistic tradition that scientists use powerful statistical tools and present precise certainty measurements within their papers and yet still use a qualifier-rich writing style. This means that when construing scientific writing, words like "suggest" or "perhaps" should effectively be treated as terms-of-art, rather than their plain-meanings of an implied scienter.

Jurists and lawyers will likely struggle to accept such a linguistic tradition for two reasons. First, lawyers write to advocate; inherent to zealous advocacy is an assurance of a one-sided presentation. From this perspective, it is difficult to fathom how a seemingly hedged statement can

---

[29] James D. Watson & Francis H.C. Crick, *Molecular Structure of Nucleic Acids*, 171 NATURE 737, 737 (1953) (emphasis added).

[30] *See* Demosthenes B. Panagiotakos, *The Value of* p-*Value in Biomedical Research*, 2008 OPEN CARDIOVASCULAR MED. J. 97, 98 (2008) (explaining the meaning of p-value and its prevalent usage in research).

[31] *See, e.g.*, Jimmy J. Zhuang, Stephen A. Banse & Craig P. Hunter, *The Nuclear Argonaute NRDE-3 Contributes to Transitive RNAi in* Caenorhabditis elegans, 194 GENETICS 117, 122 fig. 2 (2013) (figure legend describes all the relevant data points with p-value less than 0.05 or 0.01).

[32] *Supra* note 32, at 121 ("these results ***suggest*** that *sago-1* and *nrde-3* ***may*** define complementary silencing pathways") (emphasis added).

nevertheless stand for a zealous scientific belief. The conventional and trite counterargument is that scientists believe in some form of objective and progressive "truth" and therefore do not slant any wording. However, throughout this Essay, I have instead argued that, merely out of tradition, scientists write in a more measured style than lawyers do. Tentative, non-zealous language is the linguistic tradition of science. Second, lawyers and jurists focus extensively on language because it is the core of their workproduct; a poorly-written legal workproduct is a poor workproduct. By contrast, while scientists do strive for clarity, the core of their workproduct is not in its language. Rather, language is merely a means to delineate scientific experiments, which is the actual scientific workproduct.[33] In today's digital age of high resolution media files, experiments can be depicted clearly in graphical or video form,[34] which makes the language of a scientific paper an even lower priority. Therefore, courts should not close-read something not intended for close-reading.

IV. A RECONCILIATION

Before endorsing the agency's decision, the *Benzene* Court was seeking a reassuring degree of certainty for any scientific evidence that the agency relied upon. During such searches for certainty, courts are likely to envision two different standards. On one level, the courts perceive science as a progressive endeavor, where mistakes are common and their corrections are occasions for joy because scientific knowledge has advanced. At this level, the requisite certainty is low. On another level, the courts perceive legal determinations as a coercive endeavor, where mistakes are never tolerated because they carry grave costs. At this level, the requisite certainty is much higher. Using this two-tier certainty framework, it is not difficult to imagine why the *Benzene* Court reacted the way it did: findings with "likely" support are perfectly fine for science because they can always be corrected later, but "likely" findings are too tenuous to ground coercive legal powers on.

In principle, I agree with this two-tier certainty framework. Legal determinations should require higher certainty than scientific findings. However, the means by which the *Benzene* Court and subsequent courts

---

[33] *See generally* Elisa M. Stone, *Guiding Students to Develop an Understanding of Scientific Inquiry: A Science Skills Approach to Instruction and Assessment*, 13 CBE LIFE SCI. ED. 90 (2014) (describing the importance of experimental skills as the core of science abilities). The great Michael Faraday is also often cited for the famous quip that an experiment not published is an experiment not done. This anecdote reflects the primary scientific importance of the experiment and ancillary role of language (of publication) in facilitating experiments.

[34] *See, e.g.*, JOURNAL OF VISUALIZED EXPERIMENTS, www.jove.com (last accessed Jan. 7, 2016).

have chosen to evaluate certainty is grossly inappropriate. Close-reading of scientific literature for words like "probable," "perhaps," "likely," etc. is an imprecise and incorrect way to seek higher certainty because it overlooks science's linguistic tradition. Scientific writings use such halting qualifiers as terms-of-art, which are entirely divorced from actual measures of certainty. Instead, more precise certainty indicators are in fact available in the scientific literature, in the form of *p* values and other statistical metrics. Thus, future courts in search of scientific certainty should not rely on language selected and pasted from scientific literature, especially if opportunist lawyers seek to use such qualified linguistics to unfairly damage the evidence's credibility. Rather, courts should directly examine the underlying scientific findings' statistical certainty. Obviously, a proposal requiring courts to undertake such an in-depth substantive review of scientific evidence needs to be assessed in terms of costs and benefits. Regardless, courts should never rely, even partially, on the *language* of the scientific evidence within the administrative record because of the linguistic tradition discussed in this Essay.