ADMITTING SCIENCE: PROBLEMS WITH CURRENT ADMISSIBILITY STANDARDS

by
William O'Donohue & Sneha Gupta***

Attorneys use expert testimony to advance legal positions in courtrooms. Because expert testimony can have tremendous influence over judges and juries, it is admitted only if it meets specific admissibility standards, i.e., Frye *and* Daubert. *These standards provide different admissibility criteria, have been adopted by distinct sets of jurisdictions, and, we argue, are of questionable validity. These standards are attempts to ensure that only knowledge is admissible and are, then, essentially an epistemic matter. Science serves as a proxy for knowledge because science is the epistemic process that has been successful at generating knowledge. However, both epistemology and the philosophy of science are complex and unsettled. This review seeks (1) to enumerate the strengths and weaknesses of existing admissibility standards, (2) to propose an improved multidimensional set of admissibility criteria that better capture the manifold ways of evaluating science, and (3) employ these criteria to evaluate a set of claims in psychiatry known as the child sexual abuse accommodation syndrome to demonstrate the implementation of these new admissibility criteria.*

* William O'Donohue is a professor of psychology at University of Nevada, Reno, and a licensed clinical psychologist. He has received advanced degrees in clinical psychology and philosophy and has published over 90 books and 300 journal articles and book chapters. He has directed a clinic at UNR for the past 30 years that provides free assessment and psycho-therapy to children and adults who have been sexually abused, supported by a grant from the National Institute of Justice.

** Sneha Gupta is clinical psychology doctoral student at University of Nevada, Reno. Her current research focuses on serial homicide and certain crime scene behaviors exhibited by these offenders (such as trophy-taking) and their relationship with personality disorders and paraphilias.

INTRODUCTION

There is frequently a need for attorneys to use expert testimony to advance their legal positions in various courts.[1] While lay or fact witnesses are typically limited to testifying about what they have directly experienced, expert witnesses can offer opinions based on their education and experience.[2] Expert witnesses claiming to present scientific information must first be vetted under standards regarding the proper admission of expert evidence to ensure that their proffered testimony emanates from within the legitimate bounds of scientific knowledge.[3] This is particularly important

[1] Margaret A. Berger, *Expert Testimony: The Supreme Court's Rules*, ISSUES SCI. & TECH., Summer 2000, at 57, 57.

[2] Edward Imwinkelried, *Distinguishing Lay from Expert Opinion: The Need to Focus on the Epistemological Differences Between the Reasoning Process Used by Lay and Expert Witnesses*, 68 SMU L. REV. 73, 78–79 (2015).

[3] BRIAN H. BORNSTEIN & EDIE GREENE, THE JURY UNDER FIRE 130 (2017); David DeMatteo, Sarah Fishel & Aislinn Tansey, *Expert Evidence: The (Unfulfilled) Promise of* Daubert, 20 PSYCH. SCI. PUB. INT. 129, 129 (2019); Sanja Kutnjak Ivković & Valerie P. Hans, *Jurors'*

as research suggests expert witnesses are often persuasive for judges and juries.[4] Essentially, then, courts, by adopting admissibility standards, aim to exclude what the early Greek philosophers called *doxa* (mere opinions or guesses, e.g., "It is my opinion/guess that there were two additional shooters on the grassy knoll") and instead are appropriately interested in only allowing testimony that has some sort of superior warrant to mere opinion such as *episteme* or *scientia* (knowledge).[5] These latter beliefs meet additional epistemic characteristics to count as knowledge, such as sound, undefeated evidential warrant.[6]

Thus, at its core, the problem of articulating a sound admissibility standard is epistemic—every claim that appropriately counts as knowledge is potentially admissible (if relevant), while claims that are not knowledge are excluded. Whether a claim is scientific becomes a valid shorthand proxy that the claim is (at least reasonably) epistemically sound due to the view that what counts as science also counts as knowledge. It therefore follows that the most useful responses to the admissibility question ultimately must be based in epistemology—the branch of philosophy that examines the question of what knowledge is.[7] Philosophers of science and some epistemologists have argued that if one wants to find answers to questions of epistemology then it is best to study science because of its unique history of producing a multitude of knowledge claims.[8]

Unfortunately, the task of studying what counts as science or how science "works" is complicated. Philosophers of science call the attempt to identify criteria that validly distinguishes science from nonscience the "problem of demarcation."[9] Notice the complex reflexive problem involved here—if one wants to understand what is unique about science, one must study science—but to accomplish this, one

Evaluations of Expert Testimony: Judging the Messenger and the Message, 28 L. & SOC. INQUIRY 441, 442 (2003).

[4] *See, e.g.*, Richard E. Redding, Marnita Y. Floyd & Gary L. Hawk, *What Judges and Lawyers Think About the Testimony of Mental Health Experts: A Survey of the Courts and Bar*, 19 BEHAV. SCI. L. 583, 584, 591–92 (2001); Daniel A. Krauss & Bruce D. Sales, *The Effects of Clinical and Scientific Expert Testimony on Juror Decision Making in Capital Sentencing*, 7 PSYCH., PUB. POL'Y, & L. 267, 271, 274 (2001); Allen Raitz, Edith Greene, Jane Goodman & Elizabeth F. Loftus, *Determining Damages: The Influence of Expert Testimony on Jurors' Decision Making*, *in* INSTRUCTIONS, VERDICTS, AND JUDICIAL BEHAVIOR 231, 239 (Robert M. Krivoshey ed., 1994).

[5] Jessica Moss, *Plato's Doxa*, 61 ANALYTIC PHIL. 193, 193–94 (2020); *see* Baron Reed, *Certainty*, STANFORD ENCYC. PHIL., https://plato.stanford.edu/archives/spr2022/entries/certainty/ (Feb. 21, 2022).

[6] W. V. Quine, *The Nature of Natural Knowledge*, *in* MIND AND LANGUAGE 67, 74–75 (Samuel Guttenplan ed., 1975) [hereinafter Quine, *Nature of Natural Knowledge*].

[7] LARRY LAUDAN, PROGRESS AND ITS PROBLEMS: TOWARD A THEORY OF SCIENTIFIC GROWTH 1 (1978).

[8] KARL POPPER, THE OPEN SOCIETY AND ITS ENEMIES xv–xvi (Princeton University Press, 2020) (1945). *See generally* Quine, *Nature of Natural Knowledge*, *supra* note 6, at 71–72.

[9] POPPER, *supra* note 8, at 654, 658.

must first use some criteria to define what science is—which ultimately seems to be the question. For example, there are those who argue that astrology, ESP, some branches of alternative medicine, mesmerism, creation science, homeopathy, chiropractic, psychoanalysis, recovered memories, and Marxism are scientific, and thus should be included in any study of what is characteristic about science, while most others argue they are not.[10] The question becomes, which of these, if any, should be included in the study of science in an attempt to develop valid demarcation criteria?

The Nobel laureate in physics, Richard Feynman, for example, called pseudoscience "cargo cult science" after the practice of some indigenous peoples during World War II attempting to build airlines from bamboo but with none of the materials of an actual airplane such as the required engines or proper aerodynamics.[11] The Oxford English Dictionary defines pseudoscience as "A pretended or spurious science; a collection of related beliefs about the world mistakenly regarded as being based on scientific method or as having the status that scientific truths now have."[12]

These pseudosciences are not just a matter of innocuous mislabeling because pseudoscientific beliefs have done incalculable harm, such as the Nazi pseudoscience of racial differences and its resultant eugenics, advanced by them as proper science.[13] Thus, a key reason to exclude pseudoscientific content from courts is that these claims are not knowledge and therefore excluding them can prevent harm of unjust verdicts that can be produced if the claims are allowed to have a role in legal decision making. When scientific knowledge is distinguished from its pretenders, the persuasive force properly associated with actual science is not misplaced. An admissibility criterion that accurately captures knowledge is intended to accomplish this end—or in the philosopher of science Larry Laudan's colorful phrase, perform the "critical stable-cleaning chores for which it was originally intended."[14]

---

[10] *See* David K.B. Nias & Geoffrey A. Dean, *Astrology and Parapsychology*, *in* HANS EYSENCK: CONSENSUS AND CONTROVERSY 361 (1986) (discussing how the fields of astrology and parapsychology have failed to adequately account for the numerous inconsistencies that have been noted by critics); Mike Thelwall, *Alternative Medicines Worth Researching? Citation Analyses of Acupuncture, Chiropractic, Homeopathy, and Osteopathy*, 126 SCIENTOMETRICS 8731, 8737–43 (2021) (concluding that research articles regarding acupuncture, chiropractic, homeopathy, and osteopathy can be useful in advancing scientific research even though many scientists denigrate these fields); Michael Burawoy, *Marxism as Science: Historical Challenges and Theoretical Growth*, 55 AM. SOCIO. REV. 775 (1990) (discussing various arguments both for and against Marxism being a science).

[11] *See generally* Richard P. Feynman, Cargo Cult Science, Commencement Address at Cali. Polytechnic Univ., *in* 37 ENG'G & SCI., at 10, 11 (1974).

[12] *Pseudoscience*, OXFORD ENG. DICTIONARY (2d ed. 1989).

[13] *See generally* ROBERT N. PROCTOR, RACIAL HYGIENE: MEDICINE UNDER THE NAZIS (1988).

[14] R. S. COHEN & LARRY LAUDAN, PHYSICS, PHILOSOPHY AND PSYCHOANALYSIS 111, 122 (1983).

However, what properly counts as scientific knowledge also depends on at least three other additional criteria:

1. *The methodological adequacy of the relevant research*. This includes the quality of the samples, the quality of the control conditions, the quality of the statistical analyses, and many other design facets—which also vary across scientific subdomains and time as the sophistication of the science develops. As Munz has suggested, in science, we also learn how to learn.[15]
2. *The actual results of the relevant studies*. This includes whether the results corroborated, falsified, or were mixed, and possibly inconclusive—which are not entailments but matters of judgment that can change over time as more studies are conducted or methodological advances are made in the field.
3. *The variance of evaluative standards over the complex edifice of science*. This includes multiple and fundamentally diverse components such as theories, models, simulations, hypotheses, arguments about proper methodologies (which might be regarded as a key meta-issue), case studies, taxonomies, studies of the validity and reliability of measurement operations, qualitative literature reviews, meta-analyses, professional ethical codes, book reviews, unpublished conference papers, and even thought experiments.

Evaluating the epistemic merits of each of these different kinds of scientific activity can be at issue because either singularly or conjointly, these become involved in the scientist's web of belief.[16] The epistemic warrant for some claim in any part of the scientific edifice also may not remain stable as any work is done in any of these activities and thus, over time, scientific activity is constantly increasing or decreasing the relevant evidence. Therefore, judgments about epistemic warrant in science are at least partly matters of an ever-changing historical record and not just dependent on a purely logical analysis (e.g., is the claim in principle falsifiable?).

It must also be noted that for obvious practical reasons these legal admissibility criteria are ultimately constructed for a binary decision—either the claims are admissible, or they are not. However, epistemic warrant is a matter of (again, often changing) *degree* and therefore, if a binary decision must be made, an additional problem is that some cut score/demarcation must be argued which properly reduces these dynamic, possibly domain specific, and multifactorial quantitative considerations into an admissible/inadmissible dichotomy. Finally, it is critical to note that this decision is even more complicated because, at times, in science a particular theory has scientific competitors, and when this is the case, the question becomes:

---

[15] *See* PETER MUNZ, OUR KNOWLEDGE OF THE GROWTH OF KNOWLEDGE (Routledge, 2014).

[16] Quine theorized that beliefs are a web in that each statement depends on other statements to form a belief. *See* W. V. QUINE & J.S. ULLIAN, THE WEB OF BELIEF (2d ed. 1978).

Which of these relevant competing theories currently has the best epistemic warrant on these varied evaluative parameters?[17]

In light of the above, we will argue that admissibility standards regarding scientific expert evidence are not a settled area in either science or the law, and therefore two questions arise: (1) To what extent the two major admissibility standards (*Frye* and *Daubert*) currently in use are effective in accurately evaluating what constitutes science? and (2) If these are inadequate, what might be used in their place? For example, Robinson purports that *Daubert* does not provide a global definition of scientific reliability that can be applied to specific areas of expert testimony.[18] Robinson claims that said standard "neither provides a clear picture of epistemology nor realistically weighs the burdens it imposes on judicial institutions."[19] Some scholars have stated, "*Daubert* encourages a 'pragmatic constructivism' about epistemological issues, sidestepping direct consideration of epistemology and inviting lower courts to do the same."[20]

We also note that this unsettledness is due to some vagueness regarding what kind of problematic beliefs these admissibility standards are meant to exclude—e.g., all nonscience; or all pseudoscience; all bad science (perhaps because the research was poorly designed, executed, or interpreted); or even inferior science relative to a competitor; or any knowledge claim that is poorly supported.[21] These alternatives have quite different domains. For example, all would exclude astrology but only the "inferior science" criterion would exclude something like older medical interventions such as exploratory surgeries. Because the standard of care and scientific technology have improved, interventions that were once considered evidence-based, are now considered obsolete due to the scientific and technological advances in imaging, such as CAT scans or MRIs.[22] If the practical issue is to help the trier of fact by excluding expert testimony that is at variance with the most warranted

---

[17] *See* Howard Sankey, *Epistemic Objectivity and the Virtues*, PHIL. SCI., Oct. 2020, at 5, 8 (arguing there is "no universally agreed or accepted system" of ranking theoretical virtues into a fixed order of precedence and that scientists must use a process of deliberative judgment to find the most relevant scientific virtues).

[18] Robert Robinson, Daubert v. Merrell Dow Pharmaceuticals *and the Local Construction of Reliability*, 19 ALBANY L. J. SCI. & TECH. 39, 42 (2009).

[19] *Id.* at 49.

[20] *Id.* at 73; *accord* David Caudill & Richard E. Redding, *Junk Philosophy of Science?: The Paradox of Expertise and Interdisciplinarity in Federal Courts*, 57 WASH. & LEE L. REV. 685, 756 (2000).

[21] Bridget Mary McCormack, *Scientific Evidence*, *in* SCIENCE BENCH BOOK FOR JUDGES 14, 21 (Peggy Hora, Brian MacKenzie, Theodore Stalcup & David Wallace, eds., 2d ed. 2020) (ebook); Marc Picker, *Scientific Evidence*, *in* SCIENCE BENCH BOOK FOR JUDGES 30 (Peggy Hora, Brian MacKenzie, Theodore Stalcup & David Wallace, eds., 2d ed. 2020) (ebook).

[22] G. Boodman & Larry Thompson, *Where Warrior and Statesman Become One*, WASH. POST (Jan. 29, 1990), https://www.washingtonpost.com/archive/lifestyle/wellness/1990/01/30/where-warrior-and-statesman-become-one/3b7f61fb-cba9-40b5-81ae-f0c189f5728b.

knowledge about some subject matter that is currently available, and conversely admitting only the most warranted knowledge about such subject matter that is currently available, then it would seem that a proper admissibility standard should not only exclude pseudoscience but inferior and bad science too.[23]

It is important to note that there are two kinds of errors that a problematic admissibility standard can cause: (1) false positives—ruling something is properly scientific when it is not and (2) false negatives—ruling something is not scientific when, in fact, it is.[24] Perhaps a third type of error involves a problematic application of a sound admissibility standard—i.e., when a judge misapplies a valid standard. This last category of error might emphasize the standard's need for clarity and precise decision-making rules to implement it faithfully. As Pigliucci claims, because much hinges on a demarcation criterion, "This is a serious game, which ought to be played seriously."[25]

Next, we bring the issue of the appeal to scientific status as a key component of an admissibility standard into perspective by (1) providing a critical analysis of both the *Frye* and *Daubert* standards, particularly in the way these rely on a conception of science;[26] (2) providing an overview of the philosophy of science on what science is;[27] and then (3) analyzing the child sexual assault accommodation syndrome (CSAAS) that has been admitted in many courts under both the *Frye* and *Daubert* standards, and propose a standard that better captures the complexity of judgments about what constitutes science.[28]

I. STANDARDS GOVERNING EXPERT TESTIMONY

A. Frye *Standard*

In the United States, the need for expert evidence that relies on science was first recognized in 1923 in *Frye v. United States*.[29] In this landmark case, the issue of admissibility of expert evidence was raised concerning a novel test purported to

---

[23] However, another key question is, should a proper admissibility criterion also exclude warranted belief from intellectual disciplines that are not science but ones that have other reasonable epistemic practices such as testimony from historians, mathematicians, logicians, and craftsmen? For example, should testimony from the received view of history regarding the Holocaust be regarded as epistemically warranted while denialist accounts should not?

[24] Munia Jabbar, *Overcoming* Daubert*'s Shortcomings in Criminal Trials: Making the Error Rate the Primary Factor in* Daubert*'s Validity Inquiry*, 85 N.Y.U. L. REV. 2034, 2059–62 (2010).

[25] Massimo Pigliucci, *The Demarcation Problem: A (Belated) Response to Laudan*, *in* PHILOSOPHY OF PSEUDOSCIENCE: RECONSIDERING THE DEMARCATION PROBLEM 9, 21 (Massimo Pigliucci & Maarten Boudry eds., 2013).

[26] *See* discussion *infra* Section I.A–D.

[27] *See* discussion *infra* Part III.

[28] *See* discussion *infra* Section IV.A.

[29] Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).

detect deception. The defendant in the case had been convicted of second-degree murder and appealed the trial court's decision, claiming that said court made an error when the defendant's expert witness was not allowed to proffer testimony on a systolic blood pressure "deception test."[30] The defendant reasoned that the expert testimony in question should have been admitted by the trial court under the common law test for admissibility because "the question involved [did] not lie within the range of common experience or common knowledge, but [required] special experience or special knowledge."[31]

In determining the admissibility of novel scientific evidence, the appellate court ruled, "the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs."[32] This is essentially a sociological criterion and only secondarily an epistemic one.[33] The standard provided that: (1) such consensus can be accurately determined (and more precisely defined—e.g., at least 51%?) and (2) this sociological consensus closely tracks the actual evidential status of the claim.

Because the deception test in question was judged yet to have gained sufficient standing and scientific recognition in its field, the appellate court held that it was inadmissible, and the trial court was correct in denying the admissibility of expert testimony.[34] This standard became commonly known as the "general acceptance" test or the *Frye* standard, and was offered to assist judges in keeping pseudoscience, fringe, or not "sufficiently established" science out of the courtroom.[35]

Though *Frye* became the predominant standard for admitting expert evidence, courts were slow to recognize it[36] and only a few dozen cases cited it from its advent until the 1960s.[37] However, this lack of usage at that time could have also been attributed to fewer forensic/scientific advances and the general practice of disallowing expert evidence in civil proceedings.[38] In any event, the *Frye* standard remained

---

[30] *Id.* at 1013.

[31] *Id.* at 1014.

[32] *Id.*

[33] BRIAN LEITER, *The Epistemology of Admissibility: Why Even Good Philosophy of Science Would Not Make for Good Philosophy of Evidence*, 1997 BYU L. REV. 803, 818–19 (1997).

[34] *Frye*, 293 F. at 1014.

[35] *See* Zach Alter, Note, *Unpacking* Frye–Mack*: A Critical Analysis of Minnesota's* Frye–Mack *Standard for Admitting Scientific Evidence*, 43 MITCHELL HAMLINE L. REV. 626, 649–55 (2017) (explaining how the *Frye–Mack* standard allows deference to the relevant scientific community and their accepted standards, rather than requiring judges to "become amateur scientists").

[36] *Id.* at 629.

[37] David E. Bernstein, Frye, Frye *Again: The Past, Present, and Future of the General Acceptance Test*, 41 JURIMETRICS 385, 388 (2001).

[38] Alter, *supra* note 35, at 629.

largely unchallenged for over half a century until 1975 with the introduction of the Federal Rules of Evidence (FRE).[39]

The FRE were developed to guide criminal and civil litigation in federal courts.[40] Rule 702 of the first iteration provided that, "[i]f scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise."[41] This rule had a wider scope than the *Frye* standard, as Rule 702 allowed not just science but also "technical, or other specialized knowledge."[42] This is praiseworthy, although, again, the epistemic question remains: What properly counts as knowledge? Note also this standard focuses on properties of the expert rather than directly on the properties of the knowledge claims.

These rules differed in other important ways from the *Frye* standard, and the legal community remained at odds about embracing Rule 702 or the *Frye* standard.[43] The issue became further contentious when, by 1991, four federal circuits and more than fifteen states had rejected the *Frye* standard, citing vagueness and inconsistency in its application as the primary issue.[44]

B. Criticisms of the Frye *Standard*

- *Vagueness of the term "general acceptance*." There is no specification on how judges should determine general acceptance in the field.[45] The standard is devoid of relevant definitions and has left it to the courts to decide how to determine the criteria that should be employed to establish general acceptance.[46] "Courts can easily 'manipulate the parameters of the relevant "scientific community" and the level of agreement needed for "general acceptance."'"[47] Thus, the *Frye* standard

[39] DeMatteo, Fishel, & Tansey, *supra* note 3, at 129; Federal Rules of Evidence, Pub. L. No. 93–595, 88 Stat. 1926 (1975) (codified at 28 U.S.C. app.).

[40] 88 Stat. 1926.

[41] *Id.* at 1937 (Rule 702).

[42] *Id.*

[43] *The* Frye *Standard and Rule 702 of the Federal Rules of Evidence*, US LEGAL, https://forensiclaw.uslegal.com/litigation-history-of-forensic-evidence/the-frye-standard-and-rule-702-of-the-federal-rules-of-evidence/#google_vignette (last visited Apr. 9, 2025).

[44] FAUST F. ROSSI, EXPERT WITNESSES 32–42 (Am. Bar Ass'n 1991).

[45] *See* Steven J. Grossman & Christopher K. Gagne, *Science and Scientific Evidence*, 25 CONN. L. REV.1053, 1054–55 (1993).

[46] *See* Paul C. Giannelli, *The Admissibility of Novel Scientific Evidence:* Frye v. United States *a Half-Century Later*, 80 COLUM. L. 1197, 1205–06, 1208 (1980); Troy M. Horton, Comment, *The Debate is Over:* Frye *Lives No More*, 19 T. MARSHALL L. REV. 379, 384 (1994).

[47] Edward R. Becker & Aviva Orenstein, *The Federal Rules of Evidence After Sixteen Years—The Effect of "Plain Meaning" Jurisprudence, the Need for an Advisory Committee on the*

forces judges to deal with the particularly vexing epistemic/sociological problem of the (current) views of the relevant community of scientists. A judge's task is made more difficult because practicing scientists are rarely, if ever, polled regarding what they "accept" and what they do not.[48] Thus, this decision about general acceptance must be made by judges without what is, in principle, the most relevant direct evidence to correctly answer it.

Even if such poll results were available, further questions would still need to be addressed: What exactly constitutes the relevant community of scientists—e.g., all chemists?; all chemists in the relevant scientific subfield?; all chemists who have actually published peer reviewed articles on the matter?; all applied chemists who practice in the field but have not published? and so on. Does "general acceptance"[49] mean acceptance by at least 51% of the practitioners in the field?; some super majority such as nearly all?; acceptance among the scientific leaders of the field?; or higher rates of acceptance than its scientific competitors, if any? Moreover, further clarification on what "acceptance" means is also desirable: accept that it is scientific?; accept that it is at least in principle a possible or perhaps more strongly a plausible scientific finding or explanation?; accept that it is currently a promising scientific explanation but yet to be sufficiently tested?; or accept that currently it is the best scientific explanation? Further questions can be raised about the meaning of each of these predicates.

- *General acceptance of the scientific technique and/or underlying theory.* The *Frye* standard is unclear what "general acceptance" pertains to—e.g., the scientific technique in question, the underlying theory or explanation of such technique or both.[50] For example, the standard is silent on whether judges need to gauge the question of to what extent the scientific technique "works" versus why it works, which is perhaps necessary in order to fully understand the validity of the scientific technique.[51]

Rules of Evidence, and Suggestions for Selective Revision of the Rules, 60 GEO. WASH. 857, 878 (1992) (quoting United States v. Downing, 753 F.2d 1224, 1236 (3d Cir. 1985)).

[48] *See* Lisa Gonzalez, *The Admissibility of Scientific Evidence: The History and Demise of* Frye v. United States, 48 U. MIAMI L. REV. 371, 374–75 (1993).

[49] Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923).

[50] Kuhn v. Sandoz Pharms. Corp., 14 P.3d 1170, 1180 (Kan. 2000).

[51] Andrew R. Stolfi, Note, *Why Illinois Should Abandon* Frye*'s General Acceptance Standard for the Admission of Novel Scientific Evidence*, 78 CHI. KENT L. 861, 891–92 (2003).

- *Inconsistent application of the* Frye *Standard*. During the short period between 1975 (adoption of Rule 702 and the FRE[52]) and 1993 (the *Daubert* decision[53]), courts applied the *Frye* standard more frequently but inconsistently.[54] For instance, the Court of Appeals for the District of Columbia Circuit rejected voiceprint evidence citing lack of general acceptance in the field,[55] but the Court of Appeals for the Fourth Circuit admitted voiceprint evidence and rejected the applicability of the *Frye* standard in lieu of another admissibility standard.[56] This is why some scholars argue that the heterogeneous outcomes created by this standard far outweigh its advantages.[57]
- *Lack of criteria to determine relevant scientific community.* To correctly apply the *Frye* standard, judges also need to identify the appropriate field to which the expert evidence in question belongs; however, there is no guidance in the *Frye* standard on how to achieve this.[58] This is particularly problematic because scientific techniques do not always fit neatly into a single discipline or field,[59] especially when such techniques are novel and interdisciplinary.[60]
- *Gap between developing a scientific technique and its judicial acceptance*. The *Frye* standard can be insensitive to time lags between the development of novel science, its general acceptance by scientists, and finally, its judicial acceptance.[61] Judges can be ill-equipped to identify the latest research which consequently creates a conservative requirement for admission of novel science.[62] Also, some degree of lag is

---

[52] Federal Rules of Evidence, Pub. L. No. 93–595, 88 Stat. 1926 (1975) (codified at 28 U.S.C. app.).

[53] Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579 (1993).

[54] Alter, *supra* note 35, at 630.

[55] United States v. Addison, 498 F.2d 741, 743 (D.C. Cir. 1974).

[56] United States v. Baller, 519 F.2d 463, 464–67 (4th Cir. 1975); *see also* Stan Kitzinger, Note, *The Supreme Court Waves Good-Bye to* Frye: Daubert v. Merrell Dow Pharmaceuticals, Inc., 58 ALB. L. REV. 575, 583 (1994) (recognizing *Baller* as a repudiation of the *Frye* standard).

[57] Giannelli, *supra* note 46, at 1207–08; Horton, *supra* note 46, at 383.

[58] Giannelli, *supra* note 46, at 1208; Horton, *supra* note 46, at 384.

[59] Horton, *supra* note 46, at 384; Stolfi, *supra* note 51, at 888–89.

[60] Craig A. Kubiak, Comment, *Scientific Evidence in Wisconsin: Using Reliability to Regulate Expert Testimony*, 74 MARQ. L. REV. 261, 266–67 (1990).

[61] Edward J. Imwinkelried, *A New Era in the Evolution of Scientific Evidence—A Primer on Evaluating the Weight of Scientific Evidence*, 23 WM. & MARY L. REV. 261, 265 (1981); Case Comment, *The* Frye *Standard Finally Fries: Has* Daubert v. Merrell Dow *Furthered the Use of Scientific Evidence in Our Legal System?*, 14 REV. LITIG. 315, 322–23 (1994).

[62] *See* Giannelli, *supra* note 46, at 1215–16 (describing several courts' rejection of voiceprint evidence for inadequate corroboration and impartiality).

often justifiable, even necessary, to establish its replicability.[63] However, novel science can be resisted by the scientific establishment, which led the renowned physicist Max Planck to believe that science advances one funeral at a time.[64] This delayed recognition of scientific evidence can result in the exclusion of valid findings from the courtrooms.[65]

- *Questionable research practices*. The acceptance of a scientific view can be a product of many factors, including the methodological quality of the studies used to test it (including a frank discussion of the severity of the scientific tests[66]), as well as a discussion of bias that may exist in the field (for example, political bias among psychologists resulted in the creation of a right wing authoritarian measurement scale, many years before the creation of a left wing authoritarian scale[67]), an examination of any scientific misbehavior such as faked data or the frequent use of questionable research practices,[68] the clarity and effect sizes of the findings from these studies,[69] the number and importance of the untested aspects of the scientific theory (perhaps bracketed with an understanding of pragmatic, financial, and ethical constraints resulting in this),[70] the number and importance of anomalous or

[63] *See* Douglas R. Dalgleish & Teresa J. Stewart, *Thermography in Missouri's Courts: Is the* Frye *Standard Alive and Well?*, 60 UMKC L. REV. 467, 484 (1991).

[64] *See* MAX PLANCK, *A Scientific Autobiography*, *in* SCIENTIFIC AUTOBIOGRAPHY AND OTHER PAPERS 13, 33–34 (Frank Gaynor trans., 1950) ("A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it.").

[65] Grossman & Gagne, *supra* note 45, at 1054–55.

[66] POPPER, *supra* note 8, at xvi, 425. This is discussed further *infra* notes 173–74 and accompanying text.

[67] *See generally* Thomas H. Costello, Shauna M. Bowes, Sean T. Stevens, Irwin D. Waldman, Arber Tasimi & Scott O. Lilienfeld, *Clarifying the Structure and Nature of Left-Wing Authoritarianism*, 122 J. PERSONALITY & SOC. PSYCH. 135, 135–37 (2022).

[68] William O'Donohue & Akihiko Masuda, *Questionable Research Practices in Clinical Psychology*, *in* AVOIDING QUESTIONABLE RESEARCH PRACTICES IN APPLIED PSYCHOLOGY 3, 6 (2022). *See generally* Yu Xie, Kai Wang & Yan Kong, *Prevalence of Research Misconduct and Questionable Research Practices: A Systematic Review and Meta-Analysis*, SCI. & ENG'G ETHICS, June 2021, at 1–3, 25 (2021).

[69] John P.A. Ioannidis, *Why Most Published Research Findings Are False*, 2 PLOS MED. 696, 697 (2005).

[70] *See, e.g.*, Karlanne Thune Hammerstrøm & Arild Bjørndal, *If There are No Randomised Controlled Trials, Do We Always Need More Research?*, COCHRANE DATABASE OF SYSTEMATIC REVIEWS (Mar. 14, 2011), https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.ED000024/full.

potentially falsifiable results,[71] the number and seriousness of conceptual problems in the theory (e.g., such as unclear terminology; Wittgenstein famously said, "In psychology there are experimental methods and *conceptual confusion*."[72]); the generalizability of these studies to the case at hand;[73] and finally, the relative evidential status of its scientific competitors, if any, on each of these dimensions.[74] It could be time consuming and burdensome to address each of these dimensions explicitly in each case, but to do so would certainly allow a clearer, more detailed understanding of the status of the scientific view.

- *Lack of re-examination of admitted evidence*. Once expert evidence has been admitted in court under *Frye*, generally, it is not reexamined in subsequent cases where the theory or technique might be introduced as evidence.[75] This can be problematic because science changes over time, sometimes drastically, in what the influential philosopher of science Thomas Kuhn called "scientific revolutions."[76]

C. Daubert *Standard*

The issue of admitting expert evidence came under scrutiny once again in 1993 in *Daubert v. Merrell Dow Pharmaceuticals*.[77] Plaintiffs in this case were born with congenital disabilities that they claimed were caused when their mothers regularly ingested an anti-nausea drug manufactured by the defendant when pregnant with them.[78] The case was first tried in 1989 when the *Frye* standard was used to decide the admissibility of expert evidence. The plaintiffs presented eight expert witnesses, including an epidemiologist, Dr. Shanna Swan, who held a master's degree in

---

[71] THOMAS S. KUHN, THE STRUCTURE OF SCIENTIFIC REVOLUTIONS 145–46 (4th ed. 2012) [hereinafter KUHN, STRUCTURE OF SCIENTIFIC REVOLUTIONS].

[72] LUDWIG WITTGENSTEIN, PHILOSOPHICAL INVESTIGATIONS 232 (G.E.M. Anscombe trans., Basil Blackwell 1967) (1953) [hereinafter WITTGENSTEIN, PHILOSOPHICAL INVESTIGATIONS].

[73] David E. Bernstein & Jeffrey D. Jackson, *The* Daubert *Trilogy in the States*, 44 JURIMETRICS J. 351, 355 (2004).

[74] KUHN, STRUCTURE OF SCIENTIFIC REVOLUTIONS, *supra* note 71, at 109.

[75] *See* United States v Downing, 753 F.2d 1224, 1234 (3d Cir. 1985) ("Once a novel form of expertise is judicially recognized, this foundational requirement can be eliminated, as is done when, for example, fingerprint, ballistics, or x-ray evidence is offered."); United States v. Horn, 185 F.Supp.2d 530, 554 (D. Md. 2002) ("[G]iven the impact of the *stare decisis* doctrine, once a court, relying on *Frye*, had ruled that a doctrine or principle had attained general acceptance, it was all to [sic] easy for subsequent courts simply to follow suit.").

[76] KUHN, STRUCTURE OF SCIENTIFIC REVOLUTIONS, *supra* note 71, at 48.

[77] Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 591–97 (1993).

[78] *Id.* at 582.

biostatistics from Columbia University and a doctorate in statistics from University of California, Berkeley.[79] At the time of the trial, Dr. Swan was the section chief of the California Department of Health and Services section that studied causes of congenital disabilities and had also served as a consultant to the World Health Organization, the Food and Drug Administration, and the National Institutes of Health.[80]

The plaintiffs presented expert evidence to demonstrate a causal link between the defendant's drug and the plaintiffs' congenital disabilities, which included chemical analyses of the drug, studies of the drug's effects on animals in the womb and on cells in a laboratory, and a meta-analysis involving several studies with large sample sizes.[81] This meta-analysis, which was presented by Dr. Swan, revealed a small but statistically significant link between congenital disabilities and the drug.[82] However, this analysis was not published in a peer-reviewed scientific journal before the trial and was therefore ruled inadmissible.[83] On the other hand, the defendant provided more than 30 published studies showing no significant link between the drug and congenital disabilities.[84] The case was dismissed on the basis that the plaintiffs did not offer evidence meeting the *Frye* standard, i.e., Dr. Swan's meta-analysis, which demonstrated a causal link, was an aggregation of other such studies and needed to be peer-reviewed before it would be considered generally accepted in the scientific field.[85] This is despite the fact that meta-analyses, as valuable syntheses of disparate individual studies, are considered as serving a unique and useful scientific function.

The plaintiffs eventually appealed to the U.S. Supreme Court, where the contention was no longer about the causal link between the defendant's drug and congenital disabilities, but about the criteria that courts should apply when admitting expert evidence.[86] The Supreme Court addressed this issue by stating the following:

> Rule 702 [of the Federal Rules of Evidence]—place[s] appropriate limits on the admissibility of purportedly scientific evidence by assigning to the trial judge the task of ensuring that an expert's testimony both rests on a reliable foundation and is relevant to the task at hand. The reliability standard is established by Rule 702's requirement that an expert's testimony pertain to

---

[79] *Id.* at 583 n.2.

[80] *Id.*

[81] *Id.* at 583.

[82] *Id.* at 583–84; Alexis Abboud, Daubert v. Merrell Dow Pharmaceuticals, Inc. *(1993)*, ARIZ. STATE UNIV.: EMBRYO PROJECT ENCYCLOPEDIA (May 29, 2017), https://embryo.asu.edu/pages/daubert-v-merrell-dow-pharmaceuticals-inc-1993.

[83] *Daubert*, 509 U.S. at 584.

[84] *Id.* at 582.

[85] *Id.* at 584–85.

[86] *Id.* at 585.

> "scientific . . . knowledge," since the adjective "scientific" implies a grounding in science's methods and procedures, while the word "knowledge" connotes a body of known facts or of ideas inferred from such facts or accepted as true on good grounds. The Rule's requirement that the testimony "assist the trier of fact to understand the evidence or to determine a fact in issue" goes primarily to relevance by demanding a valid scientific connection to the pertinent inquiry as a precondition to admissibility.[87]

Further, the trial judge is required to

> make a preliminary assessment of whether the testimony's underlying reasoning or methodology is scientifically valid and can be properly applied to the facts at issue. Many considerations will bear on the inquiry, including whether the theory or technique in question can be (and has been) tested, whether it has been subjected to peer review and publication, its known or potential error rate and the existence and maintenance of standards controlling its operation, and whether it has attracted widespread acceptance within a relevant scientific community. The inquiry is a flexible one, and its focus must be solely on principles and methodology, not on the conclusions that they generate. Throughout, the judge should also be mindful of other applicable Rules.[88]

Accordingly, the Supreme Court formulated a two-step analysis to be used by federal trial judges: (1) that the evidence is relevant and (2) that it is reliable.[89] In determining the reliability of evidence, the Court established a separate four-part test where the judge would be required to determine: (i) whether the theory or technique in question had been, or could have been, tested with appropriate standards and controls; (ii) whether such theory or technique had been "subjected to peer review and publication"; (iii) the "known or potential error rate"; and (iv) whether such theory or technique had "widespread acceptance" within the scientific community of that field.[90] This became known as the *Daubert* standard.[91] The case was then sent back to a lower court to re-evaluate the scientific evidence per the new standard.[92] The lower court ruled in favor of the defendant because their evidence

---

[87] *Id.* at 579–80.

[88] *Id.* at 580.

[89] *Id.* at 597.

[90] *Id.* at 580.

[91] *See* Rochkind v. Stevenson, 236 A.3d 630, 633 (Ct. App. Md., 2020) (choosing to "adopt *Daubert* as the governing standard by which trial courts admit or exclude expert testimony"); Ancho v. Pentek Corp., 157 F.3d 512, 519 (7th Cir., 1998) (holding that the trial court's "articulation and application of the *Daubert* standard was sufficient" to bar an expert witness at trial).

[92] *Daubert*, 509 U.S. at 598.

outweighed the plaintiffs'.[93] Rule 702 of the FRE was then amended to incorporate the *Daubert* standard.[94]

Though some expressed satisfaction with the *Daubert* standard, there was an ongoing debate about which standard should be followed in state courts across the country.[95] This argument raged for several years until *General Electric Co. v. Joiner* in 1997.[96] This case clarified the *Daubert* standard in two important ways. First, it clarified that courts may scrutinize the reliability of an expert's reasoning process as reflected in their general methodology and not place undue reliance on *ipse dixit* (i.e., unsupported statements).[97] Therefore, judges were able to ensure that the analytical gap between the experts' data (i.e., their methodology/technique) and their proffered opinions (i.e., conclusions based on such methodology/technique) was not too large or problematic in any other way. Second, judges have wide discretion to admit a broad range of expert testimony under *Daubert* even though the standard was silent on this aspect.[98] *Joiner* clarified that decisions of trial court judges admitting expert evidence would be reviewed *only* for abuse of such discretion.[99]

In 1999, a ruling from another case, *Kumho Tire Co. v. Carmichael*, was appended to the *Daubert* standard.[100] In this case, the plaintiff offered expert evidence by a tire failure analyst and the defendant contested the evidence, arguing it was inadmissible under the *Daubert* standard on account of being non-scientific and a matter of technical expertise.[101] The matter was appealed and reached the U.S. Supreme Court, which clarified that the *Daubert* standard was to be applied to the broader universe of *all* expert testimony[102] and not just scientific expert testimony.[103] *Daubert*, *Joiner* and *Kumho* together became the "*Daubert* trilogy" which was codified in Rule 702 of the FRE.[104]

---

[93] *Id.*

[94] *See* advisory committee's note to the 2000 amendment of FED. R. EVID. 702.

[95] David Bernstein, *Hauling Junk Science Out of the Courtroom*, WALL ST. J., Jul. 13, 1993, at A16; Bernstein & Jackson, *supra* note 73, at 352–54.

[96] Gen. Elec. Co. v. Joiner, 522 U.S. 136 (1997).

[97] *Id.* at 146.

[98] *Id.* at 142.

[99] *Id.* at 146.

[100] Kumho Tire Co., Ltd. v. Carmichael, 526 U.S. 137, 141 (1999).

[101] *Id.* at 137, 145, 146–47.

[102] Bernstein & Jackson, *supra* note 73, at 353–55.

[103] Jeanne Wiggins, Kumho Tire Co. v. Carmichael*:* Daubert*'s Gatekeeping Method Expanded to Apply to all Expert Testimony*, 51 MERCER L. REV. 1325, 1328–29 (1999).

[104] Bernstein & Jackson, *supra* note 73, at 351, 354–55.

D. Criticisms of the Daubert *Standard*

A preliminary issue that must be mentioned is that it is unclear if judgments about each criterion of this standard are categorical (meets vs. fails to meet) or whether all or some subset of these criteria are not a categorical but rather a quantitative judgment (e.g., it would seem the magnitude of the error rate would be on a quantitative continuum, and the number of publications would also be quantitative). If any of these criteria are quantitative, then a further problem is what the cut score/demarcation ought to be—e.g., how much error is permissible for some theory or technique to be admissible (and does this vary across fields)? Let us now turn to an examination of each of the five criteria of the *Daubert* standard.

1. *That the evidence is relevant due to its external validity/generalizability*. A distinction should be made between *legal* relevance—which judges are well trained to make—and *scientific* relevance. Scientific relevance is a technical/methodological matter that is generally determined by the scientific study's external validity or generalizability,[105] i.e., the extent to which the study included sufficiently similar elements to the case at hand. For example, a study of the likelihood that suggestive questioning will cause false memories of sexually abusive touching that only sampled 6-year-old participants may have poor generalizability to a case that involves 15-year-olds but has better generalizability to a case involving 7-year-olds. The question of external validity requires a complex scientific judgment often on multiple dimensions (sample characteristics, similarities of the variables manipulated, similarities of the outcome variables, and so on) as every study has limited external validity.[106]
2. *That the testimony is the product of reliable principles and methods*. It is essential to note that for social scientists, the concept of reliability has a specific, technical meaning, i.e., consistency or replicability.[107] Reliability does not mean validity, although it can be mistakenly taken to.[108] Thus, a scale consistently indicating that an adult elephant weighs three pounds is reliable, but not valid. Reliability is important as it sets an upper bound to

[105] Lawrence W. Green & Russell E. Glasgow, *Evaluating the Relevance, Generalization, and Applicability of Research: Issues in External Validation and Translation Methodology*, 29 EVAL. & HEALTH PROFESSIONS 126, 127 (2006).

[106] Michael G. Findley, Kyosuke Kikuta & Michael Denly, *External Validity*, 24 ANN. REV. POL. SCI. 365, 368 (2021).

[107] Ralph O. Mueller & Thomas R. Knapp, *Reliability and Validity*, *in* THE REVIEWER'S GUIDE TO QUANTITATIVE METHODS IN THE SOCIAL SCIENCES 397, 397 (Gregory Hancock, Laura M. Stapleton, & Ralph O. Mueller eds., 2d ed. 2019).

[108] *Id.*

validity, i.e., an unreliable measure cannot be completely valid.[109] Validity, on the other hand, is more directly involved with truth—a test is valid to the extent that inferences based on its results are true.[110] Thus, it appears that the criterion probably means reliable as psychologists use this term, but it also means something more substantial, such as validity or accuracy. Let us now turn to an examination of *Daubert*'s four-part test of reliability construed as true or accurate.

i. *Whether the theory or technique in question could have been, or had been, tested with appropriate standards and controls.* First, there is a vast difference in science between whether a theory or technique had been tested and could have been tested, and it is problematic to conflate these. That a theory or technique has been tested (assuming these tests were favorable/corroborative[111]) carries far more weight than simply that these could, in principle, have been tested. The claim, "70-year-olds can make the NBA all-star team," can in principle be scientifically tested, but it is highly likely that these tests would show the claim to be false. It is problematic that this critical distinction is conflated in this criterion. It may, at times, also be important to understand why a particular claim is untestable or untested as it may have this status for diverse reasons, such as insufficient funding to conduct tests, institutional review boards have considered such testing unethical, or proponents seemingly think the truth of the claim does not require actual testing. For example, a study that randomly assigns subjects to either smoke or not smoke cigarettes to understand the causal connection between cigarette smoking and cancer has not been conducted due to its problematic ethics.

A second important issue this criterion is unfortunately silent on is the number of required tests—scientists generally place more stock on findings that have been replicated—either with direct replications (the same methods are used) or indirect replications (somewhat different methods are used).[112] As presently stated, the criterion is silent on whether a single unreplicated study or whether some number of direct and indirect studies must have occurred to show the result is replicable. The latter case is more likely to be considered scientific

[109] ANOL BHATTACHERJEE, SOCIAL SCIENCE RESEARCH: PRINCIPLES, METHODS, AND PRACTICES 56 (2d ed. 2012).

[110] *Id.* at 58.

[111] *See infra* notes 114–32 and accompanying text.

[112] *See* Seppo E. Iso-Ahola, *Replication and the Establishment of Scientific Truth*, FRONTIERS PSYCH., Sept. 2020 (discussing how exactly replicated experimental measurements that obtain the same result establish "bona fide" scientific truths).

knowledge. Third, this criterion is silent on the outcomes of these tests—it just requires that they be tested. However, this is insufficient as the outcomes of these tests are important. Questions then can be raised: Should these tests all be positive, or can the tests have somewhat mixed results, perhaps even including some falsificatory data? It would seem prima facie that the former ought to be admitted, the latter ought not to, and that the second case is unclear, forcing a more complex judgment to be made.

Finally, the important phrase, "with appropriate standards and controls,"[113] is also unclear as these methodological parameters are exactly what scientists frequently debate. Who or what is used to decide what is "appropriate" and what is not? Moreover, what exactly is meant by "standards," and what is meant by "controls"?

ii. *That the theory or technique had been subjected to publication and peer review*. This standard seems reasonable as publication in a peer-reviewed journal is generally considered a hurdle that sets the floor for a study's general quality.[114] However, there are well-known problems with peer review. For example, it is unclear who constitutes a peer—someone who does the same type of research (therefore, who also may be a direct competitor); or someone generally in the same discipline or subfield? The term "review" and what it should comprise of is equally perplexing. Would a "review" require a "peer" to just critically read the paper? Or would it require examining the raw data, rerunning all the analyses (increasingly possible and even recommended in the "open science" movement)?[115]

Almost every peer-reviewed journal follows a (mildly to extremely) different decision-making process to accept (and, therefore, publish) papers.[116] Some only use a single reviewer while others use five or more.[117] Some such reviewers are blind to the authors of the

---

[113] State v. Hull, 788 N.W.2d 91, 103 (Minn. 2010).

[114] Kelley D. Mayden, *Peer Review: Publication's Gold Standard*, 3 J. ADV. PRAC. ONCOLOGY 117, 117–18 (2012).

[115] Richard Smith, *Peer Review: A Flawed Process at the Heart of Science and Journals*, 99 J. ROYAL SOC'Y. MED. 178, 178 (2006); *see, e.g.*, Jacob T. Wittman & Brian H. Aukema, *A Guide and Toolbox to Replicability and Open Science in Entomology*, J. INSECT SCI., May 2020 (discussing the operationalization of open science principles and suggesting making research data and scripts available to the public).

[116] Smith, *supra* note 115, at 178.

[117] *See, e.g.*, *Publishing in a Scholarly Journal: Part Three, Peer Review*, AM. PSYCH. ASS'N, (Sept. 2021) [hereinafter APA, *Peer Review*], https://www.apa.org/pubs/journals/resources/publishing-tips/peer-review (stating that there can be two, three, four, or more peer reviewers for a submission); *About the Journal*, THE ASTROPHYSICAL J. [hereinafter *About the Astrophysical J.*],

paper, some are not.[118] Some may require all reviewers to agree that the paper deserves publication, some others may use one, and some a simple majority.[119]

Additionally, multiple cases of publishing misreported and/or fabricated data in peer-reviewed journals further demonstrate the problems with the effectiveness of peer review. For instance, recently, Francesca Gino, a behavioral scientist, had publications retracted because they were based on falsified data.[120] In one of the largest recent scandals in psychological research, Diederik Stapel, who had produced a large number of peer-reviewed research on stereotyping and discrimination, effectiveness of advertising, and circumstances in which people may prefer negative feedback to praise, was found to have fabricated data in at least 14 peer-reviewed publications.[121] William Fals Stewart, a well-known researcher on behavioral couples therapy for alcoholism and drug addiction (and in-charge of federally funded studies to the tune of $12 million), was found to have engaged in serious research misconduct including fabricating data for non-existent research subjects in multiple peer-reviewed publications.[122]

But the limitations of peer review are not just that these can fail to detect scientific fraud. In a classic study testing reviewer bias (also known as prestige bias), Peters and Ceci sought to republish 12 peer-reviewed studies by highly cited authors from prestigious institutions in the same journals they had originally been published.[123] The researchers only made a few key modifications to the papers they resubmitted, i.e., they substituted author names and institutions to lesser-known individuals and less prestigious institutions.[124] For three of

https://iopscience.iop.org/journal/0004-637X/page/about-the-journal#peer (last visited Apr. 9, 2025) (stating the default for submission review is a single peer reviewer).

[118] APA, *Peer Review*, *supra* note 117; *About the Astrophysical J.*, *supra* note 117.

[119] APA, *Peer Review*, *supra* note 117; *About the Astrophysical J.*, *supra* note 117.

[120] Cathleen O'Grady, *After Honesty Researcher's Retractions, Colleagues Expand Scrutiny of Her Work*, SCIENCE (Jul. 18, 2023, 4:50 PM), https://www.science.org/content/article/after-honesty-researcher-s-retractions-colleagues-expand-scrutiny-her-work.

[121] Peter Aldhous, *Psychologist Admits Faking Data in Dozens of Studies*, NEW SCIENTIST, (Oct. 29, 2024, 10:15 PM), https://www.newscientist.com/article/dn21118-psychologist-admits-faking-data-in-dozens-of-studies/.

[122] James Golden, Catherine M. Mazzotta & Kimberly Zittel-Barr, *Systemic Obstacles to Addressing Research Misconduct in Higher Education: A Case Study*, 21 J. ACAD. ETHICS 71, 73–75 (2023).

[123] Douglas P. Peters & Stephen J. Ceci, *Peer-Review Practices of Psychological Journals: The Fate of Published Articles, Submitted Again*, 5 BEHAVIORAL & BRAIN SCI. 187, 188 (1982).

[124] *Id.* at 189.

the 12 papers, the journal editors recognized that these had been published already and did not send them out for peer review.[125] However, shockingly, of the remaining nine, eight were rejected by the same editors that had previously accepted the original papers, now citing that they were of poor methodological quality.[126] A conclusion reached from this study was that in scientific peer review the methodological quality of an article, to a significant degree, can be influenced by the editors' or the reviewers' impressions of the status of the individual scientist and the prestige of their institutional affiliation rather than only by the methodological quality of the manuscript.[127]

Another common critique of the peer review system is that journals also avoid publishing negative results, i.e., results that indicate a certain technique and intervention does not work. This has become known as publication bias or the "file drawer problem"[128] which is so strong that negative results may not even be written up to be submitted to journals.[129] This problem also directly relates to the *Daubert* standard regarding error rates (discussed further below), which becomes distorted if studies with negative findings do not make their way into the scientific literature due to this bias.

Scientific journals vary in quality and prestige which is reflected by impact factors and manuscript rejection rates.[130] In exploring the quality of results reported in psychology journals, Bakker and Wicherts found that of 281 articles about 18% contained statistical results that had been incorrectly reported and, importantly, this was more common in low-impact journals (i.e., poorer quality) compared to high-impact journals.[131] Further, about 15% of the articles in journals, irrespective of their impact factor, "contained at least one statistical conclusion that proved, upon recalculation, to be incorrect."[132] Finally, in science, scholars also publish various kinds of content, including brief reports, book reviews, opinion pieces, edited books, book chapters, and introductions to special issues that have

---

[125] *Id.*

[126] *Id.* at 189–90.

[127] *Id.* at 192.

[128] Robert Rosenthal, *The "File Drawer Problem" and Tolerance for Null Results*, 86 PSYCH. BULL. 638, 638 (1979).

[129] Smith, *supra* note 115, at 180.

[130] Louis C. Buffardi & Julia A. Nichols, *Citation Impact, Acceptance Rate, and ABA Journals*, 36 AM. PSYCH. 1453, 1453–56 (1981).

[131] Marjan Bakker & Jelte M. Wicherts, *The (Mis)reporting of Statistical Results in Psychology Journals*, 43 BEHAV. RSCH. METHODS 666, 666 (2011).

[132] *Id.*

varying peer review processes. Should all of these publications be considered as having met this criterion of the *Daubert* standard?

iii. *That the theory or technique has a known or potential error rate*. Most scientific theories do not have a known error rate.[133] Moreover, questions can be raised how, even in principle, does one accurately determine a *potential* error rate. Every technique has a potential error rate of 100% if it is sufficiently poorly implemented or misapplied. A literature review using the search phrase "potential error rate" in the American Psychological Association's PsycInfo, a database of 2,319 journals and 5 million peer-reviewed articles, revealed only 58 hits from peer reviewed journals, with none of these providing a potential error rate of a theory or technique (conducted March 2, 2025).[134] Thus, one can scour the research literature without encountering the perplexing phrase "potential error rate." However, similarly, most theories and techniques do not even possess a "known error rate." What is the known error rate for Newtonian mechanics? For evolutionary theory? For the dopamine theory of schizophrenia? For classical conditioning? What would be the unit of measurement of this error rate—the percentage of variance unaccounted for, the number of studies not supporting the theory, the number of studies supporting the theory, or some other metric?

This criterion may be influenced by the fact that a minority of techniques in science have error rates, for example, cancer screens.[135] However, even in this case the problem is that there are not one but two distinct types of error rates. Error rates in cancer screens can be captured by either false positives (an indication that a patient has cancer when she does not) or false negatives (an indication that the patient does not have cancer when she does). Complicating this picture further is that there is also a tradeoff between these two error rates—it is possible to decrease one error rate by simply allowing an increase in the other. For example, a way to minimize false negative rates is to have the screen indicate that everyone has cancer as this would

---

[133] Itiel E. Dror & Nicholas Scurich, *(Mis)Use of Scientific Measurements in Forensic Science*, 2 FORENSIC SCI. INT'L. SYNERGY 333, 333 (2020).

[134] Am. Psych. Ass'n: PsycInfo Database, EBSCO HOST, https://www.ebsco.com/products/research-databases/psychology-resources-apa (search "potential error rate" and filter by "Scholarly (Peer Reviewed) Journals") (on file with author).

[135] Dror & Scurich, *supra* note 133, at 333; *see* Minetta C. Liu, Geoffrey R. Oxnard, Eric A. Klein, David Smith, Donald Richards, Timothy J. Yeatman, Allen L. Cohn et al., *Sensitive and Specific Multi-Cancer Detection and Localization Using Methylation Signatures in Cell-Free DNA*, 31 ANNALS ONCOLOGY 745, 754 (2020) (comparing cancer error-rates across screening methodologies).

produce a false negative rate of zero because mathematically it is then impossible for there to be a false negative finding. Thus, it can be simplistic to assume, even in the minority of cases where there are error rates, that there is only one error rate.

Finally, even if one can make sense of this phrase and one could determine a useful metric to capture it, there is a further question that needs to be addressed: What error rate is permitted for admissibility? Does this error rate vary from technique to technique, discipline to discipline, or from legal case to legal case? For example, it seems in principle that a theory in a newly emerging field could have a higher error rate—which faithfully represents the cutting edge of science at that time. However, should error rates of competing theories or techniques also be considered, and only the theory or technique with the lowest error rate be admissible? For example, all microscopes result in errors due to limitations in their magnification and distortion. In general, the more expensive the microscope, the better the resolution and lower the distortion, and thus the lower the error rate. However, a practical decision must be made related to margin utility—due to opportunity costs associated with any expenditure, what price results in a microscope with a "good enough" error rate in the particular circumstances?

iv. *The theory or technique had attracted widespread acceptance within the scientific community of that field.* We have already discussed the problems with this criterion in our discussion of the *Frye* standard above.[136]

E. Current Practices Across States

Although the *Daubert* Standard has been established as the admissibility criteria for expert evidence in federal courts, the situation in state courts is more unsettled.[137] As of March 2022, eight states were using the *Frye* standard, thirty-five were using the *Daubert* standard, and seven states had formulated their own admissibility standard.[138] Some states have yet to accept the *Daubert* criteria in their entirety; others view it as instructive or as consistent with their state tests but not necessarily binding.[139] For instance, the State of Indiana has prima facie accepted the *Daubert*

[136] *See* discussion *supra* Sections I.A–B.

[137] Bernstein & Jackson, *supra* note 73, at 356.

[138] *Admissibility of Expert Testimony in all 50 States*, MATTHEISEN, WICKERT & LEHRER S.C. (Oct. 29, 2024, 10:30 PM), https://www.mwl-law.com/wp-content/uploads/2018/02/ADMISSIBILITY-OF-EXPERT-TESTIMONY.pdf.

[139] Bernstein & Jackson, *supra* note 73, at 356.

standard. However, Rule 702 of the Indiana Rules of Evidence (governing expert witnesses) does not prescribe any specific tests that judges must consider to satisfy evidentiary requirements regarding expert witnesses.[140] The State of Colorado has taken a similar stance: though it has accepted the *Daubert* standard, state courts are not bound by it.[141] The State of Minnesota has incorporated a hybrid standard that is a combination of the *Frye* standard and *State v. Mack*;[142] according to this standard, expert testimony can be admitted if it has utilized a technique that is generally accepted in the scientific community and if the testing underlying expert evidence has been done properly.[143] The standard is silent on what constitutes "proper" testing.

II. FURTHER CONSIDERATIONS ON EVOLVING VIEWS OF PROPER DEFINITION

An important meta-issue must also be addressed at this point which concerns the assumptions about the definition of science itself—specifically, whether it is always possible to define a term such as science by providing necessary and sufficient conditions. Clearly, this is the assumption in both the *Frye* and *Daubert* standards.

A. The Difficulties of Definition

What is a good definition of a term like science or knowledge? That is, what in principle should a satisfactory definition of some construct like science even look like? What can one expect from such a definition? Are there different kinds of definitions that vary by what is to be defined? Also note the reflexive nature of this question—one is asking for a definition of definition. This problem has concerned scholars, particularly philosophers, who have made various distinctions.[144] For example, *lexical* definitions simply track what language users commonly mean when they use the phrase.[145] Still, in doing so, the essential properties of the construct need to be identified. For example, the definition of "bachelor" as an unmarried male adult provides the necessary and sufficient properties. In this definition,

---

[140] *See* IND. R. EVID 702.

[141] People v. Shreck, 22 P.3d 68, 70 (Colo. 2001) (finding that trial court judges have the discretion to consider *Daubert* factors when evaluating the reliability of expert testimony but are not bound by it).

[142] State v. Mack, 292 N.W.2d 764, 771–72 (Minn. 1980).

[143] *See* State v. Harvey, 932 N.W.2d 792, 806 n.11 (Minn. 2019) (citing State v. MacLennan, 702 N.W.2d 219, 230 (Minn. 2005)) (describing Minnesota's two-pronged "*Frye–Mack*" standard).

[144] Anil Gupta & Stephen Mackereth, *Definitions*, STANFORD ENCYC. PHIL., https://plato.stanford.edu/archives/fall2023/entries/definitions/ (Sept. 13, 2023).

[145] Brian Duignan, *Definition*, ENCYC. BRITANNICA (Jan. 23, 2023), https://www.britannica.com/topic/definition.

sometimes called a "conceptual explication,"[146] the word's necessary and sufficient properties are specified. The standard view of definition is that a definition should provide the necessary and sufficient properties of the word to be defined. It is exemplified by the logician Gottlob Frege's definition of "definition" in his *Foundations of Arithmetic*:

> A definition of a concept . . . must be complete; it must unambiguously determine, as regards any object, whether or not it falls under the concept . . . . [T]he concept must have a sharp boundary. . . . [A] concept that is not sharply defined is wrongly termed a concept. Such quasi-conceptual constructions cannot be recognized as concepts by logic . . . . The law of excluded middle is really just another form of the requirement that the concept should have a sharp boundary.[147]

However, subsequently, the influential philosopher Ludwig Wittgenstein suggested that many concepts do not have essential properties.[148] He thought the process of definition might involve elucidating a complex system of similarities and differences, which he called "family resemblances." He used the concept of games to illustrate this sort of definition (sometimes also called "fuzzy concepts"):[149]

> Consider for example the proceedings that we call "games". I mean board-games, card-games, ball-games, Olympic games, and so on. What is common to them all?—Don't say: 'There *must* be something common, or they would not be called "games"'—but *look and see* whether there is anything common to all.—For if you look at them you will not see something that is common to *all*, but similarities, relationships, and a whole series of them at that. . . . Look for example at board-games, with their multifarious relationships. Now pass to card-games; here you find many correspondences with the first group, but many common features drop out, and others appear. When we pass next to ball-games, much that is common is retained, but much is lost.—Are they all 'amusing'? Compare chess with noughts and crosses. Or is there always winning and losing, or competition between players? Think of patience. . . .

---

[146] Michael Bishop, *The Possibility of Conceptual Clarity in Philosophy*, 29 AM. PHIL. Q. 267, 268, 276 n.1 (1992).

[147] GOTTLOB FREGE, GRUNDGESETZE DER ARITHMETIK (1893), *reprinted in* TRANSLATIONS FROM THE PHILOSOPHICAL WRITINGS OF GOTTLOB FREGE 139, 139 (Peter Geach & Max Black eds., P. T. Geach trans., Rowman & Littlefield 3d ed. 1980).

[148] LUDWIG WITTGENSTEIN, FAMILY RESEMBLANCES (1953), *reprinted in* FUZZY GRAMMAR: A READER 41, 41 (Bas Aarts, David Denison, Evelien Keizer & Gergana Popova eds., 2004) [hereinafter WITTGENSTEIN, FAMILY RESEMBLANCES].

[149] *Id.* at 41; *see, e.g.*, Susan Byrne, Ludwig Wittgenstein's Legacy to Cognitive Psychology: Concepts as Participatory (Oct. 30, 2012) (Ph.D. dissertation, Maynooth University), https://mural.maynoothuniversity.ie/id/eprint/4395/1/PhD_SByrne_FINAL_23April2013_submitted.pdf.

> I can think of no better expression to characterize these similarities than "family resemblances" . . . . And I shall say: 'games' form a family.[150]

Therefore, a preliminary question arises: what are the assumptions about what a definition of science in an admissibility standard should look like—a lexical one à la Frege, i.e., one that provides necessary and sufficient conditions, or a "family resemblance" à la Wittgenstein, i.e., one that illustrates family resemblances, or some other kind? As we will discuss below, it is fair to say that despite numerous attempts, philosophers of science over the past century have failed to provide a lexical definition of science that provides the necessary or sufficient properties of science.

B. Prima Facie Considerations that Science May Not Share a Common Characteristic or Characteristics

The next issue concerns the likelihood that something as large, multifaceted, complex, and often as unsettled as science is likely to share a delimited set of commonalities across its diverse branches. This point leads to the second preliminary question: Are there antecedent reasons to believe that science can be neatly captured by a relatively simple enumeration of its essential properties? There are several salient reasons to suggest not, and some influential philosophers of science have argued that there is no single scientific method.[151]

Some of the difficulties of defining science can be illustrated by examining three iconic scientists, Newton, Darwin, and Einstein, as it seems reasonable to—as Wittgenstein advises above—look at the methods scientists use. Even a cursory glance reveals that each of these scientists used vastly different methods. In his *Opticks*, Newton engaged in what might be called demonstrations as opposed to experiments (involving random assignment and control conditions).[152] Newton demonstrated that white light was composed of many different colors by using a prism to refract light into seven distinct color components.[153] To show it was not the prism that was producing these colors, Newton then used a second prism to show that he could then recombine these colors back into white light, thus proving that white light consisted of seven individual color components.[154]

On the other hand, Darwin, in his classic *On the Origin of Species by Means of Natural Selection*, did not rely on such demonstrations or experiments but instead

---

[150] WITTGENSTEIN, FAMILY RESEMBLANCES, *supra* note 148, at 41.

[151] *See* BARRY GOWER, SCIENTIFIC METHOD: AN HISTORICAL AND PHILOSOPHICAL INTRODUCTION 247 (1997); PAUL FEYERABEND, AGAINST METHOD 7 (4th ed. 2010).

[152] *See generally* ISAAC NEWTON, OPTICKS, OR, A TREATISE OF THE REFLECTIONS, REFRACTIONS, INFLECTIONS & COLOURS OF LIGHT (Dover Publ'ns, Inc. 1979) (1730).

[153] *Id.*

[154] *Id.* at 77–82.

engaged in a multitude of uncontrolled but strategically chosen observations.[155] Darwin undertook several long voyages to South America, where he made numerous observations of certain natural phenomena such as variations in beak size and shape of Galapagos finches that differed from island to island.[156] Darwin also observed that different sizes and types of food were available that all required using the finches' beaks to break down. From these observations, Darwin then made the inference that islands with large-sized food benefitted the finches with larger beaks, who could eat more food, survive, and reproduce at a higher rate.[157] Over time and across islands, Darwin suggested that the available food source indicated which beak traits would be observed, leading to differentiation in Galapagos finches, which addressed Darwin's ultimate scientific question regarding the origins of speciation.[158] Finally, Einstein did not engage in revealing demonstrations like Newton or telling observations like Darwin. In fact, he collected no data during his entire life. Einstein was a theoretical physicist who relied on thought experiments. For example, he described one of his early thought experiments:

> [A] paradox upon which I had already hit at the age of sixteen: If I pursue a beam of light with the velocity *c* (velocity of light in a vacuum), I should observe such a beam of light as a spatially oscillatory electromagnetic field at rest. However, there seems to be no such thing, whether on the basis of experience or according to Maxwell's equations. From the very beginning it appeared to me intuitively clear that, judged from the standpoint of such an observer, everything would have to happen according to the same laws as for an observer who, relative to the earth, was at rest. For how, otherwise, should the first observer know, i.e., be able to determine, that he is in a state of fast, uniform motion? One sees that in this paradox the germ of the special relativity theory is already contained.[159]

Even the analysis of the variability of scientific methods taken from these three scientists must be more complex. The philosopher of science Thomas Kuhn suggested that science has a complex and multifaceted micro-community structure

---

[155] *See generally* CHARLES DARWIN, ON THE ORIGIN OF SPECIES BY MEANS OF NATURAL SELECTION (Harvard Univ. Press 1964) (1859).

[156] *See* CHARLES DARWIN, JOURNAL OF RESEARCHES INTO THE NATURAL HISTORY & GEOLOGY OF THE COUNTRIES VISITED DURING THE VOYAGE ROUND THE WORLD OF H.M.S. BEAGLE ROUND THE WORLD, UNDER THE COMMAND OF CAPTAIN FITZ ROY, R.A. 404–45 (2d ed. 1870).

[157] *See id.*

[158] *See id. See generally* Francis Darwin, *The Botanical Work of Darwin*, 13 ANNALS BOTANY ix (1899).

[159] PAUL ARTHUR SCHILPP, ALBERT EINSTEIN: PHILOSOPHER-SCIENTIST 53 (Tudor Publishing Co. 1979) (1949).

primarily due to scientific specialization.[160] One can readily see this micro-community structure in medicine—at one time, due to limited knowledge availability, there were just general physicians.[161] However, then specialties emerged, such as oncologists, and then further subspecialties emerged, such as radiation oncologists, pediatric radiation oncologists, and so on. Kuhn argues that this scientific specialization is essential for increasing the problem-solving effectiveness of science—more specialization permits a more in-depth treatment of particular scientific problems while, at the same time, the general scientific enterprise increases in breadth.[162] However, there is no guarantee that all these micro-communities of science will share the same methods—as seen in radiation oncology and surgical oncology.

III. THE VARIED CHARACTERIZATIONS OF SCIENCE FROM SOME MAJOR PHILOSOPHERS OF SCIENCE

Thus far, we have only critically examined criteria proposed in the *Frye* and *Daubert* standards. However, a more fundamental question emerges: Are these the proper criteria to distinguish science from nonscience and pseudoscience, or good science from poor science? In 1997, a survey of 176 members of the Philosophy of Science Association in the United States concluded that 80% believed that no satisfactory demarcation criteria have been found yet.[163] Next is a brief review of a more comprehensive set of such criteria gained from the criteria by the foremost philosophers of science.

A. Wittgenstein/Logical Positivists

The logical positivists had similar general motivations as the courts in that they wanted to exclude from consideration claims that might appear to have substance but did not. In the case of logical positivists, they tried to exclude metaphysical sentences that they thought were a "bewitchment of our intelligence" by seeming meaningful but were not.[164] They developed and used the verifiability criterion of meaning, which stated that meaningful sentences were either analytic sentences (e.g., "Triangles have three sides") or sentences that are, in principle, empirically verifiable.[165]

---

[160] THOMAS KUHN, THE ROAD SINCE STRUCTURE 97–99 (James Conant & John Haugeland eds., 2000).

[161] *Id.*

[162] *Id.*

[163] Brian J. Alters, *Whose Nature of Science?*, 34 J. RSCH. IN SCI. TEACHING 39, 44, 48 (1997).

[164] WITTGENSTEIN, PHILOSOPHICAL INVESTIGATIONS, *supra* note 72, at 47.

[165] *Id.* at 40–46.

However, John Passmore in the 1960s reported, "Logical positivism, then, is dead, or as dead as a philosophical movement ever becomes,"[166] largely due to three problems: (1) logical positivists could develop a version of the verifiability criterion that satisfied itself—the verifiability criterion classified itself as meaningless; (2) a general failure to find a truth-preserving and ampliative (roughly content-increasing) account of induction; and (3) cogent critiques of its analytic/synthetic distinction by the eminent philosopher Quine.[167] Thus, although this philosophy was quite influential for most of the first half of the 20th century, it provides little positive help for a satisfactory admissibility criterion. Its principal contribution may be negative—the answer may not be found in reliance on induction or characterizing science using empirical criteria of meaningfulness.

B. Popper

Sir Karl Popper thought that the logical basis of science depended not on induction but rather on the valid logical inference rule known as *modus tollens*, which reasons in the following manner: (1) if theory, then observation statement; (2) not observation statement; and (3) therefore, not theory.[168]

According to Popper, every scientific theory should divide the set of all statements derivable from it into two subsets. One set contains observation statements that are logically consistent with the theory.[169] This set is uninteresting from a scientific point of view. However, for every scientific theory, the complementary set—the set of observable states of affairs that are logically inconsistent with it should be nonempty—is the observable states of affairs that are potential falsifiers of the theory.[170] Thus, for Popper, a necessary but not sufficient criterion for a theory to be scientific is that it has a nonempty set of observable states of affairs that it rules out.

According to Popper, however, all theories that meet this criterion are not equal because the set of potential falsifiers, known as the theory's empirical content, would vary in size. Two criteria can be used to evaluate the size of the empirical content of a theory: (1) its *level of universality* (e.g., "All men are tall," has more empirical content than "North American men are tall," because the former rules out more states

---

[166] John Passmore, *Logical Positivism*, *in* 5 THE ENCYCLOPEDIA OF PHILOSOPHY 52, 56 (Paul Edwards ed., 1967).

[167] Willard V. O. Quine, *Two Dogmas of Empiricism*, 60 PHIL. REV. 20 (1951), *reprinted in* WILLARD V. QUINE, TWO DOGMAS OF EMPIRICISM 31, 51 (1976) [hereinafter Quine, *Two Dogmas of Empiricism*].

[168] POPPER, *supra* note 8, at xvi; *Modus Tollens*, MERRIAM-WEBSTER, https://www.merriam-webster.com/dictionary/modus%20tollens (last visited Apr. 9, 2025).

[169] Stephen Thornton, *Karl Popper*, STANFORD ENCYC. PHIL., https://plato.stanford.edu/archives/sum2020/entries/popper/ (Aug. 7, 2018).

[170] *Id.*

of affairs) and (2) the theory's *degree of precision* ("Men are 2.5 inches taller than women" has a more extensive set of falsifiers than "Men are somewhat taller than women")[171]

Popper thought that another necessary criterion for science is the severe testing of these theories that are in principle falsifiable. Such testing consists of taking a theory and then efficiently and ardently attempting to see if one of these potential falsifiers is obtained by using what he calls "severe testing." A severe test examines the least plausible, riskiest claims of a theory, i.e., the claims that rule out most states of affairs.[172]

Then, of course, good scientific theories survive these severe tests. Many astrological claims that are falsifiable in principle have been tested, and because they have not survived these tests, they are not part of science—instead these, as false statements, are external to science.[173] Popper suggests that a theory's evidence is calculated by the number and the severity of tests it has so far survived where survivors are considered as *corroborated* (but not confirmed) and as having *verisimilitude* or truth likeness.[174] However, he is quick to note that science is dynamic in that the next severe test might show that the theory is false, surviving all tests to date is no guarantee that a theory will survive the next test.

However, Popper's account of science has been criticized on two major grounds: (1) it is descriptively inaccurate: actual scientists are simply not doing what Popper has described—scientists are not trying to falsify their theories[175] and (2) Popper failed to account that it is not a theory alone that is used to deduce observational consequences, rather a large number of auxiliary hypotheses are needed to deduce the observational consequence.[176] Known as the Duhem–Quine problem, this criticism means that in actual scientific research, the arrows of *modus tollens* are not directed at the theory alone—but that even when observations inconsistent with the theory are found, this finding, instead of falsifying the theory (as Popper proposed), only indicates that the theory *or any one of the auxiliary hypotheses* used to deduce the observation statement is false.[177] Philip Kitcher, for example, thought that pseudosciences suffered from "too cozy a relationship with auxiliary

---

[171] *Id.*

[172] *Id.*

[173] Geoffrey Dean, *Does Astrology Need To Be True? Part 1: A Look at the Real Thing*, 11 SKEPTICAL INQUIRER 166, 166, 168–69 (1986).

[174] Thornton, *supra* note 169.

[175] THOMAS S. KUHN, *Logic of Discovery or Psychology of Research?*, *in* 4 CRITICISM AND THE GROWTH OF KNOWLEDGE 1, 3–4 (Imre Lakatos & Alan Musgrave eds., 1970).

[176] *Id.* at 4–5.

[177] Piotr K. Szałek, *The Duhem–Quine Thesis Reconsidered*, 62 STUDIES IN LOGIC, GRAMMAR, AND RHETORIC 73, 76–77 (2020).

hypotheses."[178] That is, its numerous prediction failures are blamed on these rather than the central theory (e.g., the presence of skeptics—known as "negative psi vibrations" interferes with clairvoyance).

Again, there is little left from this important philosopher of science upon which to base an admissibility criterion that elucidates a delimited set of necessary and sufficient conditions to define science. Even if a theory has many potential falsifiers, and even if the scientist engages in severe testing, all this is for naught if the scientist is logically free to save the theory from apparent falsification by blaming an auxiliary hypothesis. Perhaps even more devastating—if Popper is inaccurate about the behavior of actual scientists in that they are not attempting to falsify their theories, then his views do not seem to have much use for a sound admissibility criterion.

In *Bitler v. A.O. Smith Corp.*, the U.S. Appellate Court delved into whether experts were required to test their theory before testifying regarding it.[179] In ruling on the issue, the court recognized Popper's emphasis on falsifiability in order to determine the scientific nature of theories.[180] However, surprisingly, although the court acknowledged the importance of falsifiability, it claimed that this concept was limited to theories that purported to explain causal relations in naturally occurring phenomena.[181] The court therefore observed that the principle of falsifiability was instructive but not mandatory. However, the court's reasoning is problematic because falsifiability is not limited to causal claims—simple descriptive claims like, "All swans are white" are falsifiable as the observation of a nonwhite swan would falsify the claim.

Regarding the *Daubert* standard, in *Butler v. Union Carbide Corp.*, Georgia's Appellate Court noted that, according to Popper, the key criterion to determine scientific status of a theory was its falsifiability.[182] The court further observed that courts will find application of *Daubert* "difficult" if they treat testability as an "optional factor" because the remaining factors "all presuppose testability."[183] Because courts appeared to treat falsifiability as a prerequisite, in practice, the *Daubert* standard was interpreted as entirely consistent with Popper's logical criterion.

C. Kuhn

Thomas Kuhn also did not provide any clear, necessary, and sufficient criteria to demarcate science from pseudoscience. He also muddied these waters by indicating that there were two types of science, normal and revolutionary, and describing

---

[178] PHILIP KITCHER, ABUSING SCIENCE: THE CASE AGAINST CREATIONISM 48 (2d ed. 1982).

[179] Bitler v. A.O. Smith Corp., 400 F.3d 1227, 1235 (10th Cir. 2005).

[180] *Id.* at 1235–36.

[181] *Id.*

[182] Butler v. Union Carbide Corp., 712 S.E.2d 537, 549 (Ga. Ct. App. 2011).

[183] *Id.*

science not in logical or epistemological terms but rather in sociological and historical ones. Kuhn's account of science sees it as less rule bound and methodical.[184]

For Kuhn, the key aspect of normal science is puzzle-solving. Science emerges from a situation in which there is a series of competing schools of thought, all of which still need to solve any puzzle. According to Kuhn, this *preparadigmatic stage* is characterized by "frequent and deep debates over legitimate methods, problems, and standards of solution, though these serve rather to define schools than to produce agreement."[185] *Normal* science, or *paradigmatic* science, emerges when some problem is solved, and this solution is recognized as a notable achievement by others. Kuhn stated that scholars, who can finally be regarded as scientists, view the method, constructs, and claims of this solution to a problem as a model with great heuristic promise—the *paradigm* in his terms—for scientists to adopt to try to solve other puzzles.[186] Contra Popper, Kuhn suggests that instead of attempting to falsify their theories, good scientists attempt to opportunistically employ these to see if they can solve further puzzles.[187]

However, Kuhn then suggests that eventually in these puzzle-solving attempts, anomalies emerge because the new paradigm is not able to solve all the remaining problems in the field.[188] Eventually, these anomalies play a key role in *scientific revolutions*. Thus, for Kuhn, anomalies can result in a crisis in which there is a "blurring of a paradigm and the consequent loosening of the rules for normal research."[189] Due to these controversies, the field is now similar to the pre-paradigm period, except the differences are "smaller and more clearly defined."[190]

Sometimes, according to Kuhn, a crisis may end with the emergence of a new paradigm, which solves a puzzle refractory to a solution by the old paradigm.[191] Contra Popper, scientists do not abandon their theories because of prediction failures, but instead they wait until a better account/paradigm is available. Scientists then experience what Kuhn calls a "gestalt switch" because "[w]hen paradigms change, the world itself changes with them."[192] This new paradigm results in new ways of seeing old things, new puzzles, new exemplars, and perhaps even innovations in instrumentation and research methods. The old and new paradigms then compete for the allegiance of the scientific community.

---

[184] Thomas Nickles, *Kuhn, Historical Philosophy of Science, and Case-Based Reasoning*, 6 CONFIGURATIONS 51, 68 (1998).

[185] KUHN, STRUCTURE OF SCIENTIFIC REVOLUTIONS, *supra* note 71, at 48.

[186] *Id.*

[187] *Id.*

[188] *Id.* at 84.

[189] *Id.*

[190] *Id.*

[191] *Id.*

[192] *Id.* at 111, 118.

However, Kuhn's account of science has been criticized in numerous ways, including that: (1) he fails to consistently clearly define his fundamental construct of "paradigm"; (2) he maintains that paradigms are incommensurable and this is inconsistent with the view that there is scientific progress; and (3) his account is vague in critical ways.[193] For instance, Kuhn's account does not address what counts as a puzzle solution, nor how many or how serious the puzzle solving failures need to be to prompt a scientific revolution. Thus, Kuhn also provides no guide for developing clear and valid admissibility criteria.

Despite the above criticisms, Kuhn's work has been recognized in legal writings. For instance, Imwinkelried observed that *Frye* was problematic because it relied on general acceptance of a theory.[194] However, Imwinkelried correctly noted the popularity of a theory does not guarantee its scientific validity because, as Kuhn documented, the history of science is replete with examples of popular paradigms that were later made obsolete by subsequent scientific revolutions and their associated novel empirical findings.[195] Further, Mueller explored whether *Daubert* had a defective view of science and reiterated what philosophers of science have struggled with for decades: Science is not wholly a matter of logic.[196] Mueller discussed Kuhn's "competing paradigms" analysis to exemplify this. Scientific reliability is not seen as an invariant and objective binary (present or absent) but is, in his view, a fluid and relative judgment.[197]

D. Lakatos

Imre Lakatos, a student of Popper's, advocated what he took to be a "sophisticated methodological falsificationism" in which science is judged by several criteria related to how the scientific theory changes over time—especially how it changes when faced with anomalies.[198] Lakatos was most interested in distinguishing good science from bad science and to do so, he offered some novel technical terminology: A research program consists of a "hard core" (theory), "protective belt" (auxiliary hypotheses) and a "positive heuristic" and "negative heuristic."

---

[193] *See generally* Turkan Firinci Orman, *"Paradigm" as a Central Concept in Thomas Kuhn's Thought*, 10 INT'L J. HUMANITIES SOC. SCIENCES 47, 48–51 (2016).

[194] Edward J. Imwinkelried, *Serendipitous Timing: The Coincidental Emergence of the New Brain Science and the Advent of an Epistemological Approach to Determining the Admissibility of Expert Testimony*, 62 MERCER L. REV. 959, 963 (2011).

[195] *Id.* at 966. *See generally* KUHN, STRUCTURE OF SCIENTIFIC REVOLUTIONS, *supra* note 71.

[196] Christopher. B. Mueller, Daubert *Asks the Right Questions: Now Appellate Courts Should Help Find the Right Answers*, 33 SETON HALL L. REV. 987, 1007 (2003).

[197] *Id.* at 1008.

[198] Imre Lakatos, *Falsification and the Methodology of Scientific Research Programmes*, *in* 4 CRITICISM AND THE GROWTH OF KNOWLEDGE: PROCEEDINGS OF THE INTERNATIONAL COLLOQUIUM IN THE PHILOSOPHY OF SCIENCE, LONDON, 1965, at 122 (Imre Lakatos & Alan Musgrave eds., 1970) [hereinafter Lakatos, *Falsification*].

- "Hard core" (theory): This is a set of theoretical assumptions that are never abandoned even when anomalies are observed. It is, in principle, falsifiable, but never allowed by scientists to be interpreted as falsified.
- "Protective belt" (auxiliary hypotheses): These are statements and theories that are smaller in scope than the theory under test that are used in combination with the hard core to make experimental predictions. In addition, the auxiliary hypotheses are used to explain the evidence that could, in principle, falsify the hard core. In a situation that results in a prediction failure, the arrows of *modus tollens* are always directed toward these auxiliary hypotheses because these are seen as expendable.
- "Positive heuristic" and "negative heuristic": This is a "powerful problem-solving machinery, which . . . digests anomalies and . . . turns them into positive evidence."[199] The negative heuristic forbids scientists to question or criticize the hard core of the research program. On the other hand, the positive heuristic consists of a set of "suggestions or hints" on how to change or develop the "refutable variants" of the research program, and "how to modify [and] sophisticate, the 'refutable' protective belt."[200]

Lakatos's unit of analysis for evaluating science was how a research program develops over time. In appraising this progression, Lakatos says attributing prediction failures to problems with an auxiliary hypothesis is problematic, which he calls "degenerating."[201] Or this progression can be appropriate—his term was "progressive." To be progressive, the research program must meet three criteria. First, the modified auxiliary hypotheses need to make novel predictions (i.e., predict empirical states of affairs that the earlier research program did not). If it meets this condition, it can be said to be "theoretically progressive."[202] Second, at least one of these new predictions has to be observed to be true. The research program can then be said to be "empirically progressive."[203] Finally, successor theories in a research program can explain why their predecessors worked as well as they did.[204]

---

[199] 1 IMRE LAKATOS, THE METHODOLOGY OF SCIENTIFIC RESEARCH PROGRAMMES: PHILOSOPHICAL PAPERS 4 (John Worrall & Gregory Currie eds., 1978).

[200] Lakatos, *Falsification*, *supra* note 198, at 132–35.

[201] *Id.* at 118.

[202] *Id.*

[203] *Id.*

[204] *See id.* (reasoning that progress is measured by the degree to which progressive research programs lead to the discovery of novel facts).

Coate and Fischer cited Lakatos and his philosophy in the context of demarcating science under the *Daubert* standard.[205] These scholars argue that what the legal community requires is a tool to determine valid science in real-time because no one can wait for a degenerative theory to die. While this concern is reasonable, it misses the point that Lakatos still provides tools for appraising science; for example, a research program immediately can be seen to be degenerating if the modifications make no novel predictions. Finally, other aspects of science also can involve waiting—e.g., for a theory that is in principle falsifiable to then undergo actual testing.

E. Feyerabend

Paul Feyerabend's central claim is that if one carefully reviews the history of science, there is no one scientific method; and even stronger, whatever methodological rules philosophers of science say is essential to science, the history of science also reveals that in successful episodes of science, scientists have violated these rules.[206] For example, contra Popper and Lakatos, successful scientists ignored anomalous—potentially falsifying—observations (instead of using these to falsify their favored theories). Contra the logical positivists, scientists let external forces such as politics and metaphysics influence their interpretation of data (instead of just relying upon the raw observations). They chose to favor a theory that had much less empirical support than a rival, and so on.[207] But Feyerabend also states that even if the history of science did not reveal successful scientists violating all these supposed norms of science, it still would be good if they did.[208]

Feyerabend agrees with Popper that criticism is essential in science but maximizing criticism means that potential problems ought to be considered no matter where they originate—even from those outside of science, such as journalists, lawyers, astrologers, theologians, or indigenous accounts.[209] So, for Feyerabend, "anything goes"—scientists should simply be opportunistic puzzle solvers because there is no scientific method. Feyerabend states: "Science is an essentially anarchic enterprise: theoretical anarchism is more humanitarian and more likely to encourage progress than its law-and-order alternatives."[210] Feyerabend's account cannot help develop an account of what is scientific because he denies that such a distinction is even possible or if it is even helpful.

Thus, in this brief exposition of five major philosophies of science, one can unfortunately see no consensual criteria of what constitutes science. This issue has

---

[205] Malcolm B. Coate & Jeffrey H. Fischer, Daubert, *Science, and Modern Game Theory: Implications for Merger Analysis*, 20 SUP. CT. ECON. REV., 125, 137–38 (2012).

[206] FEYERABEND, *supra* note 151, at 7.

[207] *Id.*

[208] *See id.*

[209] *See id.* at 11–12.

[210] *Id.* at 1.

been acknowledged in legal writings[211] which have observed that the *Daubert* standard, in particular, represents a confused philosophy of science or an unsophisticated view of science, which is perhaps understandable due to the disagreement among philosophers of science themselves.[212] In fact, this sampling shows that the major philosophers of science each propose quite distinct but important ways of evaluating science. For instance, Schwartz observed that the *Daubert* standard must essentially choose between the various claims by philosophers of science—e.g., either follow Popper and therefore exclude theories that have not been severely tested, or follow Feyerabend's views that scientists are prudent to sometimes include (untested) hypotheses in the spirit of "anything goes" if these help the scientist to solve a scientific problem.[213]

It is beyond the scope of this paper to review all criteria that have been proposed for evaluating science, but this sampling also shows that evaluating science is a complex and multidimensional process. Below we propose a multi-factorial model.

IV. PROPOSED CRITERIA FOR ADMISSIBILITY

We propose a polythetic model of 39 criteria to evaluate candidates for scientific evidence. This proposal clearly is not an enumeration of a small set of necessary and sufficient conditions of science. Rather, this list is based upon Wittgenstein's conceptualization of definition through a complex, interrelated web of family resemblances.[214] We also note that although the *Frye* standard started with a single dimension, the *Daubert* standard increased this to two superordinate criteria, with the second having four additional subdimensions.[215] While the growth of the number of dimensions can make the ultimate admissibility judgment on the part of the judge more difficult, this difficulty can be justified if this larger set more accurately tracks how candidates for proper science should be evaluated and demarcated. Note that, consistent with the view of family resemblances, none of these criteria, either singly or jointly, are necessary or sufficient.

An analogy might be illustrative here. Consider deciding if a high-rise building is safe for human occupation. Certainly, it would be convenient if this was a simple unidimensional judgment—if, for example, the foundation was sound, then by this single criterion, the building could be judged as safe. The next level of convenience

---

[211] Note, *Admitting Doubt: A New Standard for Scientific Evidence*, 123 HARV. L. REV. 2021, 2026–27 (2010); Caudill & Redding, *supra* note 20, at 730–31.

[212] Joseph Sanders, Shari S. Diamond & Neil Vidmar, *Legal Perceptions of Science and Expert Knowledge*, 8 PSYCH., PUB. POL'Y, & L. 139, 148 (2002).

[213] Adina Schwartz, *A "Dogma of Empiricism" Revisited:* Daubert v. Merrell Dow Pharmaceuticals, Inc. *and the Need to Resurrect the Philosophical Insight of* Frye v. United States, 10 HARV. J. L. & TECH. 149, 182 (1996); FEYERABEND, *supra* note 151, at 12.

[214] *See* discussion *supra* Section II.A.

[215] *See* discussion *supra* Section I.D.

would be if such safety judgments could be made on a small handful of criteria, such as adding two more criteria—say, fire safety and earthquake safety. However, somewhat unfortunately, the fact of the matter is that such a judgment can only be made after considering a rather large set of safety dimensions, such as wind shear, absence of carcinogenic materials, air quality, appropriate maintenance, and so on. The proper consideration of all these dimensions makes the judgment more complex and time-consuming but it also makes the judgment more accurate.

These recommendations may not be exhaustive; these are meant to serve as a more accurate, complete, and useful proposal than existing admissibility criteria. Moreover, attorneys in conjunction with their experts would need to provide meta-arguments regarding the relative importance of some subset (proper or improper) of these recommendations in their particular situation. Experts qua experts should be able to readily do this—experts should know their field and the quality of the science involved, and thus the content demanded by each of these criteria. Judges, after hearing evidence pertaining to these criteria (including possible rebuttal testimony from the opposing side), would then be in the best position to make the decision about the admissibility of the proposed testimony. We all might wish that this matter was not so complex, but to simplify (and not oversimplify) this list, one would need to argue for the irrelevance of any individual proposed criterion judged. For example, the argument would take the position that the proposed criterion is never relevant to appraising the quality of any part of what is taken to be science. We ourselves have applied this evaluation standard and concluded that each standard was not susceptible to such criticism.

A. The 39 Criteria to Evaluate Science for Admissibility

1. *Quality of assumptions.* These are the starting points of scientific work—what is presumed. For example, scientists working on the problem, "Why do females have higher rates of depression than males?" assume that females have higher rates of depression, and more fundamentally assume that there are two genders. Thus, a key activity is to accurately identify the assumptions of a scientific work and expose these to criticism to assess their merit.
2. *Quality of logical relationships.* Scientific discourse is not a scramble of claims but, rather, there are logical relations between these claims. For example, in Carl Hempel's influential model of scientific explanation, deductive logic is used.[216] Thus, the quality of the logical relations of the claims of the scientific discourse can be evaluated.
3. *Clearly defined constructs.* Scientific discourse employs a multitude of concepts such as "force," "covalent bonding," "intelligence," "trauma," and "ego

[216] James Woodward & Lauren Ross, *Scientific Explanation*, STANFORD ENCYC. PHIL., https://plato.stanford.edu/archives/sum2021/entries/scientific-explanation (May 10, 2021).

defenses." The extent to which each of these constructs is precisely defined, vaguely defined, or well-defined can be evaluated.

4. *Valid measures of constructs.* Once a construct is defined, the next question then is whether this definition can be operationalized into a measurement operation that validly assesses the construct. Thus, any science can be evaluated on the evidence and can be accurately measured for the particular population and context at issue.

5. *Conceptual problems.* The philosopher of science Larry Laudan has argued that science has both empirical problems and conceptual problems.[217] An example of a conceptual problem is seen in Einstein's work.[218] Einstein's relativity theory dealt with conceptual problems such as, "Does motion affect how an observer moving relative to a clock measures its rate?" Thus, the number and importance of conceptual problems of some scientific discourse can be critiqued and evaluated.

6. *Clearly stated boundary conditions.* Scientific theories often have a certain limited scope (e.g., the boundary conditions of mechanical theory in physics are a vacuum).[219] The scope of scientific regularities in the social sciences may be restricted by variables such as age, gender, and culture. Thus, a theory can be critiqued regarding whether it appropriately and clearly stated these boundaries.

7. *Falsifiability of a theory.* À la Popper, a theory or model should exclude some observable state of affairs, otherwise, logically, it is compatible with all states of affairs and is not actually an empirical claim.[220]

8. *Quality of scientific explanation.* There are important issues regarding what counts as a legitimate scientific explanation. For example, Hempel's covering law model states that a necessary condition is that the putative explanation is a predictive consequence of at least scientific law.[221]

---

[217] *See* LAUDAN, *supra* note 7, at 14, 45.

[218] 3 SAMUEL J. LING, WILLIAM MOEBS & JEFF SANNY, UNIVERSITY PHYSICS 231 (2021), https://openstax.org/details/books/university-physics-volume-3 (select "Download a PDF" under "Get the book").

[219] *See* Julia R. S. Bursten, *The Function of Boundary Conditions in the Physical Sciences*, 88 PHIL. SCI. 234, 238 (2021) (defining a "boundary condition" as a "specified sets of values that a differential equation must take at the boundary region of the problem's solution space."); Charlie Wood, *The Key to Understanding the Origin and Fate of the Universe May be a More Complete Understanding of the Vacuum*, QUANTA MAG. (Aug. 9, 2022), https://www.quantamagazine.org/how-the-physics-of-nothing-underlies-everything-20220809/ (describing the foundational importance of the vacuum to physics).

[220] Sven Ove Hansson, *Science and Pseudo-Science*, STANFORD ENCYC. PHIL., https://plato.stanford.edu/archives/fall2021/entries/pseudo-science/ (May 20, 2021).

[221] Woodward & Ross, *supra* note 216.

9. *Quality of testing of scientific discourse.* Specifically, how well has it survived severe testing? How many tests has the scientific discourse been subject to? What are the results of these tests; supportive, falsification, or mixed? What is the quality of these tests?

10. *Methodological limitations.* Although related to the above, a critical discussion of all the methodological features of the tests is useful. What are the limitations of the research design used (e.g., correlational designs cannot infer causation)? Other methodological parameters include: Were the appropriate control conditions used?; Was blindness used?; Was there a representative sample?; Was the sample size sufficiently large to allow sufficient statistical power?; and so on.

11. *Problem solving effectiveness.* Beyond statistical significance, what was the practical or clinical significance? For example, O'Donohue and colleagues found that the evidence-based therapies for childhood obesity generally reduced weight by only a modest amount one-year post-treatment.[222] This at best is a partial solution to the problem.

12. *Handling the Duhem–Quine problem by avoiding* ad hoc *strategies/degenerating.* How did the study handle any anomalous results? Was the scientific discourse being tested problematically protected from prediction failures by *ad hoc* strategies blaming some auxiliary hypotheses?

13. *Use of questionable research practices.* There are a variety of questionable research practices that degrade the quality of the test. These include selectively reporting outcome variables, hypothesizing after the results are known, and p-hacking, among others.[223] Did any of these affect the scientific theory in question?

14. *Preregistered studies.* To what extent were studies relevant to the scientific discourse preregistered? Preregistration has a number of advantages, including increasing the likelihood of the reporting of null findings, and it also can prevent fishing, i.e., manipulating the discourse to achieve a desired result, because the researcher publicly describes the scientific methodology and analysis ahead of time.

15. *Open Data.* To what extent were the raw data published in a public archive to allow independent inspection and analysis?

---

[222] *See* WILLIAM O'DONOHUE, BRIE A. MOORE & BARBARA J. SCOTT, HANDBOOK OF PEDIATRIC AND ADOLESCENT OBESITY TREATMENT 190 (2008).

[223] Angelika M. Stefan & Felix D. Schönbrodt, *Big Little Lies: a Compendium and Simulation of p-Hacking*, ROYAL SOC'Y OPEN SCI. 10, 1, 2–5 (2023), https://pmc.ncbi.nlm.nih.gov/articles/PMC9905987/ (analyzing how the widespread use of questionable research practices has jeopardized the credibility of scientific results).

16. *Publication in quality peer-reviewed journals.* To what extent were the publications in peer-reviewed versus non-peer-reviewed journals, and what is the overall quality of the journals this science was published in?

17. *Replications.* To what extent have the results either been directly or indirectly replicated, preferably by independent labs?

18. *Positive outcomes in adversarial collaborations.* Science can involve competing theories. Nobel Prize laureate Daniel Kahneman proposed that scientific disputes might be best resolved by having researchers who have different positions on the issues collaborate in the design, execution, and interpretation of a study with the aim of making progress on their disputed research question.[224] Thus, the question can be: Were there any adversarial collaborations and what were the outcomes of these?

19. *No better competing theory.* This potential criticism casts a wide net and addresses the question of whether there is a competing theory in the domain that has a superior scientific status (perhaps on the dimensions discussed here).

20. *Connections to other established scientific regularities.* This refers to whether the scientific theory informs and is informed by neighboring scientific fields. Scientific theories describe reality that is not parsed into isolated academic silos. Thus, to what extent is the scientific theory congruent with other scientific regularities? For example, the scientific fields of evolutionary theory and genetics are intimately related.[225]

21. *Degree of precision.* Scientific theories can vary in the precision of their claims. Mechanics in physics makes point predictions, while feminist theories of depression generally make qualitative predictions (e.g., due to sexism, women will have a greater rate of depression than men but how much greater is not specified or known).[226]

---

[224] *Daniel Kahneman Wins Nobel Prize*, PRINCETON UNIV. OFF. OF COMMC'N (Oct. 9, 2002), https://pr.princeton.edu/news/02/q4/1009-kahneman.htm; Cory J. Clark & Philip E. Tetlock, *Adversarial Collaboration: The Next Science Reform*, *in* IDEOLOGICAL AND POLITICAL BIAS IN PSYCHOLOGY: NATURE, SCOPE, AND SOLUTIONS 906–07 (Craig L. Frisby, Richard E. Redding, William T. O'Donohue & Scott O. Lilienfeld eds., 2023).

[225] Michael Bosnall & Brian Charlesworth, *Genetics and the Causes of Evolution: 150 Years of Progress since Darwin*, 365 PHIL. TRANSACTIONS ROYAL SOC'Y B 2427, 2427 (2010).

[226] *See* Robyn Bluhm, *Gender Differences in Depression: Explanations from Feminist Ethics*, 4 INT'L J. FEMINIST ETHICS 69, 74, 78, 84 (2011); David Wallace, *The Quantitative Content of Statistical Mechanics*, 52 STUD. HIST. & PHIL. SCI. 285 (2015).

22. *Bias in the scientific discourse.* Researchers cannot only be biased in terms of wanting to confirm their favored beliefs, but can hold a number of political, cultural, gender, and other ideological biases.[227]

23. *Problematic value commitments.* Some theories can be critiqued for the normative claims they contain. For example, eugenic theories have been critiqued for problematic values concerning those with disabilities.[228]

24. *Quality of statistical analyses.* All statistical tests make assumptions about the data set (e.g., parametric tests assume that the data are normally distributed). The research can be critiqued regarding both the choice of what statistics were used, as well as the quality of the implementation or reporting of the statistical analysis.

25. *Practical/ethical constraints on research.* Some useful research cannot be conducted due to reasonable concerns about the ethics of the research.[229] For example, a randomly controlled trial of the effects of smoking on subsequent cancer rates cannot be conducted ethically as it would likely be very harmful to those randomly assigned to the smoking condition. Studies can be critiqued on these ethical grounds, but critiques of the field can themselves be criticized because they fail to recognize the legitimate limitations placed by ethical and other practical constraints.

26. *Used in harmful ways/iatrogenesis.* Applied scientific work can be evaluated on the dimension of whether it is causing harm in its practical application. For example, some branches of physics have been criticized for their involvement in weapons development.[230]

27. *Level of hucksterism.* To what extent do proponents of the research exaggerate or oversell the science, perhaps for professional or financial gain?

28. *Product of a research community*. To what extent is the science part of a scientific enterprise—or is it unconnected and thus fringe in a way that is problematic? The scientific community, through such activities as training and peer review, might be considered one large objector group providing critical feedback at

---

[227] Craig L. Frisby, Richard E. Redding & William T. O'Donohue, *Ideological and Political Bias in Psychology: An Introduction*, *in* IDEOLOGICAL AND POLITICAL BIAS IN PSYCHOLOGY: NATURE, SCOPE, AND SOLUTIONS 1, 1–2 (Craig L. Frisby, Richard E. Redding, William T. O'Donohue & Scott O. Lilienfeld eds., 2023).

[228] PAULINE MAZUMDAR, EUGENICS, HUMAN GENETICS AND HUMAN FAILINGS; THE EUGENICS SOCIETY, ITS SOURCES AND ITS FAILINGS 225–28 (1992).

[229] Goshen D. Mitue, *Ethics in Scientific Research: A Lens into Its Importance, History, and Future*, 86 ANNALS MED. & SURGERY 2395, 2395–96 (2024).

[230] Michael Atiyah, *Science for Evil: The Scientist's Dilemma*, 319 BMJ 448 (1999).

many different points. While other contexts that involve less criticism (e.g., a cult) are divorced from this crucial winnowing and corrective feedback.

29. *Nonauthoritarian.* To what extent is the degree of credence in the scientific discourse a product of data, as opposed to being believed due to some appeal to the authority of some person, persons, or institution?

30. *The extent to which scholars have been free to do research and criticize the scientific discourse.* To what extent are there personal, political, or institutional barriers to criticizing the scientific discourse? For example, a theory of evolution based on Lamarckianism was developed by the Russian scientist Lysenko.[231] This theory was endorsed by the ruling communist authorities in the Soviet Union, and those who criticized the theory received a variety of negative consequences, including execution.[232]

31. *Problematic ontology.* An ontology is a set of claims about what kinds of entities exist. Scientific discourse can be criticized for problems regarding the kinds of entities the scientific discourse claims to exist, from souls to entities that spontaneously regenerate, to vital forces, to a putative "inner child."

32. *Evidence of fraud.* Unfortunately, fraud does exist in science where researchers fabricate data. If there is evidence of this, then this is a legitimate dimension of criticism.

33. *Simplicity/parsimony/Occam's razor.* Scientific discourse should not be unnecessarily complex. To what extent does the scientific discourse have unnecessary assumptions or entities?

34. *Fecundity/fruitfulness.* Scientific theories can vary in the number and range of problems they have solved. This is similar to Kuhn's notion of the degree of problem solving of a paradigm in normal science.[233]

35. *Extent to which the scientific discourse has been shown to solve practical problems.* Does a theory only solve abstract basic scientific problems or to what extent has it been shown to be useful in addressing applied problems?

36. *The extent of problems that the scientific discourse has not been able to solve.* Conversely, what problem or problems have been refractory to solutions for the scientific discourse? This is similar to Kuhn's notion of puzzle-solving failures or anomalies.[234]

---

[231] 2 WILLIAM DEJONG-LAMBERT & NIKOLAI KREMENTSOV, THE LYSENKO CONTROVERSY AS A GLOBAL PHENOMENON: GENETICS AND AGRICULTURE IN THE SOVIET UNION AND BEYOND 180, 191 (2017).

[232] *Id.*

[233] *See generally* KUHN, STRUCTURE OF SCIENTIFIC REVOLUTIONS, *supra* note 71.

[234] *Id.*

37. *The error rate.* Some scientific discourse has error rates (e.g., the sensitivity and specificity of cancer screens).[235] The magnitudes of the error rates are subject to critical evaluation and subject to comparison of the error rates to its competitors.

38. *Adequacy of response to major criticisms.* Science is a critical process in which criticisms are made and, at times, reasonably countered. The quality of the responses to these criticisms can be evaluated.

39. *Researchers' conflicts of interest.* To what extent do researchers have conflicts of interest (e.g., financial) and have these been clearly stated and properly dealt with? For example, medical research has been compromised by payments to researchers from pharmaceutical companies.[236]

V. CSAAS UNDER *FRYE* AND *DAUBERT* STANDARDS

We next provide analyses of the admissibility of a psychiatric conjecture called the child sexual abuse accommodation syndrome (CSAAS) using the *Frye*, *Daubert*, and the polythetic model proposed here. We chose CSAAS as we believe it serves as a relevant and interesting example because it has been judged by states as both admissible and inadmissible under both *Frye* and *Daubert* standards. Thus, these standards have not yielded a reliable coherent judgment about its admissibility. Further, in analyzing the epistemic status of CSAAS with our proposed criteria, both the complexity of the task and usefulness of our model can be seen.

The child sexual abuse accommodation syndrome or CSAAS framework was introduced by psychiatrist Roland Summit in a publication in 1983.[237] As of November 2014, CSAAS was admissible in 40 states and ruled inadmissible in three states, while seven states had no statute or case law dealing with CSAAS.[238] This situation is dynamic—in 2017, New Jersey changed its status from admissible to inadmissible.[239]

---

[235] *See, e.g.*, Liu et al., *supra* note 135, at 755.

[236] Gisela Schott, Henry Pachl, Ulrich Limbach, Ursula Gundert-Remy, Klaus Lieb & Wolf-Dieter Ludwig, *The Financing of Drug Trials by Pharmaceutical Companies and Its Consequences*, 107 DEUTSCHES ÄRZTEBLATT INT'L 295, 295 (2010).

[237] *See generally* Roland Summit, *The Child Sexual Abuse Accommodation Syndrome*, 7 CHILD ABUSE & NEGLECT (1983) [hereinafter Summit, *CSAAS*].

[238] *See generally Admissibility of Expert Testimony as to Child Sexual Abuse Accommodation Syndrome*, NAT'L DIST. ATTYS. ASSOC. (Nov. 2014), https://ndaa.org/wp-content/uploads/CSAAS-Expert-Testimony-2014.pdf.

[239] State v. J.L.G., 190 A.3d 442, 446 (N.J. 2018).

A. What is the Child Sexual Abuse Accommodation Syndrome (CSAAS)?

Summit introduced CSAAS based on his belief that child victims of sexual abuse undergo secondary problems when, upon disclosure, their abuse is judged to be a lie by key adult individuals, such as their parents.[240] Despite showing no data to support this claim about this negative judgment, Summit claimed that adults do not believe children when they disclose sexual abuse by a "respectable, reasonable adult" and, in fact, "most adults who hear the accusation will fault the child."[241] He further stated this negative judgment was because, "Adult beliefs are dominated by an entrenched and self-protective mythology that passes for common sense. 'Everybody knows' that adults must protect themselves from groundless accusations of seductive or vindictive young people. An image persists of nubile adolescents playing dangerous games out of their burgeoning sexual fascination."[242]

Summit further claimed the disbelief by adults reinforces the abused "child's tendency to deal with the trauma as an intrapsychic event" and leads the child "to incorporate a monstrous apparition of guilt, self-blame, pain and rage."[243] Again, he presented no data showing these alleged outcomes. When clinicians are asked for help, Summit claimed that "[i]n present practice it is not unusual for clinical evaluation to stigmatize legitimate victims as either confused or malicious."[244] Again, he presented no data in his work that indicated the extent to which clinicians were actually making these problematic judgments.

CSAAS was generated, according to Summit, not by a presentation of new relevant scientific data but rather "in part from statistically validated assumptions regarding prevalence, age, relationships and role characteristics of child sexual abuse and in part from correlations and observations that have emerged as self-evident within an extended network of child abuse treatment programs and self-help organizations."[245] However, Summit failed to provide any specifics regarding his alleged "statistically validated assumptions" or what he means by "self-evident," although the rhetorical value of these phrases is considerable.

Summit stated that CSAAS was "derived from the collective experience of dozens of sexual abuse treatment centers in dealing with thousands of reports or complaints . . . ."[246] Although, again, there is a remarkable amount of vagueness in these numbers ("dozens," "thousands"). He further claims:

---

[240] Summit, *CSAAS*, *supra* note 237, at 177.

[241] *Id.* at 178.

[242] *Id.*

[243] *Id.* at 179.

[244] *Id.*

[245] *Id.* at 180.

[246] *Id.* at 190.

> Without a clear understanding of the accommodation syndrome, clinical specialists tend to reinforce the comforting belief that children are only rarely legitimate victims of unilateral sexual abuse and that among the few complaints that surface, most can be dismissed as fantasy, confusion, or a displacement of the child's own wish for power and seductive conquest.[247]

Summit then makes additional claims about what the typical reactions of children to sexual abuse are.[248] Interestingly, he failed to precisely specify if these reactions occur only if the child was sexually abused and not believed, or if sexual abuse itself was sufficient to produce these reactions. Further, he admitted that, in his field, there was a (problematic) general assumption that all claims of sexual abuse were true.[249]

1. *Secrecy*. This phase involves silence by the child victim about the actions of the perpetrator. Summit claims, "The secret takes on magical, monstrous proportions for the child. A child with no knowledge or awareness of sex and even with no pain or embarrassment from the sexual experience itself will still be stigmatized with a sense of badness and danger from the pervasive secrecy."[250] Obviously "magical, and monstrous proportions" is wording that is more literary than scientific. According to Summit, when the child attempts to tell an adult about the abuse, "the secret will be countered by an adult conspiracy of silence and disbelief" and will be met with responses such as "'[d]on't worry about things like that; that could never happen in our family'" or "'[n]ice children don't talk about things like that.'"[251] Again, Summit presented no data regarding the extent to which adults actually respond in this manner.
2. *Helplessness*. Because children are expected to be "obedient and affectionate" towards adults with whom they are entrusted, children will feel helpless against such adults if they are the ones who perpetrate the sexual abuse.[252] Summit claimed:

> The prevailing reality for the most frequent victim of child sexual abuse is not a street or schoolground experience and not some mutual vulnerability to oedipal temptations, but an unprecedented, relentlessly progressive intrusion of sexual acts by an overpowering adult in a one-sided victim-perpetrator relationship. The fact that the perpetrator is often in a

---

[247] *Id.* at 179.

[248] *Id.* at 181–88.

[249] *Id.* at 190–91.

[250] *Id.* at 181.

[251] *Id.*

[252] *Id.* at 182.

> trusted and apparently loving position only increases the imbalance of power and underscores the helplessness of the child.[253]

Summit went on to explain that children bear the abuse silently and while "[b]ed covers take on magical powers against monsters, . . . they are no match for human intruders."[254] For this reason, Summit claimed, "The more illogical and incredible the initiation scene might seem to adults, the more likely it is that the child's plaintive description is valid."[255] This is a particularly remarkable claim—illogical and fantastical claims of abuse (e.g., "I was abused simultaneously by Presidents Obama and Bush while we were floating in the air"), according to Summit, are more likely to be valid than claims that a stepfather touched the child while their mother was outside the home at work.

At the same time, Summit claimed, "Adults tend to despise helplessness and to condemn anyone who submits too easily to intimidation."[256] Summit further elucidated this reaction of helplessness by claiming, "If the child's testimony is rejected in court, there is more likely to be a rejection by the mother and other relatives who may be eager to restore trust in the accused adult and to brand the child as malicious."[257] For this reason, Summit stated that children who are victims of sexual assault need a "clinical advocate to translate the child's world into an adult-acceptable language."[258] Additionally, to bolster his claims about helplessness, Summit cited a letter from advice columnist, Ann Landers.[259] This is quite an unusual move in scientific discourse.

3. *Entrapment and accommodation*. Sexual abuse, when perpetrated by a known adult, typically occurs more than once. This is because even though "[t]he adult may be racked with regrets, guilt, fear and resolutions to stop . . . the forbidden quality of the experience and the unexpected ease of accomplishment seem to invite repetition."[260] As a result, the child, who is typically helpless and unable to receive protection (entrapment), is forced to accept or accommodate this abuse:

> The [abused] child cannot safely conceptualize that a parent might be

---

[253] *Id.* at 182–83. "Oedipal temptations" is a phrase from outmoded Freudian psychoanalysis. JEAN LAPLANCHE & JEAN-BERTRAND PONTALIS, THE LANGUAGE OF PSYCHO-ANALYSIS 282–83 (Donald Nicholson-Smith trans., Karnac Books 1973) (1967).

[254] Summit, *CSAAS*, *supra* note 237, at 183.

[255] *Id.*

[256] *Id.*

[257] *Id.*

[258] *Id.*

[259] *Id.* at 183–84.

[260] *Id.* at 184.

> ruthless and self-serving; such a conclusion is tantamount to abandonment and annihilation. The only acceptable alternative for the child is to believe that she has provoked the painful encounters and to hope that by learning to be good she can earn love and acceptance.[261]

Summit does not discuss any alternatives to this acceptance such as those who do report when abuse begins.

Summit claims that in order to protect the non-abusive parent, the child accommodates the abuse by potentially turning to "imaginary companions for reassurance."[262] He claims that the abused child "may develop multiple personalities, assigning helplessness and suffering to one, badness and rage to another, sexual power to another, love and compassion to another, etc. . . . The same mechanisms which allow psychic survival for the child become handicaps to effective psychological integration as an adult."[263] Again, he presents no data on the frequency that sexual abuse produces dissociative identities. Further, "[s]he may learn to exploit the father for privileges, favors and material rewards, reinforcing her self-punishing image as 'whore' in the process."[264] He presents no data regarding the extent to which sexually abused children have a self-image of "whore."

Further, "[t]he ungratifying, imperfect behavior of the young child and the diffusion of ego boundaries between parent and child invite projection of the bad introject and provide a righteous, impulsive outlet for the explosive rage."[265] Again, he relies on the outmoded psychoanalytic constructs of "projection," and "bad introject." Summit claims that even professionals may be unable to resist blaming the child victim; he states:

> It is all too easy for the would-be therapist to join the parents and all of adult society in rejecting such a child, looking at the results of abuse to assume that such an "impossible wretch" must have asked for and deserved whatever punishment had occurred, if indeed the whole problem is not a hysterical or vengeful fantasy.[266]

Again, he presents no scientific evidence of the frequency of therapists behaving in this manner.

4. *Delayed, conflicted, and unconvincing disclosure*. Summit claimed that most ongoing abuse is never disclosed. "Treated, reported or investigated cases

---

[261] *Id.*

[262] *Id.* at 185.

[263] *Id.*

[264] *Id.*

[265] *Id.*

[266] *Id.* at 186.

are the exception, not the norm."[267] For Summit, "Disclosure is an outgrowth either of overwhelming family conflict, incidental discovery by a third party, or sensitive outreach and community education by child protective agencies."[268] However, he also stated, "A child who seeks help immediately or who gains effective intervention should not be discarded as contradictory . . . ."[269] He stated that because abuse is typically disclosed years later, the child is usually an adolescent (or older) at the time of disclosure which is also a time that, according to Summit, "makes the father more jealous and controlling, trying to sequester his daughter against the 'dangers' of outside peer involvement. The corrosive effects of accommodation seem to justify any extreme of punishment."[270] He clarified, "It is worth restating that all these accommodation mechanisms—domestic martyrdom, splitting of reality, altered consciousness, hysterical phenomena, delinquency, sociopathy, projection of rage, even self-mutilation—are part of the survival skills of the child."[271] Again, no further definition of these or frequencies of these are provided.

Summit summarized this by claiming, "Whether the child is delinquent, hypersexual, counter-sexual, suicidal, hysterical, psychotic, or perfectly well-adjusted, and whether the child is angry, evasive or serene, the immediate affect and the adjustment pattern of the child will be interpreted by adults to invalidate the child's complaint."[272] Again, this is a complex empirical claim but no data are presented to support any part of it. According to Summit, the mother of the victimized child is particularly conflicted with accepting the sexual abuse allegations and is reassured when the perpetrator who is the father reassures her:

> "Are you going to believe that lying little slut? Can you believe I would do such a thing? How could something like that go on right under your nose for years? You know we can't trust her out of our sight anymore. Just when we try to clamp down and I get a little rough with her, she comes back with a ridiculous story like this. That's what I get for trying to keep her out of trouble."[273]

This claim has a lot of specificity, but again, no data are presented on the frequency at which this occurs.

---

[267] *Id.*

[268] *Id.*

[269] *Id.* at 180.

[270] *Id.* at 186.

[271] *Id.*

[272] *Id.* at 187.

[273] *Id.*

5. *Retraction*. Finally, when a child discloses their sexual abuse, their disclosure is often recanted. Summit stated, "Whatever a child says about sexual abuse, she is likely to reverse it."[274] He also made the following strong claims:

> In the chaotic aftermath of disclosure, the child discovers that the bedrock fears and threats underlying the secrecy are true. Her father abandons her and calls her a liar. Her mother does not believe her or decompensates into hysteria and rage. The family is fragmented, and all the children are placed in custody. The father is threatened with disgrace and imprisonment. The girl is blamed for causing the whole mess, and everyone seems to treat her like a freak.[275]

Note also that Summit, in claiming that these five dimensions comprise a syndrome, claimed that these five dimensions co-occur. However, he provided no evidence of such co-occurrence, and an important question is that if in some particular case all five conditions are not present, should this in itself be considered data that are contrary to his claims and thus sufficient grounds for its inadmissibility?

About ten years later, Summit published a response to what he perceived as distortions and misuse of CSAAS in courts.[276] Summit clarified that CSAAS did not originate as a scientific hypothesis or "a designated study of a defined population" but as a "summary of diverse clinical consulting experience."[277] He further stated that "CSAAS is a clinical opinion, not a scientific instrument."[278] In this publication, Summit also clarified the origination of CSAAS which was based on his "own broad consulting experience throughout Los Angeles County as well as personal discussions with . . . national visionaries."[279] What criteria one has to meet to be considered a "visionary" were not specified.

Finally, Summit stated that he wished he originally used the term "pattern" instead of the word "syndrome" because a syndrome implies that a valid inference can be made to abuse status, which he claimed could not be legitimately done from CSAAS.[280] He also acknowledged, "there is no clinical method available to distinguish 'valid' claims from 'those that should be treated as fantasy or deception,' and [CSAAS] gives no guidelines for discrimination."[281]

---

[274] *Id.* at 188 (emphasis removed).

[275] *Id.*

[276] *See generally* Roland Summit, *Abuse of the Child Sexual Abuse Accommodation Syndrome*, 1 J. CHILD SEXUAL ABUSE 153, 153–54 (1993) [hereinafter Summit, *Abuse of CSAAS*].

[277] *Id.* at 156.

[278] *Id.*

[279] *Id.* at 154.

[280] *Id.* at 157.

[281] *Id.* at 158 (quoting Summit, *CSAAS*, *supra* note 237, at 189).

B. Admissibility of CSAAS as Expert Evidence in Federal Courts

Federal courts, which are bound to follow the FRE or *Daubert* standard, notice that CSAAS is scientifically invalid or "junk" science and strongly prohibit admission on expert evidence or testimony on the matter.[282] In fact, the skepticism of CSAAS is so severe that the Second Circuit has admitted its use in a case as a ground for a valid claim of ineffective assistance of counsel, i.e., failure of defense counsel to fulfill their affirmative duty to consult medical experts in child sexual abuse cases.[283]

C. Admissibility of CSAAS as Expert Evidence in State Courts

The matter of admissibility of CSAAS in state courts, on the other hand, is problematically varied. Regardless of the admissibility standards used by the state, it is common for courts to accept the admissibility of CSAAS.[284] However, some state courts are aware of the potential abuse of CSAAS and therefore limit expert testimony based on CSAAS to explain common behavioral traits of sexually abused children in general.[285] Courts have disallowed such expert testimony when it has been used to bolster victim credibility or to prove occurrence of sexual abuse.[286] Consequently, courts have allowed such expert testimony as a rehabilitative tool, i.e., to generally speak to the common traits in child victims of sexual abuse, instead of using it as a diagnostic tool to establish proof of sexual abuse in that particular case.[287]

We present CSAAS as expert evidence in two states across the country: (1) New Jersey, where state courts have evolved in admitting expert evidence under CSAAS based on growing evidence against said syndrome and (2) California, where despite empirical findings falsifying most claims made under CSAAS, courts continue to allow expert testimony based on said syndrome.

In 1993, the State of New Jersey allowed expert testimony regarding CSAAS primarily to rehabilitate a victim's credibility (e.g., in cases where the defense was expected to claim that the child victim delayed reporting or recanted and was therefore unworthy of belief).[288] In fact, the New Jersey Supreme Court, in *State v. J.Q.*,

---

[282] Margaret Shiu, Note, *Unwarranted Skepticism: The Federal Courts' Treatment of Child Sexual Abuse Accommodations Syndrome*, 18 S. CAL. INTERDISCIPLINARY L. J. 651, 658 (2008). *See generally* Joëlle Anne Moreno, *Beyond the Polemic Against Junk Science: Navigating the Oceans that Divide Science and Law with Justice Breyer at the Helm*, 81 B.U. L. REV. 1033 (2001) (discussing the concept of "junk" science and how it permeates the legal system).

[283] Shiu, *supra* note 282, at 662–63 (citing Lindstadt v. Keane, 239 F.3d 191 (2d Cir. 2001)).

[284] *Id.* at 655–56.

[285] *Id.* at 656.

[286] *Id.*

[287] *Id.*

[288] State v. J.Q., 617 A.2d 1196, 1201 (N.J. 1993).

affirmed the admissibility of such testimony without reevaluating CSAAS evidence under the *Frye* standard (which was the standard then used by the state) that it was scientifically valid and reliable.[289] What is unclear is why the courts thought the entire complex edifice of CSAAS is necessary for this instead of simply allowing more focused scientific testimony regarding the frequency of immediate versus delayed responding.

This point was recognized in 2018, in *State v. J.L.G.*, where the New Jersey Supreme Court rightly observed that there was consensus only for one of the five conditions of CSAAS (delayed disclosure) and CSAAS did not satisfy the *Frye* standard.[290] As a result, experts would not be permitted to present evidence on any other CSAAS condition. However, the court also ruled that only when all the prongs of the admissibility criteria were met could expert evidence be presented *only* on delayed disclosure of abuse.[291] Further, the court stipulated that any expert evidence on such delayed disclosure should be accompanied with limiting instructions to the jury before expert witness testimony as part of the court's final charge which are as follows:

> Dr. [_____]'s testimony is offered only to explain his/her opinion that delayed complaints of sexual abuse are common, and not necessarily inconsistent with sexual abuse. The weight to be given to Dr. [_____]'s testimony is entirely up to you. You may give it great weight, or slight weight, or you may, in your discretion, reject it entirely.
>
> Regardless of the weight you give to Dr. [_____]'s testimony, if any, you may not consider this testimony as proof that sexual abuse occurred. You also may not consider the doctor's testimony as proof that [complaining witness]'s complaint was truthful.
>
> You may only consider Dr. [_____]'s testimony, if you accept it, for the principle that delayed disclosure commonly occurs among victims of child sexual abuse and is not necessarily inconsistent with sexual abuse. This testimony about delayed disclosure, if you choose to give it any weight, may only be considered in assessing the complaining witness's credibility and for no other purpose. You may not consider the expert testimony as in any way proving that [defendant] committed, or did not commit, any particular act of sexual abuse. The ultimate determination of whether or not the State has proven defendant's guilt beyond a reasonable doubt is to be made only by the jury.[292]

---

[289] *See id.* at 1211–12.

[290] State v. J.L.G., 190 A.3d 442, 446 (N.J. 2018).

[291] *Id.* at 464–65.

[292] N.J. CTS., MODEL CRIMINAL JURY CHARGES, DELAYED DISCLOSURE OF CHILD SEXUAL ABUSE (WHERE STATE INTRODUCES EXPERT TESTIMONY) (Apr. 8, 2019).

However, the State of California is among other states that continues to allow expert evidence and testimony under CSAAS.[293] In 1991, the California Supreme Court in *People v. McAlpin* observed that CSAAS "is not admissible to prove that the complaining witness has in fact been sexually abused; it is admissible to rehabilitate such witness's credibility when the defendant suggests that the child's conduct after the incident—e.g., a delay in reporting—is inconsistent with his or her testimony claiming molestation."[294] Unlike New Jersey, which limited testimony to only the question of delay of disclosing, this California ruling is interpreted to allow all aspects of CSAAS to be admissible for rehabilitating witness credibility.[295]

In a more recent case in 2020, *People v. Munch*, the issues with admitting CSAAS and related expert evidence was raised again in California's Second District Court of Appeals. Disregarding the defendant's claim that CSAAS is out of date and has a prejudicial effect on a defendant, said court held that "CSAAS evidence is a valid and necessary component of the prosecution case in matters involving child abuse. We conclude the reasoning of *McAlpin* is as valid today as it was in 1991."[296] The court affirmed that CSAAS is admissible because it shed light on the child's credibility and expected behavior after an assault.[297] That said, the court did note that expert evidence under CSAAS was not being offered as scientific proof that the victim had been sexually assaulted or abused; the prosecutor's expert witness even agreed with the defense that there was no research-based, clinical technique to determine whether a child had been abused or not.[298] Nevertheless, expert evidence under CSAAS was permitted in this case[299] which sets the precedent for future cases in California.

The State of California requires certain mandatory instructions to be given by the presiding judge to the jury in cases where expert evidence or testimony under CSAAS is admitted.[300] However, it is worth noting that the judge has no sua sponte

---

[293] Some other states include New York and South Dakota. GUIDE TO N.Y. EVIDENCE, ARTICLE 7, OPINION EVIDENCE (n.d.) [hereinafter GUIDE TO N.Y. EVIDENCE], https://www.nycourts.gov/JUDGES/evidence/7-OPINION/ARTICLE-7-Rules.pdf; State v. McKinney, 699 N.W.2d 471 (S.D. 2005). Further, 40 states, along with Washington, D.C. allow expert testimony under CSAAS. Joe Hernandez, *N.J. High Court Bars Longtime Behavioral Theory from Child Sexual Abuse Cases*, NEWSWORKS TONIGHT (July 31, 2018), https://whyy.org/segments/n-j-high-court-bars-longtime-behavioral-theory-from-child-sexual-abuse-cases/.

[294] People v. McAlpin, 812 P.2d 563, 569 (Cal. 1991).

[295] People v. Munch, 52 Cal. App. 5th 464, 470 (2020).

[296] *Id.* at 466.

[297] *Id.* at 468.

[298] *Id.* at 473.

[299] *Id.* at 472.

[300] ADVISORY COMM. ON CRIM. JURY INSTRUCTIONS, REVISED JUDICIAL COUNCIL OF CALIFORNIA CRIMINAL JURY INSTRUCTIONS, at xxi (2024).

duty to give this instruction, i.e., judges are required to read these instructions *only upon being formally prompted*:

> You have heard testimony from _____ <*insert name of expert*> regarding child sexual abuse accommodation syndrome.
>
> Child sexual abuse accommodation syndrome relates to a pattern of behavior that may be present in child sexual abuse cases. Testimony as to the accommodation syndrome is offered only to explain certain behavior of an alleged victim of child sexual abuse.
>
> _____'s <*insert name of expert*> testimony about child sexual abuse accommodation syndrome is not evidence that the defendant committed any of the crimes charged against (him/her) [or any conduct or crime[s] with which (he/she) was not charged].
>
> You may consider this evidence only in deciding whether or not _____'s <*insert name of alleged victim of abuse*> conduct was consistent with the conduct of someone who has been molested, and in evaluating the believability of the alleged victim.[301]

D. *The Variability in Admissibility Standards and Variability of Admissibility Judgments Using the* Frye *and* Daubert *Standards*

The table below presents examples of the variability currently produced by (1) the variability of the standards used and (2) even when using the same standard.

Frye Standard		
State	Admissibility of CSAAS	Reasoning
California	Admissible	See *Admissibility of CSAAS as Expert Evidence in State Courts* (Section V.C).
New York	Admissible	Rule 7.08 of the Guide to New York Evidence specifically allows expert testimony under CSAAS.[302] In *People v. Austen*, the state Appellate Court rejected the defendant's contention that CSAAS was no longer generally accepted in its scientific community and only a small

[301] *Id.* No. 1193.

[302] GUIDE TO N.Y. EVIDENCE, *supra* note 293, § 7.08.

Frye Standard		
State	Admissibility of CSAAS	Reasoning
		number of other courts (such as New Jersey) had rejected CSAAS testimony.[303]
Illinois	Admissible	In *People v. Hodor*, a state appellate court observed that CSAAS was a recognized and accepted form of PTSD and testimony under CSAAS was permitted because it shed light on behavioral patterns typically manifested by sexual abuse victims.[304]
Pennsylvania	Inadmissible	In *Commonwealth v. Balodis*, the state court ruled that "expert testimony as to the veracity of a particular class of people," such as victims of sexual abuse, "of which the victim is a member," would be inadmissible.[305] Further, expert testimony about general characteristics of child sexual abuse victims was rejected because CSAAS had failed to meet the standard for the reliability of expert testimony under *Frye*.[306]
Florida (at the time of ruling)[307]	Inadmissible	In *Petruschke v. Florida*, the state appellate court observed that expert testimony that an alleged victim of sexual abuse exhibits symptoms consistent with a victim who has been abused may not be used in criminal cases because it does not meet the *Frye* standard.[308] The court specifically observed that CSAAS had not been generally accepted by the scientific community; however, it would allow expert testimony based solely upon the

[303] People v. Austen, 197 A.D.3d 861, 862 (N.Y. App. Div. 2021).

[304] People v. Hodor, 792 N.E.2d 828, 861 (Ill. App. Ct. 2003).

[305] Commonwealth v. Balodis, 747 A.2d 341, 345 (Pa. 2000).

[306] *Id.*

[307] Florida adopted the *Daubert* standard in 2017. *Admissibility of Expert Testimony in all 50 States*, *supra* note 138.

[308] Petruschke v. Florida, 125 So. 3d 274, 282 n.3 (Fla. Dist. Ct. App., 2013).

Frye Standard		
State	Admissibility of CSAAS	Reasoning
		testifier's training and experience on typical behaviors exhibited by abuse victims but not a direct testimony about the victim in question.[309]

Daubert Standard		
State	Admissibility of CSAAS	Reasoning
New Jersey	Inadmissible	See *Admissibility of CSAAS as Expert Evidence in State Courts* (Section V.C).
Arizona	Admissible	In *People v. Salazar–Mercado*, the state appellate court did not agree with the concerns raised by the defendant about permitting CSAAS evidence.[310] The court ruled that expert testimony would be permitted if it "generally explain[ed] behavioral characteristics of child sexual abuse victims without offering opinions" about the case in question.[311]
Indiana	Admissible	In *Lyons v. Indiana*, the state appellate court sided with the trial court that permitted CSAAS evidence because it was not presented as a diagnostic tool to prove that sexual abuse had occurred but to explain reactions (such as recanting or delayed reporting) exhibited by sexually abused children.[312] Further, the court also observed that if a child's credibility was called into question, expert testimony deemed appropriate (under evidentiary

[309] *Id.*

[310] People v. Salazar–Mercado, 325 P. 3d 996, 1001 (Ariz. 2014).

[311] *Id.* at 999.

[312] Lyons v. Indiana, 976 N.E.2d 137, 143 (Ind. App. 2012).

Daubert Standard		
State	Admissibility of CSAAS	Reasoning
		standards) would be permitted.[313] "'Because research generally accepted as scientifically reliable recognizes that child victims of sexual abuse may exhibit unexpected behavior patterns [that are] inconsistent with claim of abuse,'" evidence of specialized knowledge that would assist the jury to understand the case would be permitted.[314]

E. Admissibility of CSAAS under Proposed Recommendations

Next, we apply the polythetic model to evaluating CSAAS:

Proposed Criteria	Evaluations of CSAAS
Quality of assumptions	Falsely assumes most adults are not believing children's allegations; falsely assumes all allegations are true.
The quality of the logic	There is not a valid logical structure between the claims of CSAAS.
Clearly defined constructs	Many constructs are left undefined—e.g., "Oedipal temptation," "accommodation," "entrapment," "delay."
Valid measures of constructs	No valid measures of accommodation, entrapment, helplessness, and many other constructs.
Conceptual problems	It is not clear what CSAAS is attempting to explain: children's reactions to abuse?; children's reactions to disclosures of abuse that are not believed?; why adults fail to believe true allegations of abuse?, etc.
Clearly stated boundary conditions	Not clear at what ages CSAAS is relevant; not clear if it only applied to familial abuse.

[313] *Id.*

[314] *Id.* (quoting IND. EVID. R. 702).

Proposed Criteria	Evaluations of CSAAS
Falsifiable	No—since it cannot measure key constructs and without valid measurement operations testing cannot occur.
The quality of scientific explanation	Poor quality of scientific explanation as it provides no evidence for any law-like regularities.
How well tested is the scientific discourse. Specifically, what is the status of how it has survived severe testing	Has never been directly tested.
The extent of methodological limitations	Not applicable because there are no direct tests.
Problem solving effectiveness	Does not solve the problem of distinguishing between true and false allegations as Summit admits.[315]
Appropriately handled the *Duhem–Quine* problem by avoiding *ad hoc* strategies/degenerating	No falsifications since not tested.
Not compromised by questionable research practices	No direct studies, so irrelevant.
Preregistered studies	No direct studies, so irrelevant.
Open data	No direct studies, so irrelevant.
Published in quality peer review journal	The two articles by Summit were published in peer reviewed journals but are conceptual not empirical.
Replications	No direct studies, so irrelevant.
Positive outcomes in adversarial collaborations	No adversarial research.

[315] Summit, *Abuse of CSAAS*, *supra* note 276, at 159–60.

Proposed Criteria	Evaluations of CSAAS
No better competing theory	Yes, the PTSD model better captures the sequelae of abuse.[316]
Connections to other established scientific regularities (informs and is informed by neighboring legitimate fields)	Connected to research such as delay of allegations but unconnected to major theories of psychopathology.
Degree of precision	Very imprecise—no quantitative predictions.
Evidence of presence of bias in the scientific discourse	Largely used by prosecutors and those biased to minimize concerns with false positives.
Problematic value commitments	Unconcerned with false accusations. Makes many negative statements about mothers and abused children.
Quality of statistical analyses	Not relevant since no empirical studies.
Practical/ethical constraints on research	Cannot do true experiments due to ethical constraints but some research that looks at the goodness of fit of this model has been done that did not support the model.[317]
Used in harmful ways/iatrogenesis	Yes, can be used to falsely convict.
Level of hucksterism	None from the original proponent but expert witnesses perhaps motivated by financial gain do not fairly describe its weaknesses.
Product of a research community	No, originally devised by a clinician; never directly tested in a focused research program.
Nonauthoritarian	Some appeals are made to the originator's clinical experience, but the actual extent of this is unclear, and was uncontrolled.
Free to do research/criticize	No problems.

[316] William O'Donohue & Lorraine Benuto, *Problems with Child Sexual Abuse Accommodation Syndrome*, 9 SCI. R. MENTAL HEALTH PRAC. 20, 26 (2012).

[317] *Id.* at 25.

Proposed Criteria	Evaluations of CSAAS
Problematic ontology	Yes, relies on several outdated psychoanalytic entities like children's self-images as "whores" and narrow conceptions of familial responses to trauma.
Evidence of fraud	None.
Simplicity/Parsimony/Occam's razor	Problematic—PTSD model is more parsimonious.
Fecundity/fruitfulness	None—has not spun off any other theories, models, or findings.
Extent to which it solves practical problems	Poor, cannot be used to determine if abuse occurred or not and cannot be used as an educational tool for any subject matter.
The error rate	Unknown, but O'Donohue and Benuto calculated the following error rate: "[E]ven if we assume . . . CSAAS's first 3 factors are universally present (100% accurate); and even if we assume that recantation rates are 20% (which is clearly high based on the empirical literature); and [Summit's] conflicting allegations error rate is 85% then the overall error rate is 1.0 x 1.0 x 1.0 x .15 x .20 which is an error rate of 97%!"[318]
Adequacy of response to major criticisms	Ignored criticisms, no revisions made.
The extent of problems refractory to solution	Problems of identifying true allegations from false remains.
Researchers' conflicts of interest	Medium—professionals can gain funds by testifying regarding it in jurisdictions where it is admissible.

CSAAS, by doing so poorly on so many of these criteria and its failure do well on any that we argue that this set of criteria is shown to be useful, at least in this complex case, depicts that it ought not be regarded as knowledge or as scientific. We understand that a single case is insufficient to support general usefulness of our proposed criteria, but this can best be evaluated by further scholarly work in diverse

[318] *Id.* at 25–26.

content areas. We welcome modifications and improvements to the model that this work may provide.

CONCLUSION

Expert evidence can have a large impact on legal verdicts, and either the admission of epistemically problematic claims or the rejection of sound information can lead to unjust outcomes.[319] Thus, valid admissibility standards are key to fair judicial outcomes. However, currently, there is considerable and undesirable variability across jurisdictions on what admissibility standards are used leading to problematic variance in the appraisal, and subsequently the admissibility, of the same set of purportedly scientific claims. Moreover, as described in this review, there are several serious, and we argue fatal, problems in both the *Frye* and *Daubert* standards. Unsurprisingly, Kovera and McAuliffe found that using these admissibility standards, many judges were still likely to admit to expert evidence findings from methodologically flawed studies.[320]

We agree with Judge Kozinski in *Daubert* who observed: "[T]hough we are largely untrained in science and certainly no match for any of the witnesses whose testimony we are reviewing, it is our responsibility to determine whether those experts' proposed testimony amounts to 'scientific knowledge,' constitutes 'good science,' and was 'derived by the scientific method.'"[321] Judge Kozinski's apt use of the phrases "scientific knowledge," "good science," and "derived by the scientific method" (although the univocal nature of the last phrase is problematic) correctly points to the fact that the core issue is an epistemic one—i.e., what are the criteria used to evaluate whether some claim or claims counts as knowledge?

However, what counts as knowledge is not a completely settled matter in either epistemology or the philosophy of science. The search for a small number of necessary and sufficient conditions for what constitutes knowledge and science has failed despite decades and even centuries of searching by many extraordinarily talented scholars. We suggest the definitional matter is itself complex and that a polythetic model of evaluative criteria based on Wittgenstein's notion of family resemblances is useful to make progress in the problem. We thus propose a network of 39 evaluative criteria for this task. Attorneys, through their experts, must make meta-arguments about the relevance and importance of each of these for the case at hand.

---

[319] *See generally* Neil Vidmar & Shari Seidman Diamond, *Juries and Expert Evidence*, 66 BROOK. L. REV. 1121 (2001) (discussing how juries use expert testimony during deliberations).

[320] *See generally* Margaret Bull Kovera & Bradley D. McAuliff, *The Effects of Peer Review and Evidence Quality on Judge Evaluations of Psychological Science: Are Judges Effective Gatekeepers?*, 85 J. APPLIED PSYCH. 574, 74, 579–81, 585 (2000) (finding that some judges are willing to admit flawed psychological studies that were not peer-reviewed, lacked control groups, were confounded, or had a biased experimenter).

[321] Daubert v. Merrell Dow Pharms., Inc., 43 F.3d 1311, 1316 (9th Cir. 1995).

They then need to argue substantively how well the claims in dispute fare on each of the criteria deemed relevant. This will allow a more complete and accurate appraisal of these claims that are currently produced by either the *Frye* or *Daubert* criteria. We then provide a case example by examining a body of claims in psychiatry known as the child sexual abuse accommodation syndrome. Currently, these claims have heterogeneous status regarding admissibility; some jurisdictions using the *Frye* standard find these inadmissible while other jurisdictions come to the opposite conclusion. Similarly, the *Daubert* standard also produces contrary conclusions regarding admissibility. This situation is incoherent and points again to the problems with current admissibility standards.

The polythetic model is a step in the direction of discerning good science and consequently, what questions to ask when in such discernment. Thus, somewhat unfortunately but we believe accurately, the appraisal of science—which is a complex, variegated, and dynamic enterprise—is not simple but rather complex and multidimensional. These recommendations could serve as a useful tool in preliminary trial hearings to determine admissibility of evidence. Further, by incorporating these criteria courts will not only be more likely to reach fair and more accurate outcomes (by weeding out anything but good science as expert evidence) but the application of this model across all states and in federal courts will achieve increased consistency in judicial decisions regarding admissibility.[322]

[322] *See* Robert F. Schopp, *Verdicts of Conscience: Nullification and Necessity as Jury Responses to Crimes of Conscience*, 69 S. CAL. L. REV. 2039, 2054 (1995).