ESSAY

SCIENTIFIC EVIDENCE: GRAND THEORIES AND BASIC METHODS

Curtis E.A. Karnow†

California law requires judges to admit expert scientific testimony without resolving scientific controversies, which are left to juries. But case law does not provide a definition of "science" verses inadmissible pseudoscience. And typically juries are asked to resolve 'scientific' controversies based on studies never provided to them.

The Essay discusses three common definitions of science, finding them insufficient to address the admissibility of scientific expert opinion. The Essay thus argues against reliance on a definition of 'science' and instead recommends the elaboration of criteria for the scientific method including statistical validity, the unreliability of single or preliminary studies, among other criteria for admissibility. In this way, judges, who do have access to the studies, make all significant reliability determinations.

The consequence of this use of scientific methodology as a test for admissibility is an expansion of the "gatekeeper" role of judges in evaluating scientific expert evidence.

✠

> *[T]he essential question. . . is whether or not the theory gives predictions that agree with experiment. It is not a question of whether a theory is philosophically delightful, or easy to understand, or perfectly reasonable from the point of view of common sense.*[1]

INTRODUCTION

This Essay develops a practical test for the admissibility of expert scientific evidence.

Judges must decide what is and is not "scientific," but case law does not provide a useful test. Indeed, California state law

---

† Judge of The California Superior Court (County of San Francisco).

1 RICHARD P. FEYNMAN, QED: THE STRANGE THEORY OF LIGHT AND MATTER 10 (1985).

provides no test at all. Below, I will juxtapose science with 'bad science' or pseudoscience, and in doing so make obvious contrasts between astrology and astronomy, for example, noting the lack of a scientific basis for aromatherapy and other easy targets. But the problem of distinguishing science from pseudoscience is far more serious than these rhetorical contrasts suggest, not only because there are no useful criteria in state law, but because both (a) much commonly used evidence doesn't meet any reasonable criteria for "science" or of the scientific method, and (b) some genuinely "scientific" work is not sufficiently reliable to be admitted as evidence. Courts use supposedly statistically validated algorithms for pretrial release in criminal cases, although the bases for those tools are unknown;[2] drug and alcohol rehabilitation programs are funded by governments, although for many there is no scientific basis to believe they are effective;[3] "pattern evidence" such as ballistics comparisons, fire spread patterns, blood splatter, and bite mark comparisons are used to convict, although there is no scientific basis to believe these tests are reliable.[4]

---

2 Michael Brenner et al., *Constitutional Dimensions of Predicative Algorithms in Criminal Justice*," 55 HARV CIV. RTS.-CIV. LIBERTIES L. REV. 267, 268 (2020) ("More than twenty states use some type of algorithmic modeling to calculate criminal defendants' recidivism risk. Some states use algorithmic models that are developed by private companies and are thus proprietary. These proprietary tools' lack of transparency makes it impossible to determine precisely how their predictive assessments are generated.") (note omitted); Tom Simonite, *Algorithms Should've Made Courts More Fair. What Went Wrong?*, WIRED (Sept. 5, 2019, 7:00 AM) (use in Kentucky), https://www.wired.com/story/algorithms-shouldve-made-courts-more-fair-what-went-wrong [https://perma.cc/RQ3N-CKTJ]; Rachael Myrow, *Report Warns A.I. Algorithms Not Quite Ready for Prime Time in Criminal Justice*, KQED (Apr. 27, 2019), https://www.kqed.org/news/11742529/report-warns-a-i-algorithms-not-quite-ready-for-prime-time-in-criminal-justice [https://perma.cc/7P3J-3GMK].

3 *See infra* note 74.

4 Radley Balko, *Opinion: Judges Are Terrible at Distinguishing Good Science From Bad. It's Time We Stopped Asking Them To Do It*, WASH. POST (Sept. 28, 2017), https://www.washingtonpost.com/news/the-watch/wp/2017/09/28/judges-are-terrible-at-distinguishing-good-science-from-bad-its-time-we-stopped-asking-them-to-do-it [https://perma.cc/44YD-PER8]; Radley Balko & Tucker Carrington, *Bad Science Puts Innocent People in Jail — And Keeps Them There*, WASH POST. (Mar. 21, 2018) (shaken baby syndrome), https://www.washingtonpost.com/outlook/bad-science-puts-innocent-people-in-jail—and-keeps-them-there/2018/03/20/f1fffd08-263e-11e8-b79d-f3d931db7f68_story.html [https://perma.cc/U7N6-QEXZ]. *See generally* Curtis Karnow, CALIFORNIA EXPERT WITNESSES § 2.11, https://works.bepress.com/curtis_karnow/39 [hereinafter EXPERT WITNESSES]. For other examples of what is likely junk science admitted in court, *see e.g.*, Benjamin Taub, *A Large Amount Of Junk Science Is Admitted As Evidence In Court Cases*, IFL SCIENCE, https://www.iflscience.com/editors-blog/a-large-amount-of-junk-science-is-admitted-as-evidence-in-court-cases

The most difficult question in dealing with scientific experts is deciding whether evidence which attacks the reliability of the opinion bars admission of the testimony (a matter for the judge), or is impeachment for admitted opinions (for the jury). A central reason for this difficulty is that the pertinent rules do not cohere. California law holds the judge's task is to distinguish science from non-science, but state law does not provide a test with which to do this. And even were we to apply a test which sorts science from non-science, it would not sort admissible from inadmissible opinions, because much of science is the development of hypotheses and preliminary studies insufficiently reliable for courtroom use.

This Essay first discusses three common theories with which to define "science," concluding they are insufficient for use in the courts. I then turn to a more targeted approach, which is to tease out the criteria of the *scientific method*. This requires the examination of scientific studies, which under current law cannot be done by the jury, but only the judge. The scientific method test described in this Essay likely would bar many opinions which are currently admissible. Thus, the Essay concludes by suggesting an adjustment between the roles of the judge and jury in the evaluation of scientific expert testimony.

Preliminarily, I address the relationship of state and federal law.

I
STATE AND FEDERAL LAW

This Essay cites both California state and federal cases. This may seem indiscriminate. Commentators have noted that California seems more "rigorously conservative" than federal

---

[https://perma.cc/9VJG-9JRZ] ("60 percent of all psychological assessments that are admitted as evidence appear to be based on junk science, although only about 5 percent of these dodgy testimonies are ever challenged by lawyers"); Cynthia Gordy Giwa, *Experts Explore the Consequences of Bad Science on the Justice System*, PROPUBLICA (Oct. 31. 2018, 4:21 PM), https://www.propublica.org/atpropublica/experts-explore-the-consequences-of-bad-science-on-the-justice-system [https://perma.cc/U2B5-R42N]; *The Verdict is in: Courtrooms Seldom Overrule Bad Science*, ASS'N FOR PSYCH. SCI. (Feb. 15, 2020), https://www.eurekalert.org/pub_releases/2020-02/afps-tvi021420.php [https://perma.cc/DF99-WGR2] ("A new, multiyear study published in Psychological Science in the Public Interest (PSPI), a journal of the Association for Psychological Science (APS), finds that only 40% of the psychological assessment tools used in courts have been favorably rated by experts. Even so, lawyers rarely challenge their conclusions, and when they do, only one third of those challenges are successful."). See *infra* note 84 for a discussion of a relatively recent case admitting pattern evidence, *People v. Azcona*, 58 Cal.App.5th 504 (2020).

courts in admitting "new scientific techniques," and more "permissive" than the "demanding federal standards" for other expert testimony.[5]

California has not fully embraced the federal *Daubert*[6] standard for the admission of expert testimony. Specifically, *People v. Kelly*, 17 Cal.3d 24 (1976) remains the law when it comes to admissions of new scientific techniques.[7] But it is also true that in *Sargon*[8] California endorsed *Daubert*'s "gatekeeper" view of the judge's role to keep out unreliable expert testimony:[9] "the *Leahy* court recognized that provisions of our Evidence Code 'seem the functional equivalent' of the Federal Rules of Evidence relied on in *Daubert*,"[10] and outside the *Kelly* context (new scientific techniques), state and federal law are "analogous."[11] This Essay does not examine federal and state law on new scientific techniques, nor with a few exceptions, *Kelly* itself. Instead, it addresses admissibility of "scientific evidence" more generally, where state and federal law are similar. Basic scientific reliability governs here. As the California Supreme Court has noted, whether or not a *Kelly* "new scientific technique" is at issue, putatively scientific evidence must still "comply with the laws of physics, chemistry, and biology. When proffered testimony describes events that exceed or violate these laws, courts must exclude it as unfounded."[12]

---

5 David L. Faigman & Edward J. Imwinkelried, *Wading into the Daubert Tide: Sargon Enterprises, Inc. v. University of Southern California*, 64 HASTINGS L.J. 1665, 1668-70 (2013).

6 Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579 (1993).

7 People v. Leahy, 8 Cal.4th 587, 591 (1994); O'Neill v. Novartis Consumer Health, Inc., 147 Cal.App.4th 1388, 1397 (2007). "New scientific techniques" in this sense are "novel devices or processes," (People v. Hill, 191 Cal.App.4th 1104, 1124 (2011)), and might address polygraphs (People v. Wilkinson, 33 Cal.4th 821, 843 (2004)), various sorts of DNA evidence (People v. Smith, 107 Cal.App.4th 646, 666 (2003)) ("deoxyribonucleic acid evidence"), but not fingerprint evidence, for example (*In re O.D.*, 221 Cal.App.4th 1001, 1007 (2013)), nor expert testimony about gangs (*Hill*, 191 Cal.App.4th at 1123-24).

8 Sargon Enters., Inc. v. Univ. of S. Cal., 55 Cal.4th 747 (2012).

9 *Id.* at 772. But this does not affect the state's reliance on the *Kelly* test. *Id.* at 772 n.6. *See generally, e.g.*, People v. Azcona, 58 Cal.App.5th 504 (2020).

10 People v. Mitchell, 110 Cal.App.4th 772, 784 (2003).

11 Apple Inc. v. Superior Court, 19 Cal.App.5th 1101, 1119 (2018). *See also* Cooper v. Takeda Pharms. Am., Inc., 239 Cal.App.4th 555, 590 (2015) (citing *Daubert* outside in context of new scientific techniques). *See* Faigman & Imwinkelried, *supra* note 5, at 1684-85) (aside from the *Kelly* context, California moving towards *Daubert* standard); James R. Dillon, *Expertise on Trial*, 19 COLUM. SCI. & TECH. L. REV. 247, 312 (2018); Eric Klein & Jessalee Landfried, *Does the Dose Still 'Make the Poison' Under Daubert?*, 31 NAT. RES. & ENV'T 1, 3 ("*Sargon* . . . appeared to adopt a standard similar to *Daubert*.").

12 People v. Jackson, 1 Cal.5th 269, 320 (2016).

II
THREE THEORIES

Appellate opinions regulating the admission of scientific expert testimony bar judges from deciding among competing scientific theories; that, we are told, is a job for the jury.[13] The judge's job as a 'gatekeeper' straining out unreliable testimony is limited to rejecting *unscientific* theories (as well as opinions which are conclusory, or where no logic ties the assumptions, theories, and facts to the opinions, etc.).[14]

But as I was drafting an informal note on handling expert witnesses,[15] I found that California state law reveals no test to determine if a theory, hypothesis, or expert conclusion is "scientific."

A. Popper Falsification

Federal law in *Daubert*,[16] for instance, cites Sir Karl Popper's formulation, which is to ask whether a theory is falsifiable by observation, such as by experiment.[17] Under this criterion, a judge might ask the proposed expert what tests would falsify a theory: if there aren't any, the theory isn't scientific. For example, Einstein's general theory of relativity calls for light to bend around gravitational centers such as stars; an experiment could be conducted to see if the light from a distant galaxy is bent around an intermediate star such that it is visible even when literally behind the star (gravitational lensing).[18] A theory that a chemical has an effect on the body can be tested by having a random group ingest the chemical and a different random group (the control group) not ingest it,

---

13 "The court does not resolve scientific controversies." *Sargon Enters.,* 55 Cal.4th at 772; *Cooper*, 239 Cal.App.4th at 590.

14 *Sargon Enters.*, 55 Cal.4th 747 (2012).

15 EXPERT WITNESSES, *supra* note 4.

16 *Daubert*, 509 U.S. at 593.

17 E.g., FED. R. EVID. 702, Advisory Committee's Notes; Bitler v. A.O. Smith Corp., 400 F.3d 1227, 1235 (10th Cir. 2005). KARL POPPER, CONJECTURES AND REFUTATIONS: THE GROWTH OF SCIENTIFIC KNOWLEDGE 37 (5th ed. 1989) ("[T]he criterion of the scientific status of a theory is its falsifiability, or refutability, or testability") (emphasis deleted*). See generally* Gary Edmond & David Mercer, *Conjectures and Exhumations: Citations of History, Philosophy and Sociology of Science in US Federal Courts*, 14 L. & LITERATURE 309 (2002).

18 Just such an experiment was conducted. Light bent; the theory was not falsified. Peter Coles, *Einstein, Eddington and the 1919 Eclipse,* NATURE (Apr. 17, 2019), https://www.nature.com/articles/d41586-019-01172-z [https://perma.cc/5T32-ZNRE]; *see also* Bianca Nogrady & Anna Salleh, *Einstein's 'Impossible' Hope: Light Bending Theory Directly Observed in Distant Stars for First Time*, ABC SCI. (June 7, 2017), https://www.abc.net.au/news/science/2017-06-08/einsteins-impossible-dream-comes-true/8598552 [https://perma.cc/P9CF-PRNJ].

and then seeing if there is a difference. The theory would be falsified if there's no difference.[19] Other theories, such that there are invisible fairies who live under plants and affect our luck, are not subject to these sorts of tests, so while they are fun, they are not scientific.

No published California case mentions Popper's falsifiability criterion. One of the most important cases in this state on scientific evidence is *People v. Kelly,* 17 Cal. 3d 24 (1976), which as it sets out the test for the admissibility of new scientific methods uses the phrases "scientific technique," "correct scientific procedures," "scientific principle or discovery," "relevant scientific community," "scientific writings," and "the standard of scientific acceptance and reliability," 17 Cal. 3d at 30, 33, 34. *Kelly* seems to require a "typical cross-section of the scientific community," and "any opposing sentiment in the relevant scientific community," id. at 37, in order for the judge to evaluate the viability of a new test. But "scientific" is never defined, as the state Supreme Court subsequently noted.[20] Other important state cases similarly take for granted the meaning of "scientific."[21]

The definition of "science" is significant, not just in the *Kelly* context regarding new scientific techniques, but in the more common *Sargon* context as well, that is, when an expert provides an opinion on scientific issues such as disease causation and treatment, environmental and workplace contamination, or principles which explain design defects in machines. In that more common expert context, courts decide (among other things) if the materials relied on by the expert

---

19 A theory can be falsifiable, and thus scientific, but just wrong—as when it's falsified. For example, the Bohr model (1913) pictured a nucleus with orbiting electrons. Experiments including on the spectra of large atoms, fine structures in spectral lines, and with multi-electron atoms generally, together with the uncertainty principle, revealed shortcomings. I mention phrenology in this connection at *infra* note 72.

20 "While the standards imposed by the *Kelly/Frye* rule are clear, the definition of a 'new scientific technique' is not." People v. Stoll, 49 Cal.3d 1136, 1155–56 (1989). *See also* People v. Lucas, 60 Cal.4th 153, 223 (2014); *disapproved of on other grounds by* People v. Romero and Self, 62 Cal.4th 1 (2015).

21 *E.g.*, People v. Venegas, 18 Cal.4th 47, 95 (1998); *Leahy*, 8 Cal.4th at 607 (which does address the term "scientific" but in an entirely different way for a different reasons-to tease out under *Kelly*, the "scientific" tests which by reason of their application and fancy sounding titles might unduly sway a jury to blindly adopt their results); *Azcona*, 58 Cal.App.5th at 512 (finding "defendant [did not] establish[] what the relevant scientific community is"); *In re* Jordan R., 205 Cal.App.4th 111, 130 (2012) ("We conclude that the juvenile court reasonably relied on the scientific literature in reaching its conclusion there was significant controversy within the relevant scientific community about the reliability of polygraph examination at the present time.").

"are of a type that reasonably may be relied on by an expert in forming an opinion" (EC 801(b)). But usually it's an expert who says what those "relied on" materials are; the inquiry presumes scientific validity. Without a workable sense of 'scientific,' it can be hard to figure out which expert should testify on even preliminary issues such as what texts and treatises are authoritative, which apparent academic degrees really mean something, which journals should count as having reliable peer reviewed articles (astrologers have their own peer review journals), as well as which materials are reasonably relied on in the discipline.[22] The case of *Jackson*[23] may be as close as we get in state law to some idea of what 'scientific' might entail, but that case just refers to the "laws of physics, chemistry, and biology," telling judges to exclude testimony on events which "exceed or violate these laws."

Should California adopt the *Daubert* standard—Popper's falsifiability approach? There are advantages: it's a plausible approach, and it's nice to have consistency between state and federal law. Falsifiability does a reasonable job distinguishing scientific disciplines like astronomy and physics from those which aren't, like astrology.

B. Kuhn's Paradigms

But falsifiability itself is a theory about science, and it is not the last word on how to define scientific disciplines. There are a lot of theories on how science develops, and what thus should count as "scientific."[24] Many of us will recall Thomas Kuhn's work.[25] He points out that, in practice, scientists do not actually reject an entire theory because one experiment seems to falsify it. Anomalous results are met with agnosticism; ideas are passed about as to why the results might have occurred; suspicion may fall on a piece of equipment, the motives and incentives of the investigators, or something funny about the statistical models used; perhaps no one can replicate the results and they perish in oblivion; sometimes the results are just ignored. Kuhn noted, "Newton's laws were retained despite the fact that they were contradicted

---

22 I raise this problem in more detail, and discuss the essentially self-referential law, in my informal note on expert testimony. EXPERT WITNESSES, *supra* note 4, at § 2.10.4.1.

23 *Jackson*, 1 Cal. 5th at 320.

24 Chaomei Chen, *Science Mapping: A Systematic Review of Literature,* 2 JDIS 1 (2017), https://content.sciendo.com/view/journals/jdis/2/2/article-p1.xml [https://perma.cc/HL5P-EH4K].

25 *See generally* THOMAS KUHN, THE STRUCTURE OF SCIENTIFIC REVOLUTIONS (2d ed. 1970).

for decades by the motions of the perihelion of Mercury and the perigee of the moon."[26] It is only after a substantial series of different experiments, over time, produce results which are incompatible with a theory, which then may produce what Kuhn termed a "crisis,"[27] which may result in a "paradigm shift" to what appears to be a new scientific theory. "Kuhn argued that falsification is no more possible than verification; each process wrongly implies the existence of absolute standards of evidence, which transcend any individual paradigm."[28] Imagine that the gravitational lensing experiment did *not* reveal bent light rays: would we have abandoned the theory of relativity? No. It had by that time already proven explanatory on other fronts, integrated into general physics. Much more would have been required to set it aside. We would, instead, have found other reasons to explain away the experimental results.

Popper's falsification theory holds no scientific theory can be definitely proven true (but only falsified); but at the same time, it entails a belief in the accuracy (or truth) of the falsifying experiment. This is awkward. For Kuhn, the shift to a new hypothesis or theory is not compelled as a matter of logic, nor by its timing; it depends on many factors, including how much ambiguity its practitioners can endure, the perceived seriousness of the anomalies and the perceived power of a proposed new paradigm to explain phenomena; even the personalities of adherents. Sometimes science changes just as older practitioners die off, replaced by those who profess a different approach.[29]

---

26 Mano Singham, *The Idea that a Scientific Theory Can Be 'Falsified' is a Myth,* SCI. AM. (Sept. 7, 2020), https://www.scientificamerican.com/article/the-idea-that-a-scientific-theory-can-be-falsified-is-a-myth [https://perma.cc/33YT-7XJ8].

27 *Thomas Kuhn*, STAN. ENCYCLOPEDIA PHIL., https://plato.stanford.edu/entries/thomas-kuhn/ [https://perma.cc/TE4A-TJFT].

28 John Horgan, *What Thomas Kuhn Really Thought About Scientific "Truth,"* SCI. AM. (May 23, 2012), https://blogs.scientificamerican.com/cross-check/what-thomas-kuhn-really-thought-about-scientific-truth [https://perma.cc/TEF9-48KP].

29 "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 . . . An important scientific innovation rarely makes its way by gradually winning over and converting its opponents: it rarely happens that Saul becomes Paul. What does happen is that its opponents gradually die out, and that the growing generation is familiarized with the ideas from the beginning: another instance of the fact that the future lies with the youth." MAX PLANCK, SCIENTIFIC AUTOBIOGRAPHY AND OTHER PAPERS 33, 97 (1949), reported at https://en.wikiquote.org/wiki/Max_Planck [https://perma.cc/SUM5-H64R]. *See e.g.*, Dalmeet Singh Chawla, *Science Really*

Kuhn has, in turn, been attacked. He suggests that no scientific theory is closer to an objective "truth" than any other; they simply compete and are eventually superseded, one after the next. Aristotle, for example, was not wrong (says Kuhn), nor less right than we are about physics; he just had a different scientific paradigm than we do.[30] Most physicists would disagree: they would say that quantum physics, for example, is closer to the truth, because it explains more, works in practice, and so on. Others have observed that shifts in the sciences are more nuanced and progressive than Kuhn, who advocated for fundamental incommensurability among scientific paradigms,[31] suggests. Science does not blow up and collapse on a scale similar to an entire world view; or if it does, that comes very, very rarely. Rather, specific hypotheses are engaged, rejected, accepted, or modified in an incremental process. The reason we might reject a hypothesis almost always has nothing to do with a paradigm shift. Many advances, such as the mechanisms of DNA and molecular biology, did not erupt as a function of paradigm shifts: they were discoveries which did not necessarily countermand previously accepted hypotheses. Global warming, for example, is now an accepted scientific fact, but this is not because we had a revolution, incompatible with prior doctrine, in thinking about the effect of humans on the environment. It's because we have more data than we did even a hundred years ago, and we can use traditional analyses to understand what it means.[32] So too with superconductivity, the remarkable property of some materials at (usually, but not always) very low temperatures to pass electrical charges with almost no resistance. The effect was initially shocking and seemed to contradict then current understandings, but the puzzle was later solved by a series of hypotheses and testing experiments, all of which relied on, but expanded, then-current views on the behavior of electrons.[33]

Does Advance One Funeral at A Time, Study Suggests, CHEMISTRY WORLD (Sept. 12, 2019), https://www.chemistryworld.com/news/science-really-does-advance-one-funeral-at-a-time-study-suggests/3010961.article [https://perma.cc/F4D9-5KZ4].

30 *E.g.*, *Was Kuhn More Wrong Than Right?*, ANTIMATTER (Feb. 1 2011), https://antimatter.ie/2011/02/01/was-kuhn-more-wrong-than-right/ [https://perma.cc/5MAG-U5T2].

31 *The Incommensurability of Scientific Theories*, STAN. ENCYCLOPEDIA PHIL., https://plato.stanford.edu/entries/incommensurability/ [https://perma.cc/28GY-CBNF].

32 *E.g.*, *Climate Change: How Do We Know?*, NASA, https://climate.nasa.gov/evidence/ [https://perma.cc/Q3CS-KFPL].

33 *E.g.*, David Goodstein & Judith Goodstein, *Richard Feynman and the*

C. Quine's Field

Avoiding most of the issues outlined above, W.V. Quine and Pierre Duhem are credited with the so-called Duhem–Quine thesis, which agrees that no experiment tests a proposition in isolation because every such test assumes the truth of background or auxiliary hypotheses; the result of every "failed" test can thus be handled either by believing the tested hypothesis is false or that one or more auxiliary hypotheses is false, similar to the approach taken by some critics of Kuhn. As a matter of practice, we usually implicitly agree that the background of auxiliary assumptions is correct; an experiment can then be conducted which tests the validity of an isolated hypothesis, and we can then decide whether the hypothesis has or has not survived the experiment. Quine writes that "our statements about the external world face the tribunal of sense experience, not individually, but only as a corporate body."[34] Citing Duhem, Quine analogizes to a force field where "a conflict with experience at the periphery occasions readjustments in the interior of the field,"[35] saying "any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system."[36] Because empirical statements are interconnected, they cannot be singly disconfirmed; if we wish to hold a particular statement true, we can always adjust another statement.[37] For convenience, I will call this—my simplified version of the Duhem–Quine thesis—the Field theory, implying a collection of connected beliefs and statements which are more or less easy to modify, causing more or less modifications of beliefs, at more or less distance from that initially modified. (By 'more or less easy to modify,' I mean that we hold some beliefs closer than others: the existence of gravity for example will not be upended when we learn that astronauts in orbit are weightless: we will eventually find explanations for that phenomenon consistent with gravity. But it won't take much to convince me that red wine is better for me than white wine, or vice versa, and I might easily change my mind from time to time.)

For example, we can imagine an experiment which seems

History of Superconductivity, 2 PHYSICS IN PERSP. 30 (2000), https://web.njit.edu/~tyson/supercon_papers/Feynman_Superconductivity_History.pdf [https://perma.cc/7J6N-24SM].

34 W. V. QUINE, *Two Dogmas of Empiricism*, in FROM A LOGICAL POINT OF VIEW 41 (Harvard 1953).

35 *Id.* at 42.

36 *Id.* at 43.

37 Some of this wording is from *Pierre Duhem*, STAN. ENCYCLOPEDIA PHIL., https://plato.stanford.edu/entries/duhem [https://perma.cc/X4UM-M3D6].

to show, unexpectedly, that carbon from fires causes, say, glaucoma, which we had believed is caused by high pressure in the eye. We have many choices here. First, perhaps the experiment was done badly, either because the equipment was faulty or the statistics used to generalize from the sample were mismanaged; perhaps the people doing the experiment were bribed to fake the results, etc. Perhaps the one experiment is an outlier, the results a matter of chance (that happens even with well-done studies), and no other study replicates it.[38] All these beliefs are preferable to giving up our belief that high pressure causes glaucoma, because the latter belief is deeply entrenched, and it will take a lot of weighty evidence to displace it and its associated networked beliefs. Instead of displacing our cherished beliefs about eye pressure, we may decide that carbon ingestion is simply one way in which eye pressure is raised, thus reconciling hypotheses. But to do that, we also need hypotheses about how carbon in the bloodstream, for instance, can affect eye pressure, implicating a large series of interconnected beliefs, which themselves will be more or less resistant to tampering.

Actually, we often reject new results. As Carl Sagan is credited with saying, extraordinary claims require extraordinary evidence. If the new hypotheses were not that carbon caused glaucoma, but, for example, looking at whether horses caused glaucoma, we would expect a truly stunning amount of persuasive evidence, because assenting to the horse hypotheses would require us to revamp a very long interconnected series of powerfully held beliefs—a field of beliefs—about how the eye works, disease mechanism, that horses are no different in this context than any other object, the physics of photons, and so on. Under Field theory, explanations for unusual experimental results are sought first among ordinary, apparently reasonable causes, and only after those explanations have been discredited (because, for example, repeated experiments have shown those seemingly reasonable causes are not related to the results) might we adjust more cherished beliefs.[39] Science in this view is in a

---

38 Quine notes, "The scientist's position is peculiarly delicate when, as here, he must decide whether to accept the testimony of his own senses to a revolutionary phenomenon, challenging entrenched scientific theory, or to dismiss the phenomenon as a presumed effect of commonplace causes which he has merely not had the wit to think up." W.V. QUINE, QUIDDITIES: AN INTERMITTENTLY PHILOSOPHICAL DICTIONARY 8 (Harvard: 1987).

39 *See* Richard Feynman, *The Unscientific Age*, in THE MEANING OF IT ALL 67 (1998). Sometimes we don't know what to believe and wait for more evidence. *E.g.*, the "Havana syndrome" when, starting in 2016, US diplomats in Cuba and

state of constant rejiggering of beliefs, sometimes displacing one with another, sometimes rephrasing a tenet with a more precisely drawn one; placing more or less weight on certain beliefs; or restating them to be more generally true after an observation has called into question an earlier view.[40] The field has edges, where it touches what we can see and hear, and it blends imperceptibly towards its centers, where we hold more abstract beliefs and theories which seem to explain what we can see and hear. The areas communicate: discernment is "theory-laden,"[41] in that we see what we are primed to see; we observe in categories, discerning what we are trained to see by the theories in which we believe. And contrariwise, sometimes our observations, however seemingly trustworthy, must give way to either other observations or to deeply held, deeply cherished abstract beliefs.[42]

Here's a nice example:

> [I]n *Truck Ins. Exch. v. MagneTek Inc.,* 360 F.3d 1206, 1211–1212 (10th Cir.2004), plaintiff attempted to introduce evidence of a novel theory, "pyrolysis," which hypothesized that wood could ignite at temperatures much lower than normal under particular circumstances. We affirmed the district court's decision to exclude this evidence because plaintiff's experts had failed to test their novel theory sufficiently to demonstrate its scientific reliability. *Id.* at 1213. When an expert proposes a theory that modifies otherwise well-established knowledge about regularly occurring phenomenon, such as the normal ignition temperature of wood, we would expect the importance of testing as a factor in determining reliability to be at its highest. Here, by contrast, plaintiffs' experts propose a theory about how the accident occurred given the known science of copper sulfide particulate contamination as a cause of propane gas leaks. What distinguishes the present case is that the need for testing is not at its highest because the reliability of the science of copper sulfide contamination

---

their families reported neurological injuries, such as buzzing noises followed by pain, vertigo, and difficulty concentrating. *Cuba Unexplained Events investigation – Final Report*, CTRS. FOR DISEASE CONTROL & PREVENTION (Dec. 3, 2019), https://beta.documentcloud.org/documents/20444983-foia-21-00330-final-report1 [https://perma.cc/GS8Q-ZKZH].

40 For example, the hypothesis that 'all living animals breathe' is displaced after observation by a more general hypothesis that 'all animals use oxygen,' which then then accounts for fish.

41 This is associated with both Kuhn and Quine. THOMAS KUHN, THE STRUCTURE OF SCIENTIFIC REVOLUTIONS 111-135 (1996); W.V. QUINE, *Empirical Content*, in THEORIES AND THINGS 25 (1981).

42 W.V. QUINE, *Five Milestones of Empiricism*, in THEORIES AND THINGS 71 (1981).

> is not in dispute, and thus the district court did not abuse its discretion in finding that the presence of a screen did not alter the reliability of the fundamental science.[43]

In this way, Field theory seems a good candidate with which to examine so-called scientific beliefs. It accounts for experiments and how they may be accepted and rejected, it acknowledges the interdependence of our beliefs about the world, and it solves problems with both Popper's and Kuhn's classic approaches.

But Field theory (at least the simplified picture I present) is logically consistent with psychosis, or indeed ordinary irrationality. The interconnected field of beliefs may be about space aliens, lizard people, and the rest. Above, I had assumed that what counts as a connection among beliefs is a matter of logic, but that's not necessarily true. There's iron in blood, but this doesn't mean you can steer blood with a magnet. People believe things irrationally, and one can maintain a vast web of beliefs which are connected by irrational association. Ordinary people, including scientists, are affected by factors like confirmation bias[44] and cognitive dissonance.[45] We should not underestimate this. A recent poll

> asked respondents whether they believe that "a group of Satan-worshipping elites who run a child sex ring are trying to control our politics and media"—the false allegation at the heart of QAnon. While only 17% said it was true, another 37% said they didn't know.[46]

---

43 *Bitler*, 400 F.3d at 1235–36 (10th Cir. 2005).

44 *Hidden Brain: Facts Aren't Enough: The Psychology of False Beliefs*, NPR (July 22, 2019), https://www.npr.org/transcripts/743195213 [https://perma.cc/N7TS-EFSK].

45 *All Things Considered: Why Many People Are Drawn to Conspiracy Theories*, NPR (Aug. 26, 2020), https://www.npr.org/2020/08/26/906333307/why-many-people-are-drawn-to-conspiracy-theories [https://perma.cc/R9F3-3DHP].

46 *Even if it's 'Bonkers,' Poll Finds Many Believe QAnon and Other Conspiracy Theories*, NPR (Dec. 30, 2020), https://www.npr.org/2020/12/30/951095644/even-if-its-bonkers-poll-finds-many-believe-qanon-and-other-conspiracy-theories [https://perma.cc/FB6X-HME3]. *See also* a 2013 poll, which asked, among other things, "Do you believe that shape-shifting reptilian people control our world by taking on human form and gaining political power to manipulate our societies, or not?" 4% said yes. But even better, 7% responded "Not sure." This is marvelous: *not sure*? *Democrats and Republicans Differ on Conspiracy Theory Beliefs,* PUB. POL'Y POLLING (Apr. 2, 2013), https://www.publicpolicypolling.com/wp-content/uploads/2017/09/PPP_Release_National_ConspiracyTheories_040213.pdf [https://perma.cc/U2AS-SCH9] (The margin of error was ±2.8). Philip Bump, *12 Million Americans Believe Lizard People Run Our Country*, ATLANTIC (Apr. 2. 2013), https://www.theatlantic.com/national/archive/2013/04/12-million-americans-believe-lizard-people-run-our-country/316706

Pause for a minute. That's over half the population at least uncertain whether Satan-worshipping pedophiles are trying to control the country. Some explanations strongly suggest that those who maintain these irrational beliefs do so precisely *because* of Field theory: the need to have an apparently consistent set of beliefs which explain all phenomena.[47]

The point of course is Field theory alone doesn't tell us what counts as scientific. As with Popper's falsification theory and Kuhn's theory of gradual encroachment on accepted doctrines subject to relatively sudden "paradigm" shifts, Field theory is useful but insufficient to distinguish admissible scientific theories from inadmissible "non" science.

It's not easy to define science.[48] So legal tests that just tell trial judges to not decide scientific controversies, but to admit science and not admit what isn't science—aren't very helpful.

III
SCIENTIFIC METHOD

I suggest that a more persuasive approach, and one more useful to lawyers and judges, is to set out what we mean by the *scientific method*. A discipline which uses the scientific method would then be 'scientific.' Also, a study or experiment within a discipline which we might otherwise call science would be rejected if the study or experiment did not, in fact, use the scientific method. With this approach, we might still invoke falsifiability and Field theory; but we constrain them.

Our hint here comes from Quine, who wrote "The edge of the system [of beliefs] must be kept squared with experience."[49] The scientific method respects and reflects *experience*; it understands that we can be fooled, mistaking cause for effect, correlation for causation, statistical significance for degree or size of effect. We make irrational associations among ideas which have no logical relationship. But science follows the rules of logical entailment: if a theory is valid, we calculate what else must be true, and then see if that is so. Science

---

[https://perma.cc/DQC3-6QN2].

47 Jan-Willem van Prooijen, *The Psychology of Qanon: Why do Seemingly Sane People Believe Bizarre Conspiracy Theories?*, NBC NEWS (Aug. 13, 2018), https://www.nbcnews.com/think/opinion/psychology-qanon-why-do-seemingly-sane-people-believe-bizarre-conspiracy-ncna900171 [https://perma.cc/UKE6-C836].

48 RICHARD P. FEYNMAN, *The Uncertainty Of Science*, in THE MEANING OF IT ALL 4 (1998); DAVID HECHT, *Pseudoscience and the Pursuit of Truth*, in PSEUDOSCIENCE: THE CONSPIRACY AGAINST SCIENCE 11 (2019) (shifting definitions of science).

49 W. V. QUINE, *Two Dogmas of Empiricism*, in FROM A LOGICAL POINT OF VIEW 45 (1953, 1961).

experiments; it tries things out; it guesses, it has preliminary, informal postulates, often wrong, but worthy nevertheless of pursuit; it makes changes to cherished beliefs grudgingly, seeks simplicity of explanation but resorts to complexity when simplicity is misleading; it checks and double checks itself; it is honest, transparent, and public. Above all, it honors reality; fact.

This section maps out practical criteria to distinguish science from pseudoscience and what I have called "Bad Science."[50] I do so by offering criteria for the scientific method.

There are four parts following this introductory comment. First, I provide general cautions on criteria that should not be used, despite the fact that some case authority suggests we do so. This demonstrates that my arguments are not all consistent with current law.

Next, I suggest criteria a judge might use in her gatekeeping capacity, deciding whether or not to admit supposedly scientific testimony. This is not easy. There is no one criterion for the scientific method: it's a constellation, and many factors are present to a greater or lesser degree. Peer review studies are usually essential, but not always,[51] and even with those studies, that's not the end of it; more is needed (a lot of "peer reviewed" journals are junk). Nor does every study distinguish statistical significance from effect, but that alone may not invalidate a study. Not every result is the subject to a meta-analysis, but that's probably not disqualifying. Statistical transparency may be a matter of degree, as is whether a result has been replicated. Some, but not all, of the underlying data may be available. Some studies have true randomization, as they allocate subjects between the test and control groups; others are not so good, with more or less room to stack the deck, as it were. Even the abstract Field theory–used as a test to determine if a scientific hypothesis should be adopted—is a matter of degree, because hypotheses fit more or less with the rest of the field: it is not a binary matter.[52]

The third part below outlines areas of impeachment—issues to be aired before the jury, after the judge has decided to admit the testimony.

---

50 I purloin the phrase from Ben Goldacre, BAD SCIENCE (2008). *See* EXPERT WITNESSES, supra note 4, at § 2.10.4.

51 *E.g.*, Clausen v. M/V NEW CARISSA, 339 F.3d 1049, 1060 (9th Cir. 2003).

52 "We have to find a new view of the world that has to agree with everything that's known, but disagree in its predictions, some way. Otherwise it's not interesting. And in that disagreement, agree with nature." RICHARD P. FEYNMAN, *Seeking New Laws*, in THE CHARACTER OF PHYSICAL LAW 165 (1994).

The line between the second and third parts is arguable: reasonable people might use a criterion from one and instead use it in the other. Indeed, this line between admissibility and impeachment is one of the most difficult problems in this area of the law,[53] and I turn to it in the fourth and concluding part of this section. As with a lot of the other choices trial judges make, many decisions are a matter of discretion—but, we hope, informed discretion. And we can always wish for more guidance from the appellate courts or legislature. In the meantime, I offer a few suggestions to get the ball rolling.

A. Science—and Science for Courts

As I've suggested, there are many approaches to defining science; and there are many criteria, most of them evident to various degrees, for the scientific method. It would be presumptuous for me to contribute to the debate at those general levels.

The concern of this Essay is more narrow: is the discussion of the scientific method sufficiently reliable to be introduced in court. Scientists engage in many activities which are 'real science' but which are not reliable enough to be the basis for a conviction in a criminal case, or the court-ordered transfer of wealth from one person to another in a civil case. Scientists guess, speculate, and begin the long journey to discovery with preliminary studies, in vitro experiments and animal studies. Perhaps substances pass initial screening tests and so are termed potential carcinogens, but without evidence that they have any adverse effect in humans. Perhaps a formal study is conducted—a perfectly good one, let us say—only to be contradicted by replication studies. A faster-than-light neutrino is discovered by eminent scientists using state of the art equipment, but months later an equipment defect is found. Cold fusion by 'real scientists' is announced, but the 'discovery' is never confirmed. "So what we call scientific knowledge today is a body of statements of varying degrees of certainty."[54]

This is the way of science, and all of these studies and reports may be the work of honest science; but that does not mean the results are reliable enough for court. Of course, courts take guidance from scientists, but the issue is for judges, and the subject of this Essay is a legal one, not exactly a scientific one.

---

53 EXPERT WITNESS, *supra* note 4, at § 2.10. *See e.g.*, David S. Caudill & Richard E. Redding, *Junk Philosophy of Science?: The Paradox of Expertise and Interdisciplinarity in Federal Courts*, 57 WASH. & LEE L. REV. 685, 722 (2000).

54 FEYNMAN, *supra* note 48, at 27.

It follows that courts must not just ask if a hypothesis, study, or report is "scientific." Contrary to the implication in opinions that juries may be asked to choose between any scientific theories and rely on whichever one they like,[55] juries should not be exposed to every scientific study or report, but only those which pass a test of scientific and legal reliability. The balance of this Essay offers an effort towards that test.

B. Cautions

1. *Not Every Issue Has Two Sides*

Journalism falls victim to this fallacy: the writer often interviews a mainstream proponent and then finds someone, somewhere, to disagree, all in an effort to seem fair and balanced—for what is balance without weight on two sides?[56] Courts tend to do this as well. We term many trials a "battle of experts," sometimes with a trace of cynicism. The calls to let the jury decide scientific issues too seem girded by this sense that there are at least two sides to every disagreement, and both should face the jury. In court, there is a felt entitlement to be able to put on an opposing case—the basic structure of a trial confirms this. Judges about to exclude evidence—including for example an expert—are often met with the incredulous plea that such a ruling will destroy a case or defense, as if that alone were a reason to let the evidence in. But if one expert says force equals mass times acceleration, and another says force equals mass times mental energy, only the first one gets to testify.

2. *Beware the Stand-Alone Expert*

There are three related problems here.

A. First, many experts have remarkable pedigrees: they went to first-rate graduate schools, have decades of experience in demanding jobs; perhaps they run a lab; and are credited with a mass of published papers. They are clearly experts of something; but the question remains: experts of what? Many

---

55 The now famous statement from *Sargon* that the "court does not resolve scientific controversies" means the jury does. *Sargon*, 55 Cal.4th at 772. This has been taken to mean that all issues which are the "subject of legitimate scientific debate" go to the jury and may not be excluded by the judge. Davis v. Honeywell Internat. Inc., 245 Cal.App.4th 477, 480 (2016).

56 *See e.g.*, SHAWN OTTO, THE WAR ON SCIENCE 21 (2016). Indeed, Otto notes that the war on science, fed by ideological, religious, and commercial (i.e., tobacco anti-science) interests, often depends on the creation of false equivalences, such as suggesting, for example, that both evolution and intelligent design be taught in schools.

scientific experts are called on to analyze research papers by others, but the witnesses' expertise may in fact be in clinical work or something else. She may know what work has been done in the field, but may or may not also be able to critically analyze and summarize the research.

B. An expert may be asked to voice an opinion based on his own expertise and experience—which seems pretty ordinary, until one realizes it may be a covert way to relay the opinions of others, the undisclosed authors of undisclosed research.[57]

C. An expert may give an opinion based on his undifferentiated experience and expertise. Ben Goldacre calls this the fallacy of relying just on the "eminence" of the witness,[58] as opposed to relying straight out on identified research.[59] Eminence of the speaker is among the weakest bases for believing a hypothesis true,[60] but courts have required trial judges to rely on this eminence despite problems with the cited research.[61] Courts do this in understandable deference to witnesses who obviously know more about a field than a judge does, but the deference may be misguided: The

---

57 *E.g.*, Strobel v. Johnson & Johnson, 69 Cal.App.5th 34, __, 284 CR3d 165, 186 (2021), review filed Nov. 1, 2021.

58 BEN GOLDACRE, YOU'LL FIND IT'S A BIT MORE COMPLICATED THAN THAT 4, 17, 203, 402 (2014) (hereinafter YOU'LL FIND IT); Wells Mangrum & Richard Collin Mangrum, *Evidence-Based Medicine in Expert Testimony*, 13 LIBERTY U. L. REV. 337, 357 (2019) ("an expert's 'experiential' opinion without evidence-based research studies to validate or refute, is ordered the least reliable of all expert opinions").

59 Eminence and associated experience are however perfectly good bases for *non*-scientific experts, such as architects, plumbers, doctors and lawyers.

60 One commentator notes a series of prestigious medical schools (such as the University of Michigan) where doctors and professors with impressive degrees teach pseudoscience. This is part of what the author describes as the encroachment of science by "magical thinking" practices via inroads with governmental licensing agencies, positions at respected institutions, the creation of professional societies, and so on. DAVID GORSKI, *'Integrative' Medicine: Integrating Quackery with Science-Based Medicine*, in PSEUDOSCIENCE: THE CONSPIRACY AGAINST SCIENCE 309 (2019). See also, *e.g.*, SETH KALICHMAN, *'HIV Does Not Come from AIDS,' A Journey into AIDS Denialism*, in PSEUDOSCIENCE: THE CONSPIRACY AGAINST SCIENCE 429 (2019) (noting a professor at Virginia Polytechnic Institute with positions in other universities in Sydney, Michigan, Southampton and Kentucky—who professes AIDS denial and authority on Loch Ness monster). Even Fair Harvard is not immune: *see* Colleen Walsh, *Panel Discusses History, Future of Alternative Therapies*, THE HARV. GAZETTE (Apr. 3, 2008) (Harvard medical professor touts nineteenth century studies on homeopathy), https://news.harvard.edu/gazette/story/2008/04/panel-discusses-history-future-of-alternative-therapies [https://perma.cc/35Y7-4XRH]. It is dangerous to admit evidence just on indicia of the proponent's apparent academic eminence.

61 Wendell v. GlaxoSmithKline LLC, 858 F.3d 1227, 1235-36 (9th Cir. 2017); *Cooper*, 239 Cal.App.4th at 589–90.

issue is not who knows more about cancer or nuclear physics. The issue is whether judges can figure out if there's an adequate basis for an opinion and whether it's based on reliable science, which of course judges must do—it's part of their job description.[62] It is not enough to have a good pedigree. As I'll explore further, a central problem is that appellate opinions conceive the issue to be one of admitting a *witness's testimony*, and so focus on the witness' credentials. But in this area, the issue is usually the reliability of the *scientific studies*, not so much the credibility of the witness.

3. *Anecdotal Evidence Isn't*

Case studies are by definition individual stories; they are the opposite of the peer reviewed papers discussed below. Among other things, they cannot pretend to statistical significance. They are preliminary, used to put a face on a hypotheses,[63] and to explore: "Case studies have often been viewed as a useful tool for the preliminary, exploratory stage of a research project, as a basis for the development of the 'more structured' tools that are necessary in surveys and experiments."[64] They might suggest a hypothesis is plausible, and they have a highly significant role in scientific research, but they are not a reliable basis for opinion in court.[65] They are a device of rhetoric; perhaps the jury might hear about them as illustrations, but they are not a basis for expert

---

62 *Sargon Enters.,* 55 Cal.4th.

63 Nicolaj Siggelkow, *Persuasion with Case Studies*, 50 ACADEM. OF MGMT. J. 50 (2007), https://www.researchgate.net/profile/Nicolaj_Siggelkow/publication/234021898_Persuasion_With_Case_Studies/links/00b7d5339bd9404622000000.pdf [https://perma.cc/F9JS-5PFW].

64 Jennifer Rowley, *Using Case Studies In Research*, 25 MGMT. RSCH. NEWS 16 (2002), http://psyking.net/HTMLobj-3843/using_case_study_in_research.pdf [https://perma.cc/EDF9-ETJ5]. *See e.g.*, Wells v. SmithKline Beecham Corp., 601 F.3d 375, 380 (5th Cir. 2010).

65 *E.g.*, Norris v. Baxter Healthcare Corp., 397 F.3d 878, 885 (10th Cir. 2005); *Wells*, 601 F.3d at 380; Kilpatrick v. Breg, Inc., 613 F.3d 1329, 1339 (11th Cir. 2010*);* McClain v. Metabolife Int'l, Inc., 401 F.3d 1233, 1253–54 (11th Cir. 2005); Kuhn v. Wyeth, Inc., 686 F.3d 618, 633 (8th Cir. 2012); Christopher R.J. Pace, *Admitting and Excluding General Causation Expert Testimony: The Eleventh Circuit Construct*, 37 AM. J. TRIAL ADVOC. 47, 56–57 (2013). State law might be more forgiving: as one court noted, some experts think case studies are good enough, and that was good enough for the court. *Davis*, 245 Cal.App.4th. Compare this view in a superseded opinion, which seems to suggest that case studies alone will not suffice. *Lockheed Litig. Cases*, 23 Cal. Rptr. 3d 762, 780 (2005), *review granted and opinion superseded sub nom. In re* Lockheed Litig. Cases, 110 P.3d 289 (2005) ("Causal attribution based on case studies must be regarded with caution. However, such studies may be carefully considered in light of other information available") (internal quotes omitted).

opinion.

4. *Suspect Implausible Extrapolation*

In a way, much expert testimony is extrapolation from what is scientifically established to the facts of the case. We extrapolate from a valid sample to a population. Even a superb study will usually only show correlation of an intervention (i.e., a drug) and an outcome; causation is an extrapolation (sometimes unwarranted[66]).

But some extrapolations are weak, such as those from sources identified above, e.g., case studies. Small scale animal studies alone may also be speculative, because they involve both extrapolations across species and of doses,[67] although there is case law to the contrary.[68] Below, I'll discuss the weakness of relying on a single study—even a sophisticated, well-done study with statistical significance. Under Field theory, extrapolations which seem incredible probably are: we should consider what other theories and beliefs we'd have to surrender or modify to accommodate the extrapolated conclusion. Scientists extrapolate, guess, and imagine all the time; that's essential work in science, but we don't admit it at trial.[69]

5. *The Role of Theory*

The scientific method requires a theory, by which I mean nothing more than an abstraction, or explanatory mechanism, of the observation or specific hypotheses. The theory is the means by which we figure out the mechanical plausibility of the observation or hypothesis; it's what lets us apply Field theory; and under falsifiability theory, it lets us devise

---

66 *Why Correlation Does Not Imply Causation in Statistics*, https://www.mathtutordvd.com/public/Why-Correlation-does-not-Imply-Causation-in-Statistics.cfm [https://perma.cc/JB6B-UQZ5].

67 Tim Sandle, *What is 'Bad Science' and How to Spot It*, DIGITAL J. (Jan. 22, 2016), http://www.digitaljournal.com/science/why-some-science-is-actually-bad-science/article/455538 [https://perma.cc/N6EZ-PZY4]. *See generally* Amanda Hungerford, *Back to Basics: Courts' Treatment of Agency Animal Studies After Daubert*, 110 COLUM. L. REV. 70 (2010).

68 *E.g.*, In re Paoli R.R. Yard PCB Litig., 35 F.3d 717, 781 (3d Cir. 1994) in which opinions based on animal studies, with other bases, were admissible*;* Milward v. Acuity Specialty Prod. Grp., Inc., 639 F.3d 11, 20 (1st Cir. 2011) (same); Roberti v. Andy's Termite & Pest Control, Inc., 113 Cal.App.4th 893, 904 (2003); Metabolife Int'l, Inc. v. Wornick, 264 F.3d 832, 842 (9th Cir. 2001) ("animal studies are not per se inadmissible"). *Compare e.g.*, Johnson v. Arkema, Inc., 685 F.3d 452, 463 (5th Cir. 2012) (noting significant distinctions in the circumstances of that case between animal and human organs).

69 Allison v. McGhan Medical Corp., 184 F.3d 1300, 1316–17 (11th Cir. 1999).

experiments.

For example, gravity is a theory which explains, and is an abstraction of, the observations we make of apples falling to the ground, the planets about the sun, and so on. The theory implies other observations or facts, and is in effect a source of predictions we can test. Feynman has a wonderful story of a mind reader: the 'theory' is that there's a guy who says he has the power of telekinesis—he can move things with his mind. If that's true, then he can make a roulette wheel come up black or red, or make other things happen with his mind, and we can test this.[70] This is the making of predications which Popper speaks of. It may not be true that a theory can be falsified with a negative experiment, for reasons outlined above, but without a theory which provides a supposed mechanism or which in effect makes predications, we don't have science. If a theory is too vague—if it isn't sufficiently specific to generate predictions which can be tested—then it's not scientific,[71] and we can't employ the scientific method.

If someone tells me an eclipse is a dragon eating the moon, I want to know what that implies, via understanding the undergirding theory: Do dragons interact with all celestial objects—why only some of them? How does all this work? How under this theory is it that we get the moon back? How does the theory tell us when the next occurrence will be? What other eating goes on in space? There are no answers to these questions: it's just a dragon eating the moon. There's no theory of which the observation (the moon being eaten) is a part; and that's the same thing as saying there's nothing else implied by the observation. So there's no science here.

Under this approach, phrenology *is* a sort of science, but it's just completely wrong. We can in fact make predictions about the relationship of the skull's shape to mental attributes, and then test these. We can have a theory based on the skull pressing on or changing the shape of parts of the brain to make changes to mental attributes.[72] Then we can experiment and

---

70 RICHARD P. FEYNMAN, THE MEANING OF IT ALL 70 (1998).

71 RICHARD P. FEYNMAN, *Seeking New Laws*, in THE CHARACTER OF PHYSICAL LAW 158 (1994).

72 The 'discipline' had all the earmarks of a scientific endeavor. "In 1840, phrenology was a confident science, promising clear and certain knowledge concerning the mental attributes and behaviors of human beings. It was a time of exhilarating new possibilities, of discoveries compounding discoveries. There were conferences and symposia. There were professional associations. There were lengthy learned tomes and scholarly journals." Pierre Schlag, *Commentary: Law And Phrenology*, 110 HARV. L. REV. 877 (1997). *See generally* O. Parker Jones, F. Alfaro-Almagro, & S. Jbabdi, *An Empirical, 21st Century Evaluation of Phrenology*,

do surveys to see if there is something right about phrenology. There isn't. It's just wrong.[73] But at least we had a theory, a scientific hypothesis, and we used the scientific method.

C. Criteria

Here, I summarize criteria a judge might use to decide if studies and reports are a sufficient basis for an opinion, and if so, admit the opinion. I offer these criteria as emblematic of valid scientific method.

Data. (A) Because we want our studies to be replicable—more on that just below—the data used must be public or at least accessible. Secret data won't do, because there's no way to check the work. This can be a serious problem, because it may be expensive to gather data, and researchers might not want to give it away for free. (B) With so-called 'big data,' there is infinite opportunity to cherry-pick data to show almost any sort of correlation one desires. Given random chance, if one wishes to show, for example, an association between eating fish and leukemia, or between height and drinking fruit juice (or, more ominously, between high tension wires and cancer), one will surely find, somewhere on earth, an increased incidence of correlation—just as one will find a decreased incidence somewhere else on the planet. So, no cherry-picking data for a given study. Cherry-picking data, selective reporting, p-hacking and other mechanisms can be used to show minimal statistical significance, but those studies are of course unreliable.[74]

Studies. Just as we shouldn't allow in a study which cherry-picked its data, we should not allow in studies cherry-picked from other studies. With the vast number of journals and studies available,[75] it is likely that *any* conclusion

---

SCIENCEDIRECT (Jan. 5, 2018), https://www.biorxiv.org/content/biorxiv/early/2018/01/05/243089.full.pdf [https://perma.cc/5G4V-SF4N].

73 Jones, Alfaro-Almagro, & Jbabdi, *supra* note 72.

74 EXPERT WITNESSES, *supra* note 4, at § 2.10.4.4. DENNIS GORMAN, *Evidence-based Practices as a Driver of Pseudoscience in Prevention Research*, in PSEUDOSCIENCE: THE CONSPIRACY AGAINST SCIENCE 263 (2019) (much of what passes for 'evidence-based,' approved and effective drug and alcohol prevention therapy has no scientific basis).

75 There are perhaps "around 30,000 [journals], with close to two million articles published each year." Philip G. Altbach & Hans de Wit, *Too Much Academic Research is Being Published*, UNIV. WORLD NEWS (Sept. 7 2018), https://www.universityworldnews.com/post.php?story=20180905095203579 [https://perma.cc/2Z45-MMV8]; *see also* Peder Olesen Larsen & Markus von Ins, *The Rate of Growth in Scientific Publication and the Decline in Coverage Provided by Science Citation Index*, SCIENTOMETRICS, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2909426

can find some support from some study.[76] If there is no reasonable explanation for choosing to rely on one or a few studies in the face of other conflicting studies, the testimony may not be admissible.

A single study can accurately report acceptable statistical significance, but, one in twenty times, the results may just be the result of random chance.[77] We don't know if a single study is the random one. '[O]ne-off studies . . . disappear in science's rearview mirror, surviving only from self-citation and eternal content of activist websites."[78]

We are looking for *replicated* results; it is shocking how often studies which seem reputable, published in reputable journals, cannot be replicated.[79] To be replicable, a study must show its work: walking the reader from the data collection to the conclusion. It's also important to know if someone has tried, and failed, to replicate the study. The problem is that with "publication bias"—which favors publishing break-out results to the exclusion of studies which merely tread old ground, and publish positive findings to the exclusion of negative findings (i.e. which find no correlation)—few scientists try for replication, and those that do, may not get published. A single study is probably insufficient to support an expert opinion.[80]

[https://perma.cc/J4H7-PJXT]; Mark Ware, *The STM Report*, STM 1, 6, https://www.stm-assoc.org/2015_02_20_STM_Report_2015.pdf ("There were about 28,100 active scholarly peer-reviewed English-language journals in late 2014 (plus a further 6450 non-English-language journals), collectively publishing about 2.5 million articles a year").

76 YOU'LL FIND IT, *supra* note 58, at 14, 26

77 EXPERT WITNESSES, *supra* note 4, at §§ 2.10.4.4.3, 2.10.4.4.7.

78 KEVIN M. FOLTA, *Food-o-Science Pseudoscience: The Weapons and Tactics in the War on Crop Biotechnology*, in PSEUDOSCIENCE: THE CONSPIRACY AGAINST SCIENCE 114 (2019), *see id.* at 113-15. *See also e.g.*, STEPHEN JAY GOULD, FULL HOUSE: THE SPREAD OF EXCELLENCE FROM PLATO TO DARWIN 208 (1996) ("One study doesn't prove a generality any more than a single swallow makes a summer").

79 "Another reason for 'bad science' is unintended and intended 'errors.' Back in 2012, the journal *Proceedings of the National Academy of Sciences* tallied up 2,047 recent retractions from journals. "Retraction" means a journal paper has been pulled because of an error. The review article found about 20 percent of the retractions were due to unintended errors (like a badly performed calculation, of the sort that should be picked up using statistical techniques like Benford's law); however, and more seriously, about 67 percent were retracted due to misconduct. Of these, 43 percent were due to deliberate fraud." Sandle, *supra* note 67. *See also*, *How to Spot Bad Science*, FARNAM STREET, https://fs.blog/2020/01/spot-bad-science [https://perma.cc/4MUE-FKA4].

80 "In the vast majority of cases, a single study is a starting point, not proof of anything. The results could be random chance, or the result of bias, or even outright fraud. Only once other researchers replicate the results can we consider a study persuasive. The more replications, the more reliable the results are. If

But case law does not reflect that: I have found no court authority which deems an expert opinion based on one study inadmissible for that reason.[81] Indeed we have this, taken out of context, which at least suggests the opposite:

> A jury may repose greater confidence in an expert who relies upon well-established scientific principles. It may accord less weight to the views of an expert who relies on a single article from an obscure journal or on a lone experiment whose results cannot be replicated.[82]

As I discuss below in the fourth part, we should rethink the basis on which a judge can reject an opinion based on only one or a few studies when either (i) other studies, including meta-studies, show the few studies to be unreliable or (ii) there are no other studies.

A few caveats. First, what appears to be one study may actually include more: multiple selection of subjects and interventions which honestly do reduce the risk of spurious results. Secondly, we may have what seems to be a long history of studies and consequent expert opinions, justifying admissibility, but then a few more recent studies which profoundly undermine the accepted wisdom. For example, pattern evidence such as blood splatter, tire tracks, ballistics comparisons, and so on have been used in court for generations, but recent studies show the techniques are unreliable.[83] Nevertheless, courts have continued to allow this sort of evidence, because of the weight of past opinions and the apparent fact that the "'clear majority' of the relevant scientific community" has, as far as one case was concerned, not yet agreed to banish the evidence.[84] The point is that sometimes

---

attempts at replication fail, this can be a sign the original research was biased or incorrect." *How to Spot Bad Science*, *supra* note 79. *See also, e.g.,* M. Munafò et al., *A Manifesto for Reproducible Science*, 1 NAT. HUM. BEHAV. 21 (2017), https://doi.org/10.1038/s41562-016-0021 [https://perma.cc/6LNV-QG32]. *See also e.g.,* KALICHMAN, *supra* note 60, at 421 ("the single study fallacy").

81 *Compare* People ex rel. Brown v. Tri-Union Seafoods, LLC, 171 Cal.App.4th 1549, 1568 (2009) (apparently relying on one study); Johnson & Johnson Talcum Powder Cases, 37 Cal.App.5th 292, 326 (2019), which approves an opinion based on two main studies (as well as more peripheral studies).

82 People v. Sanchez, 63 Cal. 4th 665, 686 (2016).

83 EXPERT WITNESSES, *supra* note 4, at § 2.11.

84 *Azcona*, 58 Cal.App.5th at 512 (2020). This is a peculiar case, perhaps because it has alternative holdings. The problem with the pattern evidence was raised by the defendant as a *Kelly* issue, and the court found (with the concurring opinion disagreeing on this point) that the "defendant did not meet his burden to show that a clear majority of the relevant scientific community no longer accepts the method as reliable." This the pattern evidence was admissible. Because it was raised as a *Kelly* issue, the court refused to review whether the evidence was

one, or very few, new studies might be reliable evidence, more reliable than the old warhorses, and sometimes showing the old studies to be unreliable.

Bad Statistics. The subject is very broad; I and many others have written about it in more detail elsewhere.[85] But I summarize a few principles.

A. The study should distinguish the p-value (statistical significance, roughly the likelihood that the result was caused by random chance) from the effect size (measuring the amount of impact, the importance, if you will, of the intervention).[86] For example, a study may show with an adequate p-value that a drug reduces pain, but the effect size might be very small, i.e. do very little to reduce pain. I don't suggest that expressing an effect size is essential (although it's a good impeachment point).

B. Samples. The study must use randomly selected samples from the population and randomly assign them to the control and test groups. Many techniques, which at first glance seem random, are not.[87] The sample must be of a

---

"reliable," as would have happened in federal court, that is, it refused to impose a generic reliability test for this supposedly scientific evidence. *Id.* at 510-11. Then the court found that the issue probably was *not* a *Kelly* matter at all, because the jury could on its own (presumably using its "common sense") figure out the reliability of the pattern evidence: "It is not clear that the technique employed here is subject to the *Kelly* standard at all, as visual comparison of marks on physical objects is not so foreign to everyday experience that jurors would have unusual difficulty evaluating it." *Id.* at 511. Having opined that the issue likely was not controlled by *Kelly*, and *was* subject to the usual reliability standards, the court noted that there was considerable evidence that undermined the reliability of the methods, but left the issue to the jury—despite the fact, unremarked by the court, that the trial judge (following the lead of counsel) had not reviewed reliability but, probably erroneously, applied the *Kelly* test. *Id.* at 512. In a final irony, the court seems to endorse the admission of the evidence under *Kelly* (which it had said was probably inapplicable), which requires a previous published California opinion endorsing the scientific technique, even though the majority and the concurrence agreed they could not find any such published opinion, because the parties, apparently mistakenly, thought there was such a published opinion. *Id.* at n.1 and 6.

85 Curtis E.A. Karnow, *Compression Algorithm: Big Data in Small Courtrooms*, 25 ABTL REPORT 1 (Winter 2016), https://works.bepress.com/curtis_karnow/43/ [https://perma.cc/2ADX-M5LY]; CURTIS E.A. KARNOW, *Statistics In Law: Bad Inferences & Uncommon Sense*, in LITIGATION IN PRACTICE (2017). I provide a list of resources at "Experts, Statistics, Science & Bad Science" in EXPERT WITNESSES, *supra* note 4.

86 *E.g.,* MEERA, "Power Analysis, Statistical Significance, & Effect Size," https://meera.snre.umich.edu/power-analysis-statistical-significance-effect-size [https://perma.cc/T4WE-XDL2].

87 E.g., Matt Downs, Kathryn Tucker, Heidi Christ-Schmidt, & Janet Wittes, *Some Practical Problems in Implementing Randomization*, 7 CLINICAL TRIALS 235—245 (2010); Lise Lotte Gluud, *Bias in Clinical Intervention Research*, 163 AM. J. EPIDEMIOLOGY 493–501 (2006) (from the abstract: "A number of methodological

minimum *size*: there are standard ways to figure that out, and sometimes a pilot study is needed.[88] The study must state the size of the population from which the sample is drawn (else the sample size may be difficult to calculate). Any study with p-values must—no exceptions here—state the confidence interval (margin of error such as ± 5) and the confidence level (e.g. 99% or 90% confident).[89]

A control group is usually needed to evaluate the impact of an intervention, or else the study won't be able to determine the impact of the placebo and nocebo effects.[90] Most interventions have some effect just as a matter of suggestion: subjects interacting with someone who looks like a doctor, or takes an inert sugar pill, may have beneficial effects (placebo[91]), and those told that an intervention may have adverse effects (e.g., medicine might make them nauseous) will report exactly that effect even when given an inert pill (nocebo[92]). Without a control group, the impact of these effects

---

studies suggest that lack of adequate randomization in published trial reports may be associated with more positive estimates of intervention effects").

88 Duran v. U.S. Bank National Assn., 59 Cal.4th 1, 22 (2014) (witness reports that "pilot study is typically done to determine the amount of variation in the underlying population"); Modaraei v. Action Property Mgmt., Inc., 40 Cal.App.5th 632, 644 (2019). *See e.g.*, Bruce P. Keller, *A Survey of Survey Evidence*, 19 LITIGATION 23, 27 (1992) (pilot study "can show the likely results of a full-scale survey. They also may help reveal flaws (*e.g.,* improper universe, suggestive questions, unforeseen responses to open-ended questions) that can be corrected in the main survey"); Vineet Chopra, et al., *Variation in Use and Outcomes Related to Midline Catheters: Results from a Multicentre Pilot Study*, 28 BMJ QUAL. SAF. 714-720 (2019) ("Given this gap, we conducted a pilot study to examine use, variation and outcomes related to midline catheters in hospitals across the state . . .").

89 These numbers are interrelated. A larger sample size may allow a larger confidence level and/or smaller margin of error; and vice versa.

90 YOU'LL FIND IT, *supra* note 58, at 401.

91 For example, it is likely that the perceived benefits of acupuncture are a placebo effect. *E.g.,* E. Ernst, Acupuncture—A Critical Analysis, 259 J. OF INTERNAL MED. 125-137 (2005). *See also* Daniel C. Cherkin, et al., *A Randomized Trial Comparing Acupuncture, Simulated Acupuncture, and Usual Care for Chronic Low Back Pain*, 169 ARCH INTERN MED (2009) ("It remains unclear whether acupuncture or our simulated method of acupuncture provide physiologically important stimulation or represent placebo or nonspecific effects." "[T]rials evaluating acupuncture for pain have failed to find that real acupuncture is superior to sham or superficial control treatments and raises questions about whether sham treatments truly serve as inactive controls").

92 *E.g.,* Paul Enck, Fabrizio Benedetti, & Manfred Schedlowski, *New Insights into the Placebo and Nocebo Responses*, 59 NEURON 195-206 (2008) http://www.sciencedirect.com/science/article/pii/S0896627308005850 [https://perma.cc/NQC7-LNL3]; Eric Manheimer et al., *Acupuncture for Hip Osteoarthritis*, COCHRANE REVIEWS (2018), https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD013010/full [https://perma.cc/3KT7-LY8Q]. *Compare* Andrew J. Vickers et al., *Acupuncture*

is unknown.

C. Peer Review. While usually considered an essential foundation to an expert's opinion, the existence of peer review studies is only a first step. As suggested above, it is important to understand follow-up reports—those which confirm or undermine a peer reviewed study. Simply being in a peer reviewed journal is no assurance of the report's validity, for many reasons. First, some essays in these journals are good science in that they present intriguing results, not from rigorous studies, but by way of preliminary reviews, case studies, and so on. These are the unreliable extrapolations I noted above. Second, there are peer reviewed studies for everyone, including astrologers. An expert dowser relying on studies of dowsing reviewed by other dowsers should give no one comfort. Third, many wonderfully sounding journals are frauds. They prey on academics looking—and willing to pay—for publication. They may say they are peer reviewed, but it's not true. "What makes these frauds so devious is that it's extraordinarily difficult to tell whether a journal is real just by looking at it online."[93] Fourth, and most troubling, is the fact that many apparently reasonable, well-done studies,

---

for Chronic Pain: Update of an Individual Patient Data Meta-Analysis, J. PAIN (2017) https://doi.org/10.1016/j.jpain.2017.11.005 [https://perma.cc/H6EV-TG39] (pain control better than placebo).

93 Bradley Allf, *Opinion: I Published a Fake Paper in a 'Peer Reviewed' Journal*, UNDARK (Nov. 26, 2020), https://undark.org/2020/11/26/fake-paper-predatory-journal/ [https://perma.cc/4QTN-E6KV]. For a nifty story of a fake paper, which easily managed to get published without meaningful peer review, *see e.g.,* Samantha Cukier et al., *Defining Predatory Journals and Responding to the Threat They Pose: A Modified Delphi Consensus Process*, BMJ OPEN (2020), https://www.nature.com/articles/d41586-020-00911-x [https://perma.cc/8TEW-58SM]; *See also*, Richard Knox, *Some Online Journals Will Publish Fake Science For a Fee*, NPR (Oct. 3, 2013, 6:37 PM), https://www.npr.org/sections/health-shots/2013/10/03/228859954/some-online-journals-will-publish-fake-science-for-a-fee [https://perma.cc/ZLC8-F96Y]; *Choosing a Journal for Publication of an Article: List of Suspicious Journals and Publishers*, YALE UNIV. LIBR., https://guides.library.yale.edu/c.php?g=296124&p=1973764 [https://perma.cc/3379-24ZH]; Alex Hern & Pamela Duncan, *Predatory Publishers: The Journals That Churn Out Fake Science*, GUARDIAN (Aug. 10, 2018, 12:55 PM), https://www.theguardian.com/technology/2018/aug/10/predatory-publishers-the-journals-who-churn-out-fake-science [https://perma.cc/53UX-BZD3]; JEFFREY BEALL, *Scientific Soundness and the Problem of Predatory Journals*, in PSEUDOSCIENCE: THE CONSPIRACY AGAINST SCIENCE 283 (2019). The Beall article notes many of these fake peer review journals have much of the indicia of legitimacy: editorial boards, PDFs on-line, volumes and issues, and peers who review (but whose comments are ignored). *Id.* at 291-92. Indeed, some of these journals are listed on PubMed, published by the US National Medical Center of Biotechnology Information, a sign of "instant legitimacy, a seal of approval from the US government." *Id.* at 293.

published in plainly reputable journals, and actually peer reviewed, turn out to be irreproducible;[94] and many of them are—years later—withdrawn.[95] The point is not precisely that some studies might be withdrawn; the point is that studies which have not been subject to replication attempts may, for that reason, not be reliable. The implications of this are explored in the fourth part below.

D. Impeachment

Assuming the court has admitted the expert opinion based on the reliability of the underlying studies, I propose appropriate attacks to be conducted before the jury.

Financial Interest. Not all researchers are independent. Readers may recall studies of tobacco conducted and paid for by the tobacco industry.[96] Drug companies fund studies of the drugs they sell.[97] There may not be practical alternatives to this practice, because, for example, clinical trial can be

---

94 YOU'LL FIND IT, *supra* note 58, at 10.

95 Jeffrey Brainard & Jia You, *What a Massive Database of Retracted Papers Reveals About Science Publishing's 'Death Penalty,'* SCIENCE (2018), https://www.sciencemag.org/news/2018/10/what-massive-database-retracted-papers-reveals-about-science-publishing-s-death-penalty [https://perma.cc/U56S-DP4A]. *See e.g.,* Monya Baker, *1,500 Scientists Lift the Lid on Reproducibility*, NATURE (2016) https://www.nature.com/articles/533452a [https://perma.cc/2UTB-L9LK] ("More than 70% of researchers have tried and failed to reproduce another scientist's experiments, and more than half have failed to reproduce their own experiments. Those are some of the telling figures that emerged from *Nature*'s survey of 1,576 researchers who took a brief online questionnaire on reproducibility in research"); Shannon Palus, *Science Under Scrutiny: The Problem of Reproducibility*, SCI. AM. (2018), https://www.scientificamerican.com/article/science-under-scrutiny-the-problem-of-reproducibility/ [https://perma.cc/Q5DV-SF9D]. Careful examination of papers, including using some of the statistics approaches urged in this Note, may be useful in predicting which studies are unlikely to be replicated. *See e.g.,* Kelsey Piper, *Science Has Been in a "Replication Crisis" for a Decade. Have We Learned Anything?*, VOX (2020), https://www.vox.com/future-perfect/21504366/science-replication-crisis-peer-review-statistics [https://perma.cc/9EZS-FMXF]. *See generally*, Kevin D. Hill, *The Crisis in Scientific Publishing and Its Effect on the Admissibility of Technical and Scientific Evidence*, 49 J. Marshall L. Rev. 727, 737 (2016).

96 Lisa A. Bero, *Tobacco Industry Manipulation of Research*, PUB. HEALTH REP. (2005), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1497700/pdf/15842123.pdf [https://perma.cc/8NSM-PEPR].

97 Sameer S. Chopra, *Industry Funding of Clinical Trials: Benefit or Bias?*, 290 JAMA 113-114 (2003), https://jamanetwork.com/journals/jama/fullarticle/196846 [https://perma.cc/9PMS-WWLG]; Shaoni Bhattacharya, *Research Funded by Drug Companies is 'Biased'* NEWSCIENTIST (30 May 2003), https://www.newscientist.com/article/dn3781-research-funded-by-drug-companies-is-biased/ [https://perma.cc/K3MC-8ZEV].

extremely expensive and no one else will fund them. But the jury deserves to know, and not all conflicts of interests are disclosed. Some experts have an even more direct financial interest, because they are selling the gear or material which, they testify, is efficacious.[98]

Other bias. The are other sorts of biases, such as publication bias. Sometimes researchers will publish results multiple times to make it look as if there are multiple studies.[99] Studies sponsored or publicized by governmental entities may be slanted or cherry-picked in order to support governmental policies. *Cui bono*?

Meta-Analyses. Any given study—including a very well-done study—could be an outlier, so we should know what the literature generally shows. Best of all are studies of studies, i.e., meta-studies or meta-analyses.[100] There should be a deep suspicion of results which contradict the results of a meta-analysis.[101]

Double blind administration. Studies in which subjects are administered a dose, such as clinical trials with a control group, should be structured so that those administering the dose don't know if they have the placebo. Subtle, unconscious behavior by the those administering the test can disrupt the

---

98 Goldacre addresses the issue in many of his books and his website, *e.g., Bad Science Book Summary, by Ben Goldacre*, ALLEN CHENG, https://www.allencheng.com/bad-science-book-summary-ben-goldacre/ [https://perma.cc/RB3Z-AQL7]. A study showing some benefit to acupuncture for migraines was apparently conducted by researchers who make a living teaching acupuncture. Ling Zhao et al., *The Long-term Effect of Acupuncture for Migraine Prophylaxis, A Randomized Clinical Trial*, JAMA INTERNAL MED., Feb. 20, 2017, at E1, https://www.researchgate.net/profile/Ling_Zhao29/publication/314087920_The_Long-term_Effect_of_Acupuncture_for_Migraine_Prophylaxis_A_Randomized_Clinical_Trial/links/5ac65abca6fdcc8bfc7f71de/The-Long-term-Effect-of-Acupuncture-for-Migraine-Prophylaxis-A-Randomized-Clinical-Trial.pdf [https://perma.cc/ZPT8-CR3V].

99 YOU'LL FIND IT, *supra* note 58, at 28.

100 *E.g.*, Jessica Gurevitch, Julia Koricheva, Shinichi Nakagawa, & Gavin Stewart, *Meta-analysis and the Science of Research Synthesis*, 555 NATURE 175, 175-182 (2018) http://unsworks.unsw.edu.au/fapi/datastream/unsworks:71830/bin4a429cdc-3669-4111-8e38-3235a5cdb724?view=true&xy=01 [https://perma.cc/972P-Y2F6]; YOU'LL FIND IT, *supra* note 58, at 25. For an example, see e.g., Feng J. He, Jiafu Li, & Graham A MacGregor, *Effect of Long Term Modest Salt Reduction on Blood Presure: Cochrane Systematic Review and Meta-Analysis of Randomised Trials*, BMJ, Apr. 4, 2013, at 1, https://www.bmj.com/content/bmj/346/bmj.f1325.full.pdf [https://perma.cc/262U-ZNDN].

101 YOU'LL FIND IT, *supra* note 58, at 349.

results.[102]

IV
PROBLEMS IN THE USE OF STUDIES

Might a study be used to impeach an admitted opinion? "An expert witness may be cross-examined about 'the matter upon which his or her opinion is based and the reasons for his or her opinion.'[103] The scope of this inquiry is broad and includes questions about whether the expert sufficiently considered matters inconsistent with the opinion."[104] At first blush that seems to allow cross examination with conflicting studies. But under Evid. C. § 721 (b), the expert "may not be cross-examined in regard to the content or tenor of any scientific, technical, or professional text, treatise, journal, or similar publication unless: [¶] (1) The witness referred to, considered, or relied upon such publication in arriving at or forming his opinion; or [¶] (2) Such publication has been admitted in evidence."[105]

So, the only way to admit a report which the expert ignored is to have it admitted while an *opposing* expert is testifying, on direct.[106] But that isn't usually allowed under current law, generally because of hearsay problems.[107] That is, the jury, which is charged under current law to decide among scientific theories and experts, is not able to review the studies and reports which putatively show a proffered approach or study, offered on direct examination, is "scientific."

This is peculiar. The peculiarity is ameliorated to some extent by the usual practice of allowing a testifying expert to recite summaries of the studies she relies on; this seems quite necessary, although it is not clear why the recitation of

---

102 *E.g.,* Double Blind, SCI. DAILY, https://www.sciencedaily.com/terms/double_blind.htm

103 Evid. C. § 721 (a).

104 People v. Doolin, 45 Cal.4th 390, 434 (2009).

105 McGarity v. Dep't of Transportation, 8 Cal.App.4th 677, 681 (1992).

106 Evid C. § 721(b)(2).

107 *E.g.,* People v. Dean, 174 Cal.App.4th 186, 201 (2009); Ellis v. International Playtex, Inc., 745 F.2d 292, 302 (4th Cir. 1984) (but finding an exception to hearsay rule in that case). It is not clear if statements in these reports, which do not relate facts specific to the case, might be admissible under People v. Sanchez, 63 Cal.4th 665 (2016). *Cf.,* People v. Azcona, 58 Cal.App.5th 504 (2020). *E.g.,* People v. Campos, 32 Cal.App.4th 304, 308 (1995) ("An expert witness may not, on direct examination, reveal the content of reports prepared or opinions expressed by non-testifying experts"). *See e.g.,* Ronald L. Carlson, *The Curious Case of Differing Literary Emphases: The Contrast Between the Use of Scientific Publications at Pretrial Daubert Hearings and at Trial*, 47 GA. L. REV. 837, 854–56 (2013).

summaries is not as much hearsay as the studies themselves.

This is an example of the incoherence cited in the opening section of this Essay: the jury must choose among competing scientific theories while being blocked from reviewing the key evidence needed to make the decision.[108]

The problem is as intractable when it comes to *impeachment* through the use of reports. Under current law, it is likely impossible to impeach based on what appear to be decisive studies if not relied on by the expert to be impeached. (Of course the expert didn't rely on them—that's the *point*.) And the other side's expert, even if *he* relied on them, likely won't provide a sufficient basis for admissibility for reasons noted just above. So there are no admissible reports, which means we can't use reports to impeach, by reason of Evid. C. § 721 (b).

This is also incoherent.

V
SOLUTIONS ON THE USE OF STUDIES

If this Essay is correct that (i) the reliability of an expert opinion should be evaluated against the backdrop of all pertinent studies and reports, including any meta-studies, and thus (ii) reliance on a single study is likely unreliable, and (iii) only a close examination of the studies will reveal if they faithfully track scientific methodology, then there are two solutions: Either the law changes to allow the jury to review in detail all the studies and have the lawyers present their arguments on all reliability issues to the jury, or the judge must undertake the task.

A. The Jury

The first option is unlikely: it would require changes in evidence law, the jury would be expected to analyze perhaps scores of studies, over days, and come to a consensus whether they satisfy the criteria of the scientific method. Some combination of training and jury instructions would be required to advise on what the scientific method entails. This is a significant change in the role of the jury.

1. Kelly *and Common Sense*

To understand how deep a shift this would represent, let us recall the rationale for California's *Kelly*[109] rule, which

---

108 See generally, EXPERT WITNESSES, *supra* note 4, at § 2.10.3.

109 People v. Kelly, 17 Cal.3d 24 (1976).

governs the introduction of new scientific techniques such as polygraphs, DNA, the horizontal gaze nystagmus (HGN) field sobriety test,[110] and other evidence.[111] (As I noted in the introduction, this Essay does not focus on *Kelly* evidence, but rather the broader issue of how to handle scientific experts in the general *Sargon* context; however, a contrast here will illuminate the issue of how to handle scientific reliability.)

Kelly requires the new technique to be shown—to the judge—to be commonly accepted by the scientific community[112] as a sound scientific method. The key aspect of *Kelly* evidence is that it emanates from what I will call a "black box," i.e. a source which might seem to the jury to be infallible, a technique which has the "aura of certainty"[113] which seems to "provide some definitive truth which the expert need only accurately recognize and relay to the jury."[114] The *Kelly* problem is that the jury doesn't see the expert witness as the source of the evidence and cannot use its oft-proclaimed "common sense"[115] to evaluate the testimony:

> The purpose underlying the rule is to protect a jury from expert testimony that conveys a "'misleading aura of certainty' about a scientific technique. The *Kelly* analysis thus is designed to address "scientific evidence or technology that is so foreign to everyday experience as to be unusually difficult for laypersons to evaluate."[116]

---

110 People v. Leahy, 8 Cal.4th 587, 591 (1994).

111 People v. Shirley, 31 Cal.3d 18, 51–52 (1982) (*Kelly* applied to polygraph examinations, "truth serum," experimental systems of blood typing, voiceprint analysis, identification of human bite marks, and microscopic identification of gunshot residue particles), as noted in People v. Garlinger, 247 Cal.App.4th 1185, 1194–1195 (2016). *See also e.g.,* People v. Hardy, 65 Cal.App.5th 312, 328–29 (2021) ("Shotspotter" technology which identifies location, time and number of gunshots fired requires *Kelly* hearing).

112 People v. Leahy, 8 Cal.4th 587, 601 (1994) ("we should make clear that "general acceptance" does not require unanimity, a consensus of opinion, or even majority support by the scientific community").

113 People v. Leahy, 8 Cal.4th 587, 607 (1994).

114 People v. Leahy, 8 Cal.4th 587, 606 (1994). *See e.g.,* People v. Garlinger, 247 Cal.App.4th 1185, 1195 (2016) ("misleading aura of scientific infallibility," quoting People v. Stoll, 49 Cal.3d 1136, 1157 (1989)).

115 People v. Daveggio and Michaud, 4 Cal.5th 790, 841 n.12 (2018) (common sense evaluating experts); People v. Richardson, 43 Cal.4th 959, 1017–1018 (2008) (same); People v. Lavender, 60 Cal.4th 679, 681 (2014) (common sense in non-expert context); Lara v. Nevitt, 123 Cal.App.4th 454, 460 (2004) (common sense in lieu of reliance on expert). The instruction to use "common sense" is of dubious utility. "To tell a juror to use common sense and experience is little more than telling the juror to do what the juror cannot help but do." People v. Campos, 156 Cal. App.4th 1228, 1240 (2007).

116 People v. Hung Tran, 50 Cal.App.5th 171, 186 (2020) (quote marks removed).

Instead of live witnesses, the source to be tested on cross examination is the black box; and there are no questions for a black box. Old fashioned cross-examination doesn't work for e.g., a "scent transfer unit" used to make scent-based identifications of a suspects: it just generates its output.[117] The *Kelly* requirements for a demonstration—to the judge and not the jury—of common acceptance kick in, because the usual assumption of human fallibility, which is available on all other examinations of witnesses, is not available here:

> When a witness gives his personal opinion on the stand—even if he qualifies as an expert—the jurors may temper their acceptance of his testimony with a healthy skepticism born of their knowledge that all human beings are fallible. But the opposite may be true when the evidence is produced by a machine: like many laypersons, jurors tend to ascribe an inordinately high degree of certainty to proof derived from an apparently 'scientific' mechanism, instrument, or procedure.[118]

Post-*Sargon* law—*Sargon*[119] itself, indeed—plainly distinguishes the *Kelly* black box problem from other scientific experts, relying on the "common sense" of juries in the latter but not former situation. But I suggest the matter is in fact not much different in the non-*Kelly* context, the subject of this Essay, were a jury to be presented with dozens of studies and reports together with meta-analyses, with conflicting conclusions, each tracking to some extent, but perhaps not in every respect, the criteria of scientific methods outlined above (or indeed any criteria set out by case law).

Who are the witnesses to be examined here? Certainly, the experts who expect to testify at trial will be available, but recall that generally it is not their work which is the subject of the studies; those authors don't testify. So once again, we are met with the futility of old-fashioned cross examination: The jury is just looking at the studies, and these may well appear for all the world to be "black boxes." Tellingly, *Sargon* itself notes something similar in this non-*Kelly* context, as it refers to an expert's report as "an array of figures conveying a delusive impression of exactness in an area where a jury's common sense is less available than usual to protect it."[120]

---

117 People v. Mitchell, 110 Cal.App.4th 772, 787 (2003).

118 People v. McDonald, 37 Cal.3d 351, 372 (1984).

119 Sargon Enterprises, Inc. v. Univ. of S. California, 55 Cal.4th 747 (2012).

120 *Sargon Enterprises*, 55 Cal.4th 747, 769 (2012) (quoting Judge Friendly in Herman Schwabe, Inc. v. United Shoe Mach. Corp., 297 F.2d 906, 912 (2d Cir. 1962)).

Relying on the jury's "common sense" for non-*Kelly* scientific experts as we do for other witnesses is dubious: much of widely accepted science is *not* a matter of common sense, and common sense—at least lay common sense—won't help anyone figure it out.[121] The big bang theory of the beginning of the universe, quantum mechanics, relativistic effects near the speed of light (i.e. that mass and time dilation approach infinity) are counterintuitive; no one actually understands dark matter,[122] and even the simple Venturi effect (where the constriction of a conduit both speeds the flowing fluid and reduces its pressure) seems peculiar.[123] For millions of people, aromatherapy seems reasonable, but science doesn't support the notion,[124] any more than it does commonly accepted practices in naturopathy[125] homeopathy,[126] or e.g., the anti-vaccine movement.[127] There is "overwhelming

---

121 "The most reasonable possibilities often turn out not to be the situation," Richard Feynman, *Probability and Uncertainty*, in THE CHARACTER OF PHYSICAL LAW 141-142 (1994). "One odd characteristic [of advances in science] is that they often seem to become more and more unreasonable and more and more intuitively far from obvious." *Id.* at 121. "[T]he non obvious is so often true," STEPHEN JAY GOULD, FULL HOUSE 155 (1996) (the book explains how the popular conception of Darwinian evolution—millennia of ever-upward progress leading to the current pinnacle of achievement of mankind—is entirely wrong).

122 Brian Resnick, *Dark Matter Holds Our Universe Together. No One Knows What It Is*, VOX (November 25, 2020), https://www.vox.com/science-and-health/21537034/dark-matter-unexplainable-podcast [https://perma.cc/C7F6-WF3F].

123 *E.g.*, Hayes Spray Gun Co. v. E. C. Brown Co., 291 F.2d 319, 321 (9th Cir. 1961); Melancon v. W. Auto Supply Co., 628 F.2d 395, 398 (5th Cir. 1980).

124 Joy Victory, *Why Health Claims About Essential Oils Deserve More Scrutiny from Journalists*, HEALTH NEWS REV., (Oct. 25, 2017), https://www.healthnewsreview.org/2017/10/why-health-claims-about-essential-oils-deserve-more-scrutiny-from-journalists/ [https://perma.cc/UVE7-9YGS].

125 BRITT MARIE HERMES, *An Inside Look at Naturopathic Medicine: A Whistleblower's Deconstruction of its Core Principles*, in PSEUDOSCIENCE: THE CONSPIRACY AGAINST SCIENCE 137 (2019); Kimball Atwood, *Naturopathy, Pseudoscience, and Medicine: Myths and Fallacies vs Truth*, 6 MEDGENMED 33 (2004), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1140750/ [https://perma.cc/R8M8-FNZT].

126 FERNANDO BLANCO et al. *The Illusion of Causality: A Cognitive Bias Underlying Pseudoscience*, in PSEUDOSCIENCE: THE CONSPIRACY AGAINST SCIENCE 45, 56 (2019) (homeopathy with no effect other than placebo). *See Homeopathy: A Case Study* in EXPERT WITNESSES, *supra* note 4, at §2.10.4.2.1.

127 "Despite scientific studies clarifying that vaccines are not linked to autism in children, 33 percent of parents of children under the age of 18 and 29 percent of all adults continue to believe 'vaccinations can cause autism.'" *Survey: One Third of American Parents Mistakenly Link Vaccines to Autism*, NAT'L CONSUMERS LEAGUE (Apr. 2, 2014), https://nclnet.org/survey_one_third_of_american_parents_mistakenly_link_vaccines_to_autism/ [https://perma.cc/WNC8-AXMY]. *See generally*, JONATHAN HOWARD et al., *The Anti-Vaccine Movement: A Litany of Fallacy and Errors*, in

evidence that commonsense reasoning often contributes to grossly mistaken conclusions."[128] So-called common sense heuristics that serve us well in everyday life "lead us astray in scientific contexts."[129]

Common sense isn't.[130]

2. *Source of Evidence*

Tendencies carried over from evaluations of other sorts of witnesses infect our evaluation of expert scientific witnesses. The test for lay or "percipient" witnesses looks at their personal experience: did they see the accident? Did they personally overhear the conversation or sign the document? For non-scientific experts, too, we look to the witnesses' personal experience and expertise: pilots testifying on the correct way to handle an instrument landing, or doctors and plumbers who testify on standards of care, are themselves the source of the testimony—so it is that their "eminence" might be important, and we might not care much about the ability of the jury to review manuals and reports in the area.

But it's different with scientific experts: unless they themselves undertook the studies the subject of the relied-on reports, they are only conductors of the information. In those instances, we ask the jury to rely on the studies and the literature; not, exactly, the expert herself. Case law, however, does not make this distinction: as with other sorts of witnesses, by habit case law focuses on the human witness, her credentials, and whether her opinion might be impeached.[131] Concomitantly juries, generally speaking, never see the reports: they just see the witness.

But the true import of the scientific evidence is usually the

---

PSEUDOSCIENCE: THE CONSPIRACY AGAINST SCIENCE 195 (2019).

128 SCOTT LILLIENFIELD, *Foreword: Navigating A Post-Truth World: Ten Enduring Lessons from the Study of Pseudoscience*, in PSEUDOSCIENCE: THE CONSPIRACY AGAINST SCIENCE xi (2019) (referring to, among things, the works of Kahneman, *see infra* note 130).

129 EMILO J.C. LOBATO et al, *The Psychology of Pseudoscience*, in PSEUDOSCIENCE: THE CONSPIRACY AGAINST SCIENCE 29 (2019). The infirmities are suffered by scientists, too. *Id.*

130 *See* Curtis E. A. Karnow, *The Temptation of Common Sense*, DAILY J. (2014), https://works.bepress.com/curtis_karnow/14/ [https://perma.cc/H7MG-DHZ3]. What passes for common sense is often the result of cognitive biases. An entire discipline in behavioral economics and related areas has grown to explore this, much of it inspired by the work of Daniel Kahneman such as his book THINKING, FAST AND SLOW (2013).

131 Miranda v. Bomel Construction Co., Inc., 187 Cal.App.4th 1326, 1343 (2010).

substance of the relied-on reports.[132] Scientific experts are not like percipient witnesses. Rules governing the admissibility of their testimony should be more like those governing the *Kelly* "new scientific techniques".

3. *Analogy to Right to Jury*

In *Nationwide Biweekly*,[133] the California Supreme Court recently detailed a test for deciding when claims are for the jury and when they are for the judge. The opinion addresses causes of action, and has nothing to do with experts, but it gives good guidance on the underlying capabilities of the two sorts of decision-makers which is useful here.

In addition to the routine test for whether a claim was triable to a jury when the state was formed in 1850,[134] the Court explains the "gist" test, segregating "legal" from "equitable" claims triable respectively by the jury and the judge. *Nationwide Biweekly* held that judges are uniquely suited to the determination of issues requiring "broad discretion to consider a nonexclusive list of factors"[135] and "expansive and broadly worded substantive standards."[136] This Essay has argued that the criteria for reliable expert testimony are studies and reports which qualify as valid scientific methodology, a series of criteria (present to varying degrees in the studies) such as proper attention to data collection and statistics, general rejection of one-off studies, a connection with an explanatory theory, and so on. These are a broad set of nonexclusive factors, best suited to a judge, and it would be difficult or impossible to draft a useful set of jury instructions on these.

For all these reasons, the sorts of scientific disputes which are likely to erupt in the courtroom, where experts fundamentally disagree on e.g., disease mechanism are, I suggest, analogous to the *Kelly* "new scientific technique" context. The science will usually be at the current edge of research, with studies and reports coming out on different sides, literally the competing scientific approaches

---

132 It's not always this straightforward, and to make my point I ignore hybrid contexts such as some medical testimony which for its strength relies both on studies and the personal experience of the testing doctor.

133 Nationwide Biweekly Administration, Inc. v. Superior Court of Alameda County, 9 Cal.5th 279 (2020).

134 *Id.* at 315. The federal historical test of what sorts of claims are triable to a jury isn't much different, but it uses a different year, 1791. *E.g.,* Burch v. P.J. Cheese, Inc., 861 F.3d 1338, 1347 (11th Cir. 2017).

135 *Nationwide Biweekly Administration,* 9 Cal.5th at 326.

136 *Id.* at 327.

contemplated by *Sargon*. Although *Sargon* can easily be read to require the reliability of all these opinions all to be handed over to the jury, I suggest that in those cases the argument for a *judge's* analysis of reliability is as strong as it is in the *Kelly* context.

B. The Judge

It is, then, the judge, who does have access to the studies, and before whom the studies need not be admissible as they do at trial,[137] who must undertake the task. That is, part of the judge's gatekeeper role should include a review of conflicting studies and meta-studies, and the judge may block opinions which rely on, for instance, one study, such as a cherry-picked study. This might block use of a study which was properly done, that is, has true statistical significance, true random allocation of subjects, and the rest. That result likely conflicts with current law, because current law tells judges not to choose sides in scientific controversies and seems to define a plausible (and so admissible) scientific opinion as one which relies on a valid study. Current law passes muster with Popper's falsification approach, but not necessarily Field theory, which in effect asks how the theory (and a study's results) fit into a larger rubric. The trial judge, presumably trained, or at least briefed, on the elements of the scientific method (whether they are the criteria I propose or as established otherwise by case law) determines the legal reliability of the putatively scientific work. This is, I emphasize, a legal decision, not a scientific one, although it of course depends on the law's views—which I have pursued here—on the nature of the scientific method.

In view of the discussion above, the trial judge would, I suggest, evaluate the proffered expert opinion to decide if it is (1) pseudoscience, such as naturopathy, homeopathy, etc.; (2) bad science: studies and reports with no serious statistical analysis, no randomly selected subjects, obvious self-selection bias, studies published in predatory "peer review" journals, and so on; (3) real science but likely wrong, such as reports of faster than light travel;[138] reports which, while seemingly valid

---

137 In pretrial hearings judges may consider evidence not admissible at trial. *E.g.,* In re Paoli R.R. Yard PCB Litigation, 35 F.3d 717, 739 n.4 (3d Cir. 1994); U.S. v. Posado, 57 F.3d 428, 435 (5th Cir. 1995).

138 CHAD ORZEL, *Scientific Failure as Public Good: Illustrating the Process of Science and its Contrast with Pseudoscience*, in PSEUDOSCIENCE: THE CONSPIRACY AGAINST SCIENCE 243 (2019) (experiment with neutrinos initially showing they travelled faster than light, error in equipment later detected). *See also, e.g.,* Geoff Brumfiel, *Neutrinos Not Faster Than Light*, NATURE (16 March 2012),

when examined in isolation, are so outweighed by commonly accepted science that on balance they are untrustworthy at present; opinions based on one or a few studies which ignore contrary studies; preliminary reports, in effect test balloons based on anecdotal evidence, or small studies, or animal studies where the *relevant* similarities between humans and animals are not evident; and (4) results shown with true scientific methodology, with results buttressed by multiple studies, or meta-studies, and where opinions account for contrary studies; also opinions based on new studies which contradict an established consensus, where the new studies reveal significant flaws in the studies which supported the established consensus.

The jury would see only opinions based on the last category (4) of studies.

As intimated in my discussion of *Nationwide Biweekly*, this can be a complex task: there are many criteria, some of them may point in different directions, and some are extant along a spectrum. One may ask whether the theory meets the falsification test, whether the results mesh with accepted theories and findings under Field theory, or simply whether the one or two studies offered as basis for expert opinion were more than preliminary scientific efforts. The issue, I have argued, is not so much whether the work is scientific (as current law seems to suggest), which is not too hard to decide when done by eminent, patently qualified scientists. The issue is more difficult: it is whether the scientific work has advanced to scientific, and legal, reliability, considered in context of the rest of the scientific work in the area.

My fellow judges may not thank me for the additional work I suggest they shoulder, generally to be handled in a pretrial hearing. Most judges, including me, find great comfort in the use of juries: they handle the difficult, sometime inchoate, incalculable decisions of weight and credibility of evidence, forging a simple verdict from a grand, complex, and conflicting amalgam of facts and law. No one really wants to know how that sausage is made. But in the end, someone—the judge or the jury—must surely undertake detailed reliability evaluations of scientific opinion. Better, surely, to have the judge do this work, done in open court, noting her findings on the record or written order, overtly considering the criteria for scientific methods and reliability, with full access to all studies

---

https://www.nature.com/articles/nature.2012.10249 [https://perma.cc/5PMM-TY4X].

at issue, than to have the jury do it, *sub silentio,* with little or no access to the studies, evaluating (or not) the reports, or more likely someone's summary of the some of the reports, in the secrecy of the jury deliberation room.[139]

VI
IMPACT ON CURRENT LAW

To put some perspective on the problem with current law, we should review the impact of ordinary proof requirements at trial with the rules governing the admission of expert testimony. Assume a plaintiff's expert wishes to testify that a [fictitious] chemical, say, Soylant Green, causes heart attacks. She relies on one study, well done, from a reputable source, peer reviewed, with a p-value showing statistical significance. Or she relies on a series of in vitro and animal studies.[140] The other side—the manufacturer of Soylant Green—has dozens of studies which don't show the correlation; perhaps there's also a meta-study which reaches that conclusion, too. Under current law, the state trial judge is likely to (1) admit the expert's testimony, because the judge has been told not to resolve scientific controversies, and (2) tell the Soylant Green company to try to impeach the expert with the contrary studies. The judge may find that the expert's opinions, based on the one good study or the animal and in vitro studies, are not "*clearly invalid and unreliable,*"[141] and so the judge perforce must pass the torch on further findings on reliability to the jury. Because the "testimony of a single witness may be substantial evidence, including the testimony of an expert,"[142] the jury may rely on the one expert, ignore *all* the rest of the evidence, and still be held to have returned a verdict based on

---

139 My suggestions are likely consistent with those of an important paper, David L. Faigman, et al., *Gatekeeping Science: Using the Structure of Scientific Research to Distinguish Between Admissibility and Weight in Expert Testimony*, 110 NW. U. L. REV. 859 (2016). In brief, the authors suggest that all reliability issues concerning scientific experts should be handled by the judge, with the exception of "factual disputes that relate solely to the case at hand are for the jury to assess. Thus, whether an expert in the instant case actually applied the methodology that the judge found valid generally is a matter of weight [for the jury], as is any conclusion the expert reaches that is applicable only to the litigants." *Id.* at 865.

140 Roberti v. Andy's Termite & Pest Control, Inc., 113 Cal.App.4th 893, 903–04 (2003) (because *Kelly* "new technique" was not at issue, trial judge should not have rejected expert testimony based on animal and in vitro studies).

141 Cooper v. Takeda Pharmaceuticals America, Inc., 239 Cal.App.4th 555, 590 (2015) (quoting and emphasizing *Sargon*, 55 Cal.4th at 772).

142 Johnson & Johnson Talcum Powder Cases, 37 Cal.App.5th 292, 314 (2019).

"substantial evidence"[143] which will not be reversed on appeal. That is, even weak expert testimony, outweighed by very strong contrary evidence, is enough to sustain a verdict.[144] And of course, precisely the same evidence from the same experts may lead to opposite verdicts in different trials, one jury finding Soylant Green causes heart attacks, and another unconvinced.[145] Both verdicts may be upheld on appeal.

That's current law.

Under the proposals of this Essay, it is unlikely the jury would see the one-off study, given the context of the other studies; or indeed, if there were no other studies. It's not that plaintiff is wrong that Soylant Green is a killer; we just don't know, yet, that she's right.

CONCLUSION

Current law on the admissibility of expert scientific option is not coherent. It relies on but does not define "science," and

---

143 The notion of "substantial evidence" is one of life's little mysteries. As with many terms in the law, "substantial" doesn't have the ordinary English meaning of e.g., "large in size, value, or importance," as manifested in dictionary examples such as "The findings show a substantial difference between the opinions of men and women. She inherited a substantial fortune from her grandmother. The first draft of his novel needed a substantial amount of rewriting." *Substantial*, CAMBRIDGE ONLINE DICTIONARY. Rather, it's just *some* evidence, "such relevant evidence as a reasonable [person] might accept as adequate to support a conclusion," as opposed to "speculation and conjecture." Flagship Theatres of Palm Desert, LLC v. Century Theatres, Inc., 55 Cal.App.5th 381, 413 (2020) (internal quotes removed); Frausto v. Dep't of California Highway Patrol, 53 Cal.App.5th 973, 996 (2020). "Substantial evidence" can be disputed and contradicted. Tansavatdi v. City of Rancho Palos Verdes, 60 Cal.App.5th 423, 438 (2021); Pilliod v. Monsanto Co., 67 Cal.App.5th 591, 282 Cal. Rptr. 3d 679, 704 (2021). For these reasons, as soon as the judge admits an expert opinion—by definition having found it not speculative—the 'substantial evidence' threshold is met. And when that's true, the verdict can't be reversed for e.g., insufficient evidence. Shirvanyan v. Los Angeles Cmty. Coll. Dist., 59 Cal.App.5th 82, 98 (2020); Enplas Display Device Corp. v. Seoul Semiconductor Co., Ltd., 909 F.3d 398, 406 (Fed. Cir. 2018).

144 To be sure, a motion for a new trial may be granted, in which the judge sits as the "13th" juror and is entitled to give no weight to the winning party's evidence. CCP §§ 657 *ff.*; Diemer v. Eric F. Anderson, Inc., 242 Cal.App.2d 503, 505 (1966); Ryan v. Crown Castle NG Networks Inc., 6 Cal.App.5th 775, 784 (2016). But what then? Suppose the second jury does as the first? Will the judge keep granting new trials until some jury does as the judge desires? This seems an abuse of the new trial motion procedure.

145 Although this is true, the practical risks for the defendant are *far* more serious. A dozen juries may find for Soylant Green in a dozen trials with a dozen different plaintiffs, but if one jury finds for one plaintiff, the doctrine of offensive collateral estoppel may bar Soylant Green from contesting the issue in all subsequent cases. *See generally, e.g.,* Imen v. Glassford, 201 Cal.App.3d 898, 906 (1988); Smith v. ExxonMobil Oil Corp., 153 Cal.App.4th 1407, 1414 (2007); RESTATEMENT (SECOND) OF JUDGMENTS, § 29 (Issue Preclusion in Subsequent Litigation with Others) (1982).

juries are asked to evaluate expert opinion which relies on studies without access to the studies. Current law allows juries to base their verdicts on testimony which, while "scientific" in the sense that it may be the work of genuine scientists, relies on scientifically unreliable reports.

The distinction in state law between the scrutiny given to (i) new scientific techniques under *Kelly* and (ii) other scientific evidence is not generally supportable, although this Essay does not argue that the scientific evidence must be commonly accepted (as in *Kelly*), only that the same relatively deep level of scrutiny is appropriate.

These problems can be ameliorated by establishing criteria for the scientific method; the Essay proves a first approximation of those. The test for scientific reliability based on scientific methodology comprises many elements, to various degrees; requires close review of the underlying studies as well as other studies and meta-studies in the area; and is not suited to jury determination. Rather this is the work a judge should do in the judge's "gatekeeper" role prior to trial.