Scientific Evidence: A Need for Caution in Decision-making

Jonathan Beach*

Presented at the 'Judicial Reasoning: Art or Science ?' Conference (February 2009) National Judicial College of Australia/ ANU College of Law/Australian Academy of Forensic Sciences

Techniques of judicial reasoning are context dependent. So:

(a) non-expert fact finding uses rational epistemic methods to resolve competing representations of historic facts; expert fact finding involves similar methods but tailored to resolving competing theories or applications;

(b) contrastingly, applying legal principle to facts to draw conclusions is a largely deductive exercise, provided that the degrees of freedom are minimal and no discretion is involved; but for greater degrees of freedom or discretion, the deductive approach shifts to a "goal oriented" rational synthesis of potentially dissonant elements;

(c) dissimilarly, developing a common law principle is an inductive exercise[1], with rational values employed such as, inter alia, internal consistency, coherence with other legal principles, greater simplicity and unification to justify any step forward.

The purpose of my paper is to focus on judicial reasoning which analyses scientific evidence. Unsurprisingly, the epistemic lens applied to that task correlates with that used by scientists. Whether judicial reasoning is an art, and problems with that predicate, are issues defined out of that context. For my context, judicial reasoning resonates with the epistemological themes that underpin the scientific method. But such reasoning requires a circumspect application of some of these overlapping epistemic values.

Now, there are different modes for judicial analysis of scientific evidence, viz:

(a) The judge might hear expert evidence in the usual way. And one way to proceed may be for one of the experts (as agreed) to give a relevant presentation on the non-contentious theories and to explain the general science relevant to the questions.[2] This may assist understanding of the context and content of the contentious scientific issues.

---

[1] There are several dimensions. One dimension takes individual cases and fact specific results to support the new induction. The other dimension, which is a derivative, takes separate legal principles (themselves previously formulated by separate inductions) and seeks, inter alia, to unify them under a broader induction.

[2] This procedure was used in the Longford Royal Commission (1999). There seems to be no good reason why such a procedure could not usefully be employed in litigation. The presentation could be given in a sworn form at the outset of all evidence on the issue, with the right to supplement it or subject it to cross-examination at a later stage.

*Queen's Counsel, Victorian Bar

(b) The judge might obtain assistance from an expert assessor/assistant who may assist with any explanation of the scientific evidence; see for example Section 217 of the *Patents Act* 1990 (Cth)[3] and Order 34B of the Federal Court Rules (the latter by consent only). So a similar function may be performed to (a).

(c) The judge might receive a report on the scientific question from an independent expert appointed by the court. Procedural rules such as Order 34 of the Federal Court Rules allow for such appointments; this mode is to be expanded in the Federal Court context to referees.[4] But this procedure may not have the same direct educative function because the process is, in the first instance, remote from any direct judicial analysis of the scientific question. And that may create difficulties if a challenge to the independent report is made; at that later time the judge may not have the same level of educative assistance available.

But whatever the mode, this does not change the epistemic values and method used to assess scientific evidence. Further, whatever be the role and standard of performance of the advocates in challenging scientific evidence, judges must apply their own questioning epistemic lens to any such analysis. The advocate's lens has an astigmatic characteristic which is necessarily absent from a judge's lens.

It is the principal purpose of this paper to discuss some themes relevant to the judicial lens to be applied to scientific evidence. For convenience, this paper is divided into the following sections:

A. Epistemic Lenses:
   - I General;
   - II Lay witnesses;
   - III Scientific witnesses.

B. Scientific theories:
   - I The tentative nature of any scientific theory;
   - II Science – a rational objective activity?;
   - III Contestable theories;
   - IV Models;
   - V Mathematical components of a theory;
   - VI Statistics;

---

[3] Heerey J used such an assessor in a patent case involving recombinant DNA techniques: *Genetic Institute Inc v. Kirin-Amgen Inc (No 2)* [1997] FCA 1058 and *(No 3)* [1998] FCA 740. See also the Hon. Michael Black: *Genetics in the Courtroom* (2003) 26(3) UNSWLJ 755 and *Beecham Group Ltd v. Bristol Myers Co (No 2)* [1980] 1 NZLR 185 at 192.

[4] See Items 6 and 7 of Schedule 1 to the *Federal Justice System Amendment (Efficiency Measures) Bill (No 1) 2008* and paragraphs 22-32 of the associated Explanatory Memorandum.

VII The theory laden quality of observation;

VIII How to assess the reliability of a scientific theory or its application;

C. Conclusion.

A Epistemic Lenses

I General

At a general level there are differences between legal and scientific inquiry and reasoning. For example:

(a) First, the objectives differ. Science searches for increased knowledge, with truth as the ideal; its coverage is more comprehensive and not time sensitive. But in the forensic legal context, the law's objective is to resolve conflicts; its coverage is limited and time sensitive. And in terms of expounding legal principle, the law is concerned with regulating human affairs in accordance with certain values and objectives. So it deals with what "ought to be". But, as Korn explains,[5] science is purely descriptive of the natural world. Scientific laws do not have the objective of controlling or judging. Rather they describe and explain. But at the general level of their respective theories, each discipline provides a body of "generalized, systematized, and transmissible knowledge",[6] although the inductive processes used to produce such knowledge differ.

(b) Second, scientific theories are always at risk of being falsified. So scientific "truth" is contestable, revisable and consequently always tentative.[7] Contrastingly, legal principles are at risk of being found to be irrelevant or re-interpreted.[8] Further, forensic "truth" is static.

(c) Third, in scientific theory development, the facts relevant to an old theory cannot contradict the new theory. Moreover, the new theory ought also explain all the facts explained by the old theory. But for the development of legal principle, the facts of an older case are either irrelevant or at most of secondary significance in understanding the foundation of the principle which is to be modified.

(d) Fourth, the pattern of legal reasoning is by way of example. Examples are used by lawyers to support a principle. But falsification of an example does not negate the principle. But in scientific reasoning, theory is not developed from ad hoc example. Moreover, one falsified deductive consequence negates the theory.

(e) Fifth, forensic analysis in litigation proceeds by way of a dialectical process. Further, the analysis takes place within the idiosyncratic constraint of how the issue is

---

[5] Korn, H.L.: "*Law, Fact, and Science in the Courts*": Columbia Law Review (1966): 66(6) 1080 at 1093-4.

[6] Korn (supra) at 1101.

[7] Foster and Huber have expressed the view that "the ultimate test of the 'validity' of a theory or of some data is time". Science is said to be "cumulative" and "self correcting". Foster, K.R. and Huber, P.W. (1997): "*Judging Science: Scientific Knowledge and the Federal Courts*" p17 MIT Press, Cambridge.

[8] Fienberg, S.E. and Straf, M.L.: "*Statistical Assessments as Evidence*": J.R. Statist. Soc. A (1982) 145(4) p410 at 414-5.

formulated and evidence is adduced. Contrastingly, for scientific analysis, there is only one epistemic objective. Further, the inquiry is not so issue or evidence constrained.[9] Further, the method of reasoning is more linear, viz:

Observation →Hypothesis →Testing →Theory →Further confirmation or attempted falsification.

Now accepting that there are these and other general differences does not gainsay the point that when judicial analysis of scientific evidence is required, the epistemic values that need to be applied correlate with values underpinning the scientific method. An epistemic lens needs to be used which provides a sophisticated picture of the contextual development, content and reliability of the scientific evidence and any theories or applications included within it.

A failure to analyze scientific evidence through the appropriate epistemic lens can lead to outcomes "determined by intuitive perceptions of the weight of authority rather than by reasoning from evidence".[10] And complex issues of technology and science require reasoning not just from facts to law but reasoning among purely technical facts analyzed through complex models and (in some cases) statistical inference.[11] Indeed, in many cases mathematical instruments are required which have no direct comparator in legal reasoning. The language, premises and analytical styles between the scientific method and the judicial method have their differences. But when the latter is required to evaluate the former, correspondence of epistemic values arises. Judges may be required to "retrace and evaluate the technical analysts' logic, from empirical data to subjective judgment".[12] Now how is this to be done? And what epistemological themes should underpin that appraisal?

II Lay witnesses

Now to suggest the use of an epistemic lens to analyze scientific evidence should not be surprising. Non expert evidence is assessable through an epistemic lens. For example, one way to analyze the evidence of a lay witness is to:

(a) first, examine his sworn evidence in chief to assess the prima facie plausibility of each component separately and cumulatively[13], including internal consistency;

(b) second, compare each component against:

 (i) other versions given under cross-examination, including assessing the plausibility of those other versions;

---

[9] There may be practical constraints (e.g. funding) or paradigm constraints for scientists conducting normal science within the bounds of their discipline.

[10] Yellin, J.: "*High Technology and the Courts: Nuclear Power and the Need for Institutional Reform*": Harvard Law Review (1981): 94(3) 489 at 491.

[11] Yellin at p495.

[12] Yellin at p520. Yellin's context encompassed U.S. judicial review of regulatory decision-making, but his comment applies *a fortiori* to direct factual adjudication between competing scientific evidence.

[13] Perhaps a consideration of what might reasonably have been expected to have been said by the witness but was not said may also go to this aggregate assessment.

(ii) previous written or oral statements made by the witness;

(iii) the broader contemporaneous documentary record; and

(iv) facts not in dispute;

(c) third, compare each component against the evidence of other lay witnesses similarly subjected to steps (a) and (b);[14]

(d) fourth, compare each component against any expert evidence filtered through a different but related epistemic lens (to be discussed);

(e) fifth, and as a consequence of the above steps, form a prima facie view about the acceptability of part or all of the evidence of that lay witness;

(f) sixth, form a prima facie view about the acceptability of part or all of the evidence of other lay witnesses by applying similar techniques;

(g) seventh, determine the best fit for the views reached in steps (e) and (f) – this may require several iterations;

(h) finally, accept or reject part or all of the evidence of that lay witness by reference to this best fit.

Moreover, within this lens there may be sub-elements. For example, plausibility may examine whether:

(a) the level of detail is what one would expect from the witness as a function of either the time that has elapsed since the relevant events or the capacity to refresh memory from legitimate sources (as distinct from reconstruction based upon a hand-held course in document education);[15]

(b) the evidence is consistent with the physical or attitudinal perspective of the witness at the time apparent personal knowledge was said to be acquired.[16]

(c) the evidence is for or against the interests of the witness or those with which the witness perceives he is associated;[17]

Now this is one possible lens. Other lenses can be used.[18] But my purpose is not to focus on appraisals of lay evidence but rather scientific evidence to which I will now turn.

---

[14] Either too little correlation or too much correlation between the various versions may raise forensic doubts.

[15] Within this category, for example, one would include a non-responsive answer, alternatively an exaggerated version, alternatively a version which had a surprising level of detail. These content questions go directly to plausibility.

[16] Ready examples of the importance of physical perspective relate to witness observation. Less obviously, the importance of attitudinal perspective may arise, for example, where there are conflicting versions of what was said at a meeting, where the content and power of recollection is a function of the contemporaneous attitudinal perspective of each witness (the *Occidental and Regal v. Bank of Melbourne* (1992) litigation provides a good example where multiple and conflicting versions were given by numerous witnesses concerning what occurred at a settlement meeting).

[17] Divergence between version and interest may enhance plausibility; consistency may detract.

[18] For example, other lenses may incorporate the notion of "witness demeanour", but this may be of limited forensic utility (save and except a non-responsive answer that some might class as a demeanour

III Scientific witnesses

At the federal level and in some states, the statutory requirement for the admissibility of expert evidence is contained in Section 79 of the *Evidence Act* 1995 (Cth) or state equivalents.[19] The expert must have "specialised knowledge" and the relevant opinion must be wholly or substantially based upon that knowledge. So there must be a field of "specialised knowledge" in which the witness "by reason of specified training, study or experience" has become an expert (*Makita (Australia) Pty Ltd v. Sprowles* (2001) 52 NSWLR 705 at [85] per Heydon JA). There is also a form requirement. Any expert report has to be presented in a form which enables such requirements to be assessed and satisfied (*HG v. The Queen* (1999) 197 CLR 414 at [39] per Gleeson CJ). More specifically, scientific witnesses must furnish a judge with "the necessary scientific criteria for testing the accuracy of their conclusions" (*Davie v. Lord Provost, Magistrates and Councillors of the City of Edinburgh* [1953] SC 34 at 39-40, per Lord President Cooper, cited with approval by Heydon JA in *Makita* at [59]). Moreover, any "inability to articulate the principal tenets that need to be understood, to describe in ordinary language the methods used and the reasons that point to a particular conclusion" are the "hallmarks of unreliable science and the not-so-qualified expert" (*Lewis v. The Queen* (1987) 88 FLR 104 at 123-4 per Maurice J). So "the opinion of an expert requires demonstration or examination of the scientific or other intellectual basis of the conclusions reached" (*Makita* at [85] per Heydon JA).

I will return to the concept of "specialised knowledge" in Section B III below. But assuming that these general requirements for admissibility are satisfied, how should scientific evidence then be assessed? What epistemic lens should be used for analyzing its subject matter and underlying theories? The elements of a lens that I want to suggest, include considering:

(a) the tentative nature of any scientific theory;

(b) the sociological framework within which theories are constructed and contested;

(c) theory development as otherwise than a purely rational objective activity;

---

issue). First, objective demeanour is a function of psychological state which is unknown. To seek to infer the subjective from the objective when the objective base is itself nebulous may be an exercise in futility or at the least apt to mislead. Moreover, reasonable minds may differ regarding analysis of the objective signs. For example, long pauses may be interpreted as a sign of evasion, alternatively as a sign of careful reflection. Anxiousness to assist may be perceived as self-serving, alternatively a sign of nerves, alternatively a selfless desire to assist. More generally, any assessment of personality manifestations is informed by the idiosyncratic value system of the trier of fact and so further unreliability is injected. Second, assume that the elements of the above forensic lens (or something similar) are applied and that they suggest acceptance or rejection of part or all of the evidence of the witness. Such an analysis is likely to bias any additional "demeanour" assessment. But regardless, if the "assessment" of demeanour is consistent with such rejection or acceptance, what does it meaningfully add? And if inconsistent, should not that demeanour assessment be discarded as being less helpful than the more rigorous epistemic analysis? Assessment of demeanour is a forensic tool of last resort in the epistemic ranking.

[19] "If a person has specialised knowledge based on the person's training, study or experience, the opinion rule does not apply to evidence of an opinion of that person that is wholly or substantially based on that knowledge". The opinion rule (s.76) is the general exclusionary rule applying to opinion evidence.

(d) the contestable nature of theories;

(e) how scientific models are used for both explanatory and predictive power;

(f) how the significance of mathematical or other quantitative components of scientific evidence should be assessed;

(g) the theory laden quality of observation;

(h) some techniques for assessing the reliability of a theory.

Now these elements are generic rather than specific to particular scientific fields. But without the focus provided by such a lens, detailed examination of the weight and reliability of the scientific evidence in an individual case may not proceed with the necessary degree of informed circumspection.

B Scientific Theories

I The tentative nature of any scientific theory

Any scientific theory, no matter the degree of its acceptance, is necessarily tentative and revisable. That proposition is established by what defines "science". The demarcation between science and other fields of human inquiry is a contentious issue. There is no satisfactory definition. But there is some common ground among lawyers, scientists and philosophers of science. The scientific method involves the positing of testable hypotheses based upon empirical observation. But any theory formed is only tentative. Any theory gives rise to the problem of under-determination. For any set of empirical observations, there is more than one theory which can explain them or is empirically adequate. All scientific theories are tentative and revisable, as the history of science demonstrates. Indeed, history provides support for the pessimistic meta-induction argument. The induction that all our scientific theories today will be replaced by later "truer" theories is supported by what has occurred for all past theories.[20]

Some U.S. Courts have discussed what is "science" when ruling on the constitutionality of state legislation which mandated a balanced treatment for the teaching of "creation-science". For example, Judge Overton in *McLean v. Arkansas Board of Education* 529 F. Supp 1255 (1982) formulated five characteristics of a science or a scientific theory, viz, that it is:

(i) "guided by natural law";

(ii) "explanatory by reference to natural law";

(iii) "testable against the empirical world" (a concept derived from the philosophy of verifiability and more generally logical empiricism);

[20] As each present or past theory is itself an induction, the global statement is a meta-induction.

(iv) falsifiable (a concept derived from Karl Popper's philosophy of falsifiability, viz, that any scientific theory has to be, in principle, falsifiable);

(v) tentative.

On this test, the relevant Arkansas Act, which required such balanced treatment, was ruled unconstitutional. Creation science conveyed "an inescapable religiosity" and did not satisfy all five elements. Rather, it was a "religion". In 1985, a similar Louisiana Act was ruled invalid by Judge Duplantier in *Aguillard v. Treen* 634 F. Supp 426.[21] This matter came before the U.S. Supreme Court (*Edwards v. Aguillard* 482 US 578 (1987)) where the majority (7 to 2) ruled that the Louisiana Act violated the Establishment Clause of the First Amendment.[22] But at no level in this later litigation did any court rule upon the essential characteristics of a science, even though 72 Nobel laureate scientists and 24 state academies and other organizations of science submitted various amicus curiae briefs to the Supreme Court which sought to define the scientific method.[23] The majority avoided the question, ruling that the Act was invalid on purposive grounds to be gleaned from the statute's text and context.[24] The minority (Scalia J, with whom Rehnquist CJ joined) also avoided giving any definition. But they referred to some of the evidence suggesting that "creation-science" had some elements of a science (properly so called) and that "evolution-science" was a theory rather than a proven fact. Later, the Supreme Court in a separate context in *Daubert v. Merrell Dow* 509 US 579 (1993) considered what the scientific method involved.[25] The Court considered the adjective "scientific" and endorsed Popper's philosophy of falsifiability [31]. But at best this is only a necessary rather than a sufficient condition for defining the scientific method.[26] Moreover, the Court did not set out

---

[21] Shermer, M.B.: "*Science Defended, Science Defined: The Louisiana Creationism Case*" Science, Technology and Human Values, (1991) 16(4) 517 at 522.

[22] This forbids the enactment of any law "respecting an establishment of religion".

[23] See Shermer at p529. Their themes were as follows. The scientific method involves the making and accumulation of observations of the natural world and its phenomena. Such observations give information about underlying facts. On the basis of well-established facts, testable hypotheses are formed. An hypothesis is transformed into a theory when it "explains a large and diverse body of facts" and "consistently predicts new phenomena that are subsequently observed". But "even the most robust and reliable theory … is tentative. A scientific theory is forever subject to re-examination".

[24] The Court applied a three-limbed test for invalidity, one element being whether the legislature adopted the law for a secular purpose. The majority held that the Act lacked a clear secular purpose.

[25] The context was the admissibility of expert evidence in litigation involving the alleged effects of a prescription drug causing serious birth defects in children of mothers who had ingested the drug.

[26] Not only is falsifiability only one of the elements suggested above, it is an imperfect "test". First, scientific theories are usually only testable in combination with the utilization of auxiliary hypotheses, theories and a metaphysical framework. For example, Newton's law, F = MA, is not testable without a metaphysical theory for two of the variables, a theory of testing and a theory of measurement. Moreover, if an observation falsifies the deductible consequences of the combination, there is always an issue as to which component of the combination has been falsified; it may not be the theory being tested but rather the auxiliary hypotheses, testing procedures or initial conditions. Second, falsifiability is not a sufficient test for demarcation of a scientific claim from a religious claim. Many claims in religion are in principle falsifiable, e.g., Noah's flood; *Fasold v. Roberts* [1997] FCA 439 turned on different issues, though its sub-stratum touched upon the underlying debate about evidence of the flood. Third, normal science and its development does not proceed in the idealized manner suggested by Popper, as Thomas Kuhn later explained. Moreover, much of scientific practice appears to relate to the seeking of confirmatory evidence rather than falsifying evidence. Fourth, many theories have been accepted even though at the time they were apparently falsified. Newton's inverse square law was accepted even though at the time it was posited it was apparently refuted by the orbit of Uranus. Later it was discovered that the description

Popper's detailed propositions which contextualized and explained the phrase cited by the Court. Popper said:[27]

> "I often formulated my problem as one of distinguishing between a genuinely empirical method and a non-empirical or even a pseudo-empirical method – that is to say, a method which, although it appeals to observation and experiment, nevertheless does not come up to scientific standards …
>
> … If observation shows that the predicted effect is definitely absent, then the theory [in question] is simply refuted. The theory is *incompatible with certain possible results of observation* … This is quite different from the situation … when it turned out that the theories in question were compatible with the most divergent human behavior **[e.g. Popper's reference to Marxism and Freud's theory of psycho-analysis]**, so that it was practically impossible to describe any human behavior that might not be claimed to be a verification of those theories. **[Popper then listed the following elements derived by him as early as 1919-1920]**.
>
> 1 It is easy to obtain confirmations, or verifications, for nearly every theory – if we look for confirmations.
>
> 2 Confirmations should count only if they are the result of *risky predictions*; that is to say, if, unenlightened by the theory in question, we should have expected an event which was incompatible with the theory – an event which would have refuted the theory.
>
> 3 Every 'good' scientific theory is a prohibition: it forbids certain things to happen. The more a theory forbids, the better it is.
>
> 4 A theory which is not refutable by any conceivable event is non-scientific. Irrefutability is not a virtue of a theory (as people often think) but a vice.
>
> 5 Every genuine *test* of a theory is an attempt to falsify it, or to refute it. Testability is falsifiability; but there are degrees of testability: some theories are more testable, more exposed to refutation, than others; they take, as it were, greater risks.
>
> 6 Confirming evidence should not count *except when it is the result of a genuine test of the theory*; and this means that it can be presented as a serious but unsuccessful attempt to falsify the theory. (I now speak in such cases of 'corroborating evidence').
>
> 7 Some genuinely testable theories, when found to be false, are still upheld by their admirers – for example by introducing *ad hoc* some auxiliary assumption, or by re-interpreting the theory *ad hoc* in such a way that it escapes refutation. Such a procedure is always possible, but it rescues the theory from refutation only at the price of destroying, or at least lowering, its scientific status …
>
> One can sum up all this by saying that the *criterion of the scientific status of a theory is its falsifiability, or refutability, or testability*." **[only these last two lines were cited by the Court]**.
>
> **[The following is also from footnote 3.]**
>
> "'Clinical observations,' like all other observations, are interpretations in the light of theories … and for this reason alone they are apt to seem to support those theories in the light of which they were interpreted. But real support can be obtained only from observations undertaken as tests (by 'attempted refutations'); and for this purpose *criteria of refutation* have to be laid down beforehand: it must be agreed which observable situations, if actually observed, mean that the theory is refuted …
>
> … [I]f a theory is found to be non-scientific, or 'metaphysical' (as we might say), it is not thereby found to be unimportant, or insignificant, or 'meaningless', or 'nonsensical'. [footnote omitted] But it cannot claim to be backed by empirical

---

of the initial conditions was false, rather than the theory, because of the non-inclusion of Neptune. Fifth, there are many theories that one would consider to be scientific (and accepted as such) but which have not been able to be verified or falsified. String theory is a case in point.

[27] See Popper, K.: "*Conjectures and Refutations*", fifth edition, (London: Routledge and Kegan Paul, (1989)) p37. Foster and Huber (1997) (supra) at Chapter 3 discuss other concerns with the philosophy of both Popper and Carl Hempel (a logical positivist) who was also cited in *Daubert*.

> evidence in the scientific sense – although it may easily be, in some generic sense, the 'result of observation.'"

And when one appreciates this detail, the statement of Rehnquist CJ (in dissent) at [50] that "I am at a loss to know what is meant when it is said that the scientific status of a theory depends on its 'falsifiability'" loses some of its lustre.[28] I do not need to take falsifiability further in the present context or to discuss the *post-Daubert* U.S. jurisprudence including *Kumho Tire Co Ltd v. Patrick Carmichael* 526 US 137 (1999). It is sufficient for my purposes to adopt one description given by Edmond and Mercer, namely, that recourse to Popper's philosophy of science is appropriate "as a legal literary technology capable of assisting with strategic articulations of science (and non science)".[29] Falsifiability generally is an important emphasis because it overcomes the effects of confirmatory bias. As Foster and Huber explain, scientists, like others, may have a tendency to settle on a theory quite early, and then by doing so look for confirmatory evidence rather than discrediting evidence.[30] Falsifiability is, accordingly, both a test and a check, and coupled with revisability, provides sufficient insight for present purposes.

Generally, what can be distilled from these judicial considerations relevant to the epistemic lens is that any scientific theory must be testable (and consequently falsifiable), but regardless, is only ever tentative and contestable.[31] As said in *Daubert* at [39], scientific conclusions are subject to "perpetual revision". The dogmatism of any scientific witness for a particular theory must consequently be viewed sceptically. The certitude for any theory at any time is problematic. At most, what can be said of any scientific theory is that it is empirically adequate, but only at a particular point in time. And the high point for the empirical adequacy of a theory is only that the theory is consistent with and explains the observed data or observable (not yet observed) data in that same class. But no monopoly on truth is given to any theory. There can be multiple theories of equal and greater empirical adequacy discoverable. Hence the evolution of scientific theories, and their reflection of a continuous Darwinian pattern.

II Science – a rational objective activity?

But not only is science always tentative and revisable. Science is also not solely a rational objective activity faithfully preserving the subject/object distinction.[32] Theory selection is intertwined with the subjective frameworks of the scientists and their communities. Theories are formulated and selected utilizing values that are partly subjective in content, prioritization

---

[28] The matter was remitted to the U.S. Court of Appeals for the Ninth Circuit (43 F. 3d 1311) where the evidence was held to be inadmissible, partly because "no tested or testable theory" was offered.
[29] Edmond, G. and Mercer, D.: "*Conjectures and Exhumations: Citations of History, Philosophy and Sociology of Science in U.S. Federal Courts*": (2002) Law and Literature: 14(2) 309 at 339.
[30] Foster and Huber (1997) (supra) pp44-45.
[31] See also *Tobacco Institute of Australia Ltd v. Australian Federation of Consumer Organizations Inc* (1992) 38 FCR 1 at 7 per Sheppard J.
[32] Indeed, the Copenhagen interpretation of quantum mechanics demonstrates otherwise at the sub-atomic level in any event.

and application. Moreover, there are problems with the concept of objectivity as explained by Edmond.[33] The "extra-social image of objectivity" is "untenable"; "what is considered objective depends on the stance, commitments and assumptions of the observer (or adjudicator)". There is no such "extra-social image" at any level of scientific discourse.

Values, talents and circumstances guide an expert's preference between qualitative or quantitative methodologies. Further, predispositions of an individual scientist may change over time as experience is accumulated.[34] More generally, value-laden or self interested reasons guide prevailing research agendas and how questions are framed.[35] Indeed, research agendas that seek to challenge the prevailing "wisdom" in an area, for example, string theory,[36] may not be funded or the contrary scholars disadvantaged in obtaining research positions.

More generally, "facts" in science are the product of social construction. They are produced "by human agency through the institutions and processes of science".[37] Prior agreements about the correctness of theories, methods, instrumentation, validation and review are necessary. But these are socially derived "through continual negotiation and renegotiation among relevant bodies of scientists" as Banks explains. "Truth" becomes contingent on experimental or interpretative conventions. Moreover, scientific facts are presented and accepted in a social context. Foster and Huber explain that they may have to fit the social objective at hand or the particular paradigm within which the scientists work.[38]

At an even more general philosophical level, Thomas Kuhn[39] expounded the notion of paradigms as a conceptual tool to explain how scientific theories come to be accepted and further evolve. A paradigm in a particular subject area embraces the carrying out of what is described as normal science in that field and its then temporarily embedded theories. But when a crisis in theory development or application occurs, with competing theories put forward to solve that problem, there may be a paradigm shift. More generally, there may be different paradigms at work in the same subject area at the same time within different scientific

[33] Edmond, G.: "*After Objectivity: Expert Evidence and Procedural Reform*": (2003) 25 Sydney Law Review 131. Moreover, if scientific evidence has specifically been gathered for litigation or a theory formulated or developed in that context, *a fortiori* "objectivity" develops an illusory quality.
[34] Kousser, J.M.: "*Are Expert Witnesses Whores? Reflections on Objectivity in Scholarship and Expert Witnessing"*: The Public Historian (1984) 6(1) 5 at 14.
[35] Kousser (1984) (supra) at 13.
[36] Smolin, L. (2007): "*The Trouble with Physics: The Rise of String Theory, the Fall of a Science, and What Comes Next*": Mariner Books; New York.
[37] Banks, N.K.S.: "*Trials and Tribulations: Social Science Evidence, Expert Witnesses, The Voice of Authority and the Discourse of Ideology in the Courts*": [1999] Mur UEJL 42 at [19].
[38] See Foster and Huber (1997) (supra) at 24-27.
[39] Kuhn, T. (1962): "*The Structure of Scientific Revolutions"*: University of Chicago Press (3rd Ed., 1996). Kuhn's philosophy, which posited a more sociological and relativistic picture of theory development, may be contrasted with Popper's philosophy which posited theory development as proceeding in a more homogenous and rational way through repeated linear applications of falsification, with each new successor theory falsifying its predecessor (or reducing its predecessor to a special case) whilst increasing the scope of its empirical adequacy, and in turn subjecting itself to being falsified.

communities. Edmond and Mercer give an example of this in the field of epidemiology which displayed opposing views on the standards for epidemiological analysis to be reliable in the Bendectin/Debendox litigation; the various knowledges also came with "no 'neutral' means of determining their appropriateness, comparative advantages and limitations".[40] Generally, "accepted" scientific theory used by a witness at a particular point is a function of the particular paradigm, itself a function of the particular sociological conditions applying to the scientific community within which that scientist works. Further, evaluation and acceptance of theories is complicated by problems associated with incommensurability, translation and understanding. When different theories purport to explain the same data, conceptual differences may manifest themselves at the semantic level.[41] Incomparability arises. Generally, scientific theories ought to be seen in a relativistic context rather than in some objective absolute framework.

The short point is that any accepted scientific theory at a particular time has not been chosen and accepted on a purely rational objective basis. Its formulation and acceptance is substantially affected by its sociological context. Now this is not to say that theory selection is not substantially a rational process. Undoubtedly, rational epistemic values are at work in theory selection, viz, internal consistency, simplicity, unification, coherence with other theories, explanatory power, predictive power, etc. But it is to say that theory selection and acceptance is affected by non-epistemic contextual factors, notwithstanding the certitude that scientific experts have as to the purely rational objective foundations of their theories.

Generally, sciences are as fragmented and pluralistic as other fields of intellectual inquiry. As Korn[42] explains, there are "wide variations both in the general level of certainty of the various sciences called upon in lawsuits and between the axiomatic and frontier areas" of a given discipline. This is readily apparent in any so-called "battle of the experts".[43] And moreover, if any consensus does exist, it may itself be rationally undermined by the adversarial process which:

(a) deconstructs "facts";

(b) exposes contingencies and hidden assumptions which underlie claims; and

(c) "privileges scepticism over consensus".[44]

But experts may differ not just on theoretical grounds, but also because of the mode in which their expertise is being used. A scientist may have generalized knowledge which is not in dispute, but differences of opinion may arise in its application to litigation facts which are "multifarious, haphazard and often unique" or "incomplete" or require "close questions of

---

[40] Edmond, G. and Mercer, D.: "*Litigation Life: Law-Science Knowledge Construction in (Bendectin) Mass Toxic Tort Litigation*": (2000) Social Studies of Science: 30(2) 265 at 298-9.

[41] Sankey, H.: "*Incommensurability, Translation and Understanding*": The Philosophical Quarterly: 41 (165) at 414.

[42] Korn (1966) (supra) at p1092.

[43] Banks (1999) (supra) at [17].

[44] Banks (1999) (supra) at [18].

judgment".[45] So too, differences may legitimately arise where predictive elements are involved such as risk assessment. But my focus is more on analyzing contentious scientific theory and general application questions rather than differences at a lower level explicable by idiosyncratic, incomplete or vague foundational facts in the individual case. But to talk of applying an epistemic lens to contentious scientific theories or applications presupposes their admissibility in evidence. Is that a good assumption?

III Contestable theories

Let me begin by making a distinction between:

(a) a theory well accepted by the majority practising in an organized field of knowledge such as physics or chemistry;

(b) competing theories within that field representing the separate views of competing paradigms; or

(c) a novel theory within that field that has not yet reached a level of general acceptance within the field or a competing paradigm within that field.

In the U.S., prior to *Daubert*, theory (a) would have been admissible, theory (b) may have been admissible, and theory (c) would not have been admissible under the "general acceptance" test in *Frye v. U.S.* 293 F1013 (1923). Subsequent to *Daubert*, and more importantly under Rule 702 of the U.S. Federal Rules of Evidence,[46] each of theories (a), (b) and (c) would be admissible provided that scientific reliability was established. And in assessing reliability,[47] factors to consider would be testability/falsifiability including actual testing, peer review, known or potential error rates, existence and maintenance of technical standards and whether there was "general acceptance".

But in Australia, at the federal level and in certain states, the relevant test for admissibility is that contained in Section 79 of the *Evidence Act* 1995 (Cth) rather than *Daubert*.[48] And Section 79 does not expressly refer to any concept of reliability. But having said that, the concept of

---

[45] Korn (1966) (supra) at p1092.

[46] The Rule provided that "If scientific, technical or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue …" then it may be admissible. *Daubert* in essence held that *Frye* had not survived the 1975 codification. Rule 702 was altered in December 2000 to expressly refer to reliability at two levels, viz, first at the level of principles and methods, and second, to the application of those principles and methods to the facts of the case – See Freckelton, I. and Selby, H. at [2.10.680]: "*Expert Evidence*": Thomson-Lawbook Co. (looseleaf). This change does not affect my analysis.

[47] I am talking of reliability here in its broader sense rather than the narrower sense, viz, as to just whether the same test carried out numerous times delivers consistency of result. The broader sense includes the concept of validity i.e. whether the theory or its application delivers results the interpretation of which supports the assertion the theory endeavours to prove. So, for an IQ test, the issue is not just whether repeated applications deliver the same result (scientific reliability), but whether the results accurately display a person's intelligence (scientific validity).

[48] Gleeson CJ in *HG v. The Queen* (1999) 197 CLR 414 at [40] (footnote 37) referred to *Daubert* but found it unnecessary to discuss its principles. NSW has similar provisions (*Evidence Act* 1995), so too *Tasmania (Evidence Act* 2001*)*. Victoria has similar provisions (*Evidence Act* 2008), but not yet commenced.

"specialized knowledge" does not exclude a consideration of the *Daubert* factors as matters to be taken into account.[49] Moreover, the language of Section 79 and the U.S. Rule are "similar in content" although there are textual differences.[50] It seems to me that reliability (in its broader sense) would be the or a touchstone. First, the section refers to "knowledge". Philosophically that usually is taken to be justified true belief, although *Daubert* uses the concept slightly differently. The concept of justification entails that there is a conscious rational basis for the belief, with that foundation rationally linked to the holding of the belief; it is not enough to just have a true belief, because you might have a true belief in something based upon purely erroneous grounds and therefore cannot be said to have knowledge. Further, the reference to "true" is not an absolute, but does suggest some apparent correspondence with the real world. Second, the concept "specialised" indicates that it is not just an idiosyncratic characterization. The relevant speciality must have some rational and objective status. Third, the phrase "based on the person's training, study or experience", by referring to the source of knowledge, implies rationality of foundation which has substantial objective components to which the witness must have been exposed. Moreover it is the composite phrase that has to have that foundation. Fourth, the broader context (Sections 79 and 135-137) is the treatment of relevant and sufficiently probative evidence, again suggesting rationality in theory and application. The theme of reliability may be derived from or is at least consistent with the cumulation of these factors. Alternatively, reliability may be at least a predominant factor. And injecting the concept of reliability does not re-write the statutory language but rather provides a test for its satisfaction or a factor to consider. After all, the legislature left it open as to what it precisely meant, indicating that a degree of flexibility was intended. Moreover, as Sections 135-137 contain potential exclusions, this suggests that Section 79 ought not unnecessarily be read down; but equally such an argument could be used for saying that not even "reliability" need be shown for the purposes of Section 79.

The debate is still open concerning the meaning of Section 79.[51] I do not need to resolve it. What can be said is that there is nothing in Section 79 that mandates any "general acceptance" test as a necessary condition for a scientific theory.[52] It would seem, like *Daubert*, that theories

---

[49] Cf. Edmond, G: "*Deflating Daubert: Kumho Tire Co v. Carmichael and the Inevitability of General Acceptance*": (2000) 23(1) UNSWLJ 38.

[50] Odgers, S.J. and Richardson, J.T.: "*Keeping Bad Science out of the Courtroom – Changes in American and Australian Expert Evidence Law*": (1995) 18(1) UNSWLJ 108 at 109 and 128.

[51] See for example *R v. Tang* [2006] NSWCCA 167, *The Queen v. Murdoch* [2005] NTSC 78, *Murdoch v. The Queen* [2007] NTCCA 1 and S. Odgers' discussion in: "*Uniform Evidence Law*": (looseleaf) at [1.3.4260].

[52] Previously, Dixon CJ in *Clark v.Ryan* (1960) 103 CLR 486 at 491 did not mandate any "general acceptance" test. And although he referred to expertise in "the nature of a science", he did not define what this meant. Menzies J at p501 referred to an "organized branch of knowledge" within which the witness was an expert. But this did not necessarily negate the admissibility of a highly contentious theory within such a branch. *Frye* has previously been applied in Australia, as Freckelton and Selby point out ([2.10.140]-[2.10.180]). But the *Frye* test is not mandated, albeit that "general acceptance" is a relevant consideration. See also Odgers, S.J. and Richardson, J.T. (1995) (supra) at pp122-129.

(a) and (b) would be admissible under Section 79. The more problematic area is theory (c). But it would seem that such a theory could also be admitted; see *The Commissioner for Government Transport v. Adamcik* (1961) 106 CLR 292 where a novel theory of the cause of leukaemia was held to be admissible. More generally, the level of acceptance of a particular theory is not determinative of admissibility even if the theory is farfetched or implausible (*R v. Parenzee* [2007] SASC 143 at [62]-[74] per Sulan J), although it may be relevant to weight. Generally, contentious scientific evidence is admissible and appropriate for adjudication (*Chamberlain v. The Queen (No 2)* (1984) 153 CLR 521 at 598-9 per Brennan J).[53] That also applies to novel scientific theories or novel applications and I proceed on that basis. But what are the elements of such theories? How should they be analyzed? What type of circumspection should be used?

IV Models

Scientists use models. Models can be physical (Watson's and Crick's model for DNA) or theoretical. Theoretical models are used by scientists as primarily representational entities to represent aspects of the world.[54] Models can either constitute the scientific theory or they can mediate between the theory and the world.[55] But models, and their corresponding theories, only describe idealized, manufactured and controlled situations. And theories relying upon such models only apply in situations that resemble their models.

Theoretical models use abstract concepts or objects which are usually unobservable, for example, sub-atomic particles. There is a debate between scientific realists and anti-realists about the truth of theories which posit such unobservables. Scientific realists posit the approximate truth of such theories and the ontology of such unobservables. Anti-realists accept the reality of observables and the empirical adequacy of theories (i.e. that they are adequate to explain empirical observations), but are agnostic about the truth of theories and the ontology of unobservables. Such a philosophical debate underlies any scientific theory, although most scientists are scientific realists. But it is important to bear in mind this observable/unobservable distinction when analyzing a theory. Are unobservables used as part of the model? If so, is their ontology asserted? Or are they just used as constructive components upon which the

---

[53] But one caveat may be that in the criminal law sphere, if the expert evidence is as to a scientific method of a novel kind, the Crown may have to demonstrate the "scientific reliability of the evidence" (*Murdoch v. The Queen* [2007] NTCCA 1 at [275]). Such observations, if as to admissibility, may also reflect a consideration of a combination of admissibility and exclusionary rules. But query whether the observation was made as to admissibility or as to weight. See also *Mallard v. The Queen* [2003] WASCA 296 at [270], *Lewis v. The Queen* (supra), and Freckelton and Selby at [12.10.01] and [12.10.10]. I do not need to resolve this.

[54] Giere, R. (1999): "*Science without Laws*": p5 University of Chicago Press.

[55] Crasnow, S.L.: "*Models and Reality: When Science Tackles Sex*": Hypatia: (2001) 16(3) 138 at 142-3. Further, to bridge the gap between high level theory and lower level phenomena, you may have multiple models, viz: (i) the high level theory and its attendant model; (ii) an experimental theory with its attendant model; and (iii) a theory of the base phenomena and its attendant model. Moreover there may be shared structural components between the layers. See generally: Brading, K. and Landry, E. (2005): "*A Minimal Construal of Scientific Structuralism*": Philosophy of Science Assoc. 19th Biennial Meeting: PhilSciArchive (http://philsci-archive.pitt.edu).

model is based, with the model only used for a theory which itself is only said to have explanatory or predictive power in relation to observables? To properly analyze any theory posited by a scientific expert, you need to understand the structure of the associated model used and the ontological status of the entities used. But what are unobservables?

Let me discuss van Fraassen's example of light.[56] Common sense tells us that it is an observable. Many models have been put forward to explain light including the particle theory, the wave theory and a dialectical synthesis of both. But is it an observable at all? Objects are visible when the light switch is on. But what about the light itself? Search lights and car lights produce visible beams. But they do so only in dirty air, not in space or pure air. What is visible are particles that are illuminated, not the light itself. An experiment can be set up to show that you cannot see light itself. Set up a light bulb and lenses so that light shines into a box without illuminating anything in the box or the interior walls. If there is a view-port, you will only see pure darkness. But assume there is a lever that can move objects into view in the box. The objects will be illuminated. But you won't see light as such. The retina of the eye, however, will detect what is a theorized unobservable which has wave/particle duality.

In summary, when a scientific theory is being discussed or analyzed:

(a) it is a model or is based upon a model;

(b) such a model will only deal in an idealized way with observational data (empirical phenomena);

(c) such a model will also posit or rely upon theoretical entities or unobservables that have questionable ontology,[57] and which the history of science shows are replaced in each substantive successive theory.[58]

It is worth making one other point. There is a difference between the use of models in science and their use in economics. Contemporary economics utilizes models. But their function in that context demonstrates that field's pseudo-scientific method.[59] Such models rely upon inadequate data sets, artificial assumptions of behaviour and idealized statistical theory. Moreover, the models are either static or inappropriately simplistic dynamic models. Further, their derivation

---

[56] van Fraassen, B. (2000): "*Constructive Empiricism Now*": Paper delivered to a symposium of the American Philosophical Association, Albuqueque 2000; http://webware.princeton.edu/vanfraas/mss/APPSection II (or www.princeton.edu/~fraassen).

[57] For example, the imaginary gravitational force "at a distance" in Newton's inverse square law, which was replaced by the concept of a force field, and has since then been replaced by Einstein's theory of general relativity which posits "gravity" as the curvature of space and time produced by objects' mass (itself a metaphysical concept built upon yet other theories of unobservables).

[58] There may, however, be some continuity in the structural elements of these unobservables and their relationship with each other in each successive theory (structural realism).

[59] See further on theory use, McMahon, K.: "*Competition Law, Adjudication and the High Court*" [2006] MULR 25, particularly Section IVB – Economic Analysis as Authority and Section IVC – Choice of Economic Theory.

is also value laden by predominantly non-empirical questions.[60]They have little explanatory power; precision regarding empirical observation is absent. Moreover, they have even less predictive power. Any model not rigorously calibrated against, and explanatory of, precise data is devoid of the necessary foundation for predictive utility. But even if that foundation was established, under-determination of theories still arises, particularly when the variables used are magnified in number and inter-dependent. Similar criticisms have also been made of models used in the competing paradigms of climate change theory.[61] But none of this is to criticize the use of models to aid comprehension by providing a conceptual structure within which to place and manipulate empirical data.

V Mathematical components of a theory

Scientific evidence may use mathematical formulae. But several preliminary observations need to be borne in mind when analyzing such use viz:

(a) First, a formula is usually used to explain an apparent regularity between empirical observations. But a formula (and the underlying data) does not of itself demonstrate any necessary causal connection. From a correlation between the movement in two variables, sometimes such a link is inferred. But philosophically this is always contestable in a Humean sense. And practically may be contestable in many individual cases. The correlation may be due to a separate but common cause. Moreover the fact that movement in one may precede movement in the other is only the starting point for the analysis.

(b) Second, a formula only establishes a regularity observed from present or past observed data. Whether it can be used to make a judgment about a possible future regularity may be a contentious point. Moreover, even as an explanation for past data, the formula may have its limitations. Past data upon which the formula is built may only be a limited sub-set of the available data. The complete data set (if available) may change the apparent pattern of that regularity.

(c) Third, and related to the second point, the formula may only be one of many ways to express the apparent regularity of the observed phenomena. So a formula may be a useful explanatory tool to explain observed phenomena, yet as a predictive tool it may be of little utility. For example, take 10 empirical observations which are plotted on a graph showing for each observation variables x and y measurements. Assume that this graphing shows that the 10 points can be connected by a straight line. One might conclude that there is a linear correlation between variables x and y and posit a formula

[60] So, for example, one has the "Chicago School" of economic theory underpinned by the philosophy of promoting consumer welfare through productive efficiency ("market as regulator"), itself an ideologically driven liberal theory – see McMahon (supra). Indeed, some theories even descend to "rules of thumb", a level of imprecision recently criticized by the Australian Competition Tribunal in *Application by Chime Communications Pty Ltd* [2008] A Comp T4 at [60].

[61] Consideration of problems with computer simulations may also be necessary but it is beyond the scope of this paper.

for the straight line function (y = ax + b) to make further predictions. But theoretically there are an infinite number of lines (including curved) that could be drawn to join the 10 points, with correspondingly different formulae. And a further measurement (the 11^{th} point) may show y = ax + b not to hold. More generally, for any set of empirical data, there are multiple potential theories/formulae to explain the same, viz, the under-determination problem. Generally, one theory/formula consistent with the data may usefully explain observed data but may be a questionable predictive tool. Whether it is also a successful predictive tool may only be ascertained by the longevity of its success, e.g. Newton's inverse square law, although this is now seen as a special case of Einstein's field equations supporting general relativity.

(d) Fourth, a mathematical formula strictly takes its subject matter (viz, one or more of its variables) and their measurement as idealized and precise. But empirical data may be imprecise in quality or measurement. At most, any formula is only an idealized representation of the apparent regularity consistent with the model underpinning the theory.

Generally, mathematical equations should be seen in their limited context as imperfect tools. Their apparent precision ought not to be taken as giving a greater air of verisimilitude to a scientific theory (or the results therefrom) than is warranted having regard to all qualitative and quantitative aspects of the theory.

In forensic analysis of scientific evidence, useful steps to analyze a theory's mathematical components include the following:

(i) First, a definition of each of the terms used in a formula (variables or constants), i.e., the relevant dictionary, should be obtained. Further, if any unusual notation is used (e.g. Greek letters), find out what they symbolize. They usually have straightforward conventions for use.

(ii) Second, an explanation of the objective that the formula is seeking to achieve or display should be obtained. For example, in the simple equation "y = ax +b", where "y" is time and "x" is the position of an object, the plain explanation is that the formula shows how the position of an object changes with time; its gradient is the object's velocity.

(iii) Third, an explanation of how the formula achieves the relevant objective may be necessary when the formula is complex. To begin, a further explanation of each of the factors in the formula may be necessary; these are the elements of the formula usually joined together by the arithmetic operators. Now one of the factors may be a constant (a fixed value). In many cases a constant may be readily accepted as based upon prior theory and observation. But there is a need for caution. Sometimes a constant can be used as a "fudge" factor to enable the observed data to "fit" the theorist's formula. Another factor may be a variable (a changing but measurable observation which the dictionary should have defined). Yet another factor may be a combination of one or more variables or constants. This may be readily comprehensible. But some factors may have complexity requiring an explanation. This may arise, for example, where the factor uses derivatives (the change in one variable as a function of a change in another variable). But generally each factor should be able to be explained in straightforward terms by the expert using the formula. And if not, that may indicate inferior expertise. The best experts are usually those that can simplify complexity, including reducing complex terms to comprehensible lay explanations. Now once each factor has been explained, the next step is to enquire as to why they have been put together in the order of the formula. It is pleasantly surprising to find that in many cases, if you have proceeded this far, the justification for the inclusion and ordering of each factor is readily appreciable. Indeed, with the definitions and a general understanding of the factors, you may perceive that the formula's derivation is trivial, at least in hindsight (cf. *AGL v. ACCC* (2003) 137 FCR 317 at [518] per French J).[62]

(iv) Finally, the computations applying the formula to the data are usually non-contentious. In any event, any such difficulties would be disclosed within competing experts' reports and usually then readily resolvable. Scientific debates are not so much about mechanical computational questions, but rather:

A. the validity of the formula; or

B. the utility of the formula.

It is worth briefly elaborating on utility. First, and generally, the mathematics may only be a small part of a predominantly qualitative theory. If so, the use of the mathematics ought not to be seen as adding a greater air of verisimilitude to the overall theory than is warranted. Second, the mathematics may only be useful for understanding a static state, whereas what might be required is an understanding of a dynamic state. In the Longford Royal Commission, which investigated the cause of a gas plant explosion, what needed to be investigated was the dynamic state of the physical chemistry within the plant in the four hours preceding the explosion. But the then mathematics only allowed for static or "snap shot" exercises which were of limited utility. Dynamic mathematic models are now deployed more readily. But where what is being modelled involves numerous variables which are inter-dependent, such models can produce significant errors within very short time runs no matter the integrity of the independently verified initial conditions and no matter how well calibrated with historical data.

VI Statistics

There are at least two types of uses of statistics as evidence, viz:

---

[62] Sometimes, though, the meaning and utility of the mathematical formulae may remain elusive. See for example *Australian Retailers Association v. Reserve Bank of Australia* (2005) 148 FCR 446 at [481]-[485] per Weinberg J.

(a) the analysis of surveys or experiments, to draw cause and effect inferences e.g. the potential mutagenic or carcinogenic effects of a chemical;

(b) the presentation or interpretation of statistics to analyze a particular characteristic of a population.[63]

But no statistical evidence is based upon a purely probabilistic analysis. Issues of substance concerning the data and its context, whether economic, genetic or medical, are "inevitably intertwined with the statistical issues".[64] Another intertwining occurs between a particular theory (scientific or economic) and statistics. So, a theory may be mathematically posited as to the relationship between variables "x" and "y". Such a theory may be described as "deterministic".[65] But in the real world, data on "x" and "y" may be imprecise or difficult to generate. Statistics may need to be gathered and assessed, and assumptions made as to their distributions or pattern.[66] The combination of the theory with the statistics then provides the relevant explanatory or predictive tool.

Now as Imwinkelried explains,[67] any mathematical formula used in statistics may be derived:

(a) by theoretical mathematical reasoning from a priori principles; or

(b) experimentally;

depending upon the particular case. But whichever way, the formula may need to be validated and issues may arise as to whether:

(i) the formula is generally accepted by scientists/statisticians in the field;

(ii) the formula omits factors that are relevant but too soft to quantify;

(iii) the formula is, for example, the multiplication or product rule, but there is no evidence that the probabilities being multiplied are independent;

(iv) the sources for the values inserted in the variables in the formula are unreliable.

One classic use of statistics is in epidemiology.[68] Epidemiology is usually used as evidence to infer a cause and effect relationship between an exposure to a particular chemical and subsequent illness or death. But it is well accepted that epidemiological studies ought to be viewed through a sceptical epistemic lens which considers, inter alia, the following elements relevant to cause and effect inferences:

---

[63] Fienberg, S.E and Straf, M.L. (1982): See footnote 8.

[64] Fienberg, S.E. and Straf, M.L.: "*Statistical Evidence in the US Courts: an Appraisal*": J.R. Statist. Soc. A (1991) 154(1) p49 at 53.

[65] *AGL v. ACCC* (supra) at [495] per French J.

[66] Such an assumption may involve a stochastic conjecture about a particular probability distribution for the data relevant to a particular variable.

[67] Imwinkelried, E.J. (2004): "*The Methods of Attacking Scientific Evidence*": Chapter 6 - Lexis Nexis.

[68] See generally *Seltsam Pty Ltd v. McGuiness* (2000) 49 NSWLR 262 at [138]-[147] per Spigelman CJ; Bradford-Hill, A. (1965): "*The Environment and Disease: Association or Causation*": (1965) 58 Proc. R. Soc. Medicine 295; Evans, A.S. (1976): "*Causation and Disease: The Henle-Koch Postulates Revisited*": (1976) Yale Journal of Biology and Medicine 49: p175.

(a) Whether there is any strength of association and its degree. The higher the risk estimate, the less chance of a confounding element or bias.

(b) Whether any dose response effect is demonstrated. If the level of risk is directly proportionate to the level of exposure, a causal link is more plausible.

(c) Whether the time sequence is correct. The exposure or risk factor must precede the disease. But sometimes this may not easily be resolved. For example, does a particular diet lead to the disease or does the disease also cause the particular dietary habit?

(d) Whether there is specificity. There may be a strong argument for causation where the association between exposure and disease is limited to specific workers and there is no other explanatory factor.

(e) Whether there is any misclassification of subjects. Morling J in *Australian Federation of Consumer Organizations Inc v. Tobacco Institute of Australia Ltd* (1991) 27 FCR 149 at 164 gave the example of the possibility of smokers, who are being surveyed, claiming for social reasons to be non-smokers. Such a misclassification may inflate the calculated risk of lung cancer amongst non-smokers.

(f) Whether other potential confounding factors have been considered.[69]

(g) Whether there is data dredging. If within the one study, many hypothesized associations are tested, then a high or low result in a particular case might be expected to occur by chance.

(h) Whether there is consistency of result over a meta-analysis of two or more epidemiological studies. If similar results apply over similar populations, well and good. If similar results apply over diverse populations, even better.

(i) Whether there is any publication bias. Positive results studies may be written up more than negative results. This may bias any meta-analysis.

(j) Whether animal toxicology studies are consistent with the epidemiology.

(k) Whether there is a plausible or demonstrated human biological pathway such as to provide a foundation for the causal inference suggested by the epidemiology.

(l) Whether there is any relevant analogy. Sir Austin Bradford Hill gave the example: “With the effects of thalidomide and rubella before us one would surely be ready to accept slighter but similar evidence with another drug or another viral disease in pregnancy”.

(m) Whether there is any evidence inconsistent with the association.

As to probabilistic evidence, it is a truism that concentration on mathematical probabilities can prejudice common sense.[70] Wrong intuitions can arise, an example being the so-called “prosecutor’s fallacy” in relation to DNA evidence.[71]

---

[69] Graziano, A.M. and Raulin, M.L. (1993): “*Research Methods: A Process of Inquiry*”: 2nd edition, Harper Collins pp170-182. The authors provide a list of some of the major confounding variables.

In terms of evidence couched in numerical terms, the following issues may generally arise:

(a) First, even accepting the calculation of the number at face value, it may be insignificant when compared with the relevant population. For example, Imwinkelried gives the example[72] of a serologist giving an opinion on blood type relevant, say, to a paternity suit (although DNA testing is now available[73]). If the test shows that there is a 95% probability that the test would exclude the defendant if he were not the father, that sounds impressive. But if the relevant "universe" was two million males in the same city, that would leave 100,000 possible candidates including the defendant who, all else being equal, could be the father. A less impressive statistic.

(b) Second, a data base may be lacking which is needed to evaluate the significance of the test. Say it is asserted that two samples are identical in characteristics or at least indistinguishable. You cannot properly conclude that such samples have a common source without carrying out an extensive empirical investigation. You need a data base of population frequencies for the relevant characteristics before a "match" can be concluded in terms of source. Such data bases are used in the DNA context.

(c) Third, the test result has to be compared with a data base that is comparable. Say a technique has gone through several generations of development. A data base generated from the use of the first generation technique may not be comparable to a test result produced from the use of a third generation technique.

(d) Fourth, the final number yielded and its foundation may be such that the evidence ought be excluded because its probative value is substantially outweighed by the danger that the evidence might be unfairly prejudicial or misleading or confusing (Sections 135-137 of the *Evidence Act* 1995 (Cth)).[74]

Further, another issue arises concerning the quantitative assessment of statistical significance.[75] Statistics can determine the probability of an observed outcome being due to chance rather than real association. As Hill J explained in *Tobacco Institute of Australia Ltd v. Australian*

---

[70] Hodgson, D.H.: "*Scales of Justice: Probability and Proof in Legal Fact-Finding*": (1995) 69 ALJ 731 at 736. See also Murphy J in *Perry v. The Queen* (1982) 150 CLR 580 at 594.

[71] Say one person in a million has a DNA profile which matches the crime stain. Say the accused has a DNA profile which matches that stain. The "prosecutor's fallacy" is to conclude that there is a million to one probability that the accused left the crime stain. But what if you were told that the relevant population consisted of 50 million males. Then it may be said that there is only a 1 in 50 chance that the accused left the crime stain. You need to know something more about the circumstances of the accused and his relationship to the crime scene. See generally Mason P's discussion in *R v. GK* (2001) 53 NSWLR 317 at [47]-[58].

[72] See generally Imwinkelried, E.J. (supra) at Chapter 13.

[73] See for example *R v. GK* (supra). *G v. H* (1994) 181 CLR 387 at 398 discussed a probabilistic analysis, but not concerning DNA tests.

[74] See analogous arguments discussed in the DNA context in *R v. GK* (supra) at [35]-[58] per Mason P and *R v. Jarrett* (1994) 62 SASR 443 at 455. But exclusion is unlikely.

[75] In the U.S. there is a discussion of epidemiology in the Federal Judicial Center's *Reference Manual on Scientific Evidence* in the chapter *Reference Guide on Epidemiology*. See also the discussion by Spigelman CJ in *Seltsam* (supra) at [59] et seq.

Federation of Consumer Organizations (1992) 38 FCR 1 at 60, a null hypothesis is constructed (the opposite of the theory or link sought to be shown); this is the "chance" component. The data is statistically analyzed to work out the probability of the null/chance hypothesis. If this probability (p) is equal to or less than 0.05, the "positive" result is said to be statistically significant. The figure 0.05 means that the relevant result is expected to happen by chance once in 20 times. Flipping it, the probability (p) of a real association is 0.95. The result is said to be shown at the 95 per cent confidence level. But this significance level is only a convention. There is nothing to enshrine it. Some epidemiologists use a P value (for the null hypothesis) of 0.10. Some tort lawyers have suggested, erroneously, that it should be at 0.5.[76] It would seem that there could be room for some flexibility at the lower end of the number so long as the consequence of adopting a P value higher than 0.05 is understood i.e. the higher the number, the less the weight to be given to any statistical correlation.

Finally, Bayes' theorem should briefly be mentioned. This is a theory reflected in a mathematical formula which takes the prior probability of an hypothesis (based upon certain evidence) and then seeks to adjust that probability based upon the addition of new evidence. That is, does the new evidence make the hypothesis more or less probable than the hypothesis without the evidence? The formula (for present purposes) is simply expressed as:

$$P(h/e) = \frac{P(h) \times P(e/h)}{P(e)}$$

P(h/e) is the probability of the hypothesis "h" in the light of new evidence "e". It is the conclusion being sought. P(h) is the prior probability of the hypothesis (i.e. absent reliance on the new evidence). P(e) is the prior probability of the new evidence (absent any assumption about the truth of the hypothesis). P(e/h) is the probability of the new evidence on the assumption that the hypothesis is true.[77] Now such a theorem has obvious forensic problems. First, you need estimates of all prior probabilities. These may not be available or even calculable. Second, in any event, any calculation of prior probability estimates is substantially subjective and usually highly contentious. There are usually little rigorous objective standards

---

[76] See Foster and Huber (1997) (supra) at p79.

[77] The intuition is that the extent to which new evidence adds support to an hypothesis is proportional to the degree to which the hypothesis predicts or entails the new evidence. Hence the position of P(e/h) in the numerator. P must be equal to or less than one. If P(e/h) = 1, that indicates that if the hypothesis is true then the new evidence necessarily follows. The position of P(e) in the denominator can also be shown intuitively. As said, P(e) measures how likely the new evidence is considered to be when the truth of the hypothesis is not assumed. If the new evidence was certain to occur regardless of the truth of the hypothesis, then intuitively the new evidence should add little if anything to the hypothesis. But if the new evidence is very unlikely to occur, but it does occur, that should add support to the hypothesis. So if P(e) is in the denominator and substantially less than one, then mathematically P(h/e) is increased, the result you would expect intuitively. In summary then, the scaling factor P(e/h)/P(e) demonstrates that P(h/e) is proportional to P(e/h) and inversely proportional to P(e), consistent with intuition. See, generally, Chalmers, A.F. (1976): "*What is this Thing Called Science*": Chapter 12; University of Queensland Press.

to assess them. Third, the compartmentalized assessments of each of the relevant probabilities (prior or posterior) is necessarily artificial and their utility questionable.

VII The theory laden quality of observation

Scientific observation is often portrayed as an objective tool separate from theory. But it is very much theory-laden. And any scientific expert who gives evidence of observations or their interpretation needs to be closely tested on any underpinning theoretical assumptions. Scientific observation is usually performed with instruments. But for a reliable observation one needs:

(a) a theory of what is being observed;[78]

(b) a theory as to its means of measurement[79] - for example, it may not be directly measurable but only indirectly inferred and measured from other more direct data (such as traces left in a cloud chamber to infer the existence and movement of a particle);

(c) a theory as to the design and construction of an instrument which accords with theories (a) and (b);

(d) the reliable operation of the instrument to produce the derived measurement;

(e) a theory as to the interpretation of the results of measurement.

But even visual observations performed by the unaided eye can depend upon theory or the perceptual experience and knowledge of the observer. Polanyi[80] gives the example of a medical student who is taught to make a diagnosis based upon an x-ray picture. Initially the student may perceive shadowy blotches and possibly shadows of the heart and ribs. After several lectures the student may begin to see "through" the ribs to "observe" the lungs. Eventually, after more lectures the student may begin to see a rich panorama of detailed physiological variations in the lungs.[81] Another example is provided by the Longford Royal Commission and subsequent litigation where particular microscopy on a fracture point in the weld of a heat exchanger[82] was open to conflicting and theory-laden interpretations as to whether the "observations" displayed brittle fracture or ductile failure in the weld. Indeed the interpretation of these pictures very much displayed more of an art than a rigorous scientific method. Similar issues arise in the biological sciences in identifying cells, mutations, bacteria, viruses, etc. The short point is that it is not justifiable, in generality, to give scientific observation some status independent of theory. The former is built upon and interpreted by the latter, particularly where mediating observational or measuring instruments are involved. And both the quality of the observation

[78] I have not used Bas van Fraassen's definition of observables which is confined to objects that are directly observable by the unaided human senses (van Fraassen, B. (2000) – *Constructive Empiricism Now* (supra)).
[79] For example the theory of optics underpinning microscopes and telescopes.
[80] Polanyi, M. (1973) – *Personal Knowledge* p101. Routledge and Kegan Paul.
[81] See Chalmers, A.F. (1976) (supra) at Chapters 1 and 2.
[82] This exchanger in a lean oil absorption gas processing plant in Longford, Victoria failed, with consequential vapour release and then ignition.

and its theoretical underpinnings need to be subjected to scrutiny. And problems are even more acute when the interpretation of observation involves subjective considerations injected because of:

(a) the individual expertise and experience of the witness, including the level and diversity of experience of making and interpreting relevant observations;

(b) the subconscious biases of the witness in interpreting such observations, including the priority given by the witness to qualitative tools over quantitative forensic techniques.

In summary, interpretation of observation may become as much an "art" as a "science".

VIII How to Assess the Reliability of a Scientific Theory or its Application

Let me set out some check-lists that may be used to assess a theory or its application. Suitable questions to pursue include the following:

(a) Has the theory demonstrated explanatory success? Has the theory demonstrated novel predictive success? More specifically, is the theory testable empirically? Is there any experimental verification of the theory? If so, is it adequate?[83]

(b) Has the theory been peer reviewed? What is the breadth and expertise of those doing the peer review? What are the results of such peer review? Generally, the greater the scrutiny, the greater the likelihood of detection of substantial flaws.

(c) Has the theory been published? If so, where and how broadly? In peer reviewed journals for example? Now some theories may be too new or of too limited interest to be published. If so, other means are going to have to be used to determine the degree of acceptance within the or a scientific community which is relevant to weight. As said in *Daubert* at [32], and equally applicable to Section 79 of the *Evidence* Act 1995 (Cth), publication is not a sine qua non of admissibility.[84] But it may be relevant to weight.

(d) What have been the scientific community's responses to the theory or its publication? What are the criticisms or its weaknesses? One question to ask an expert who uses a particular theory is whether there are any known or published deficiencies or criticisms and if so, how they have been addressed. And for a new theory, the question may be whether there has been a sufficient opportunity or time to challenge the new theory. If not, the question may be what potential weaknesses or criticisms might suggest themselves as worthy of further investigation. More generally for any theory, is there a large number of treatises which reject the theory? Do any professional organizations reject the theory? If so, why? Are the opponents of the theory better credentialled?

(e) If the theory is new and purports to "replace" an existing theory, justification for the replacement may be necessary. What were the flaws in the existing theory? How are

[83] Imwinkelried, E.J. (2004) (supra) at Chapters 4 and 10.

[84] More generally, but in a separate context, the more widely an expert witness has published in the relevant or associated field, the more credible the expert witness (Kousser (1984) (supra) at p16).

they solved by the replacement theory? Does the replacement theory create new problems not raised by the old theory? If so, what is their significance?

(f) Moreover, for a particular expert, a relevant question may be how frequently the expert has used the theory. Infrequent use may suggest a doubt as to the breadth of the experience of the witness, or that the witness has doubts as to the theory, or that there is a more preferable theory, or that it is simply a very new and contestable theory. Moreover, a review of the publications of the witness may be illuminating on such matters. Further, although a witness may know of the theory, yet he may be relatively unfamiliar with the relevant instrument or technique to apply it. This may need to be investigated if the witness is also giving evidence as to the latter.

(g) Further, are there any logical flaws? Now it is unlikely that any theory put forward will contain such flaws. But flaws may exist in their application or other aspects. Freckelton and Selby ([7.10.210]-[7.10.230]) discuss logical flaws flowing from the following:

(i) First, *post hoc ergo propter hoc* or arguing that because B follows A that there is a causal link.

(ii) Second, affirming the consequent, arguing for example: "If the heart is weak then the pulse is feeble. The pulse is feeble. Therefore the heart is weak". Other causes are not excluded.

(iii) Third, denying the antecedent, by arguing for example: "If schizophrenia is hereditary, lineal descendants will inherit it. Schizophrenia is not hereditary. Therefore, lineal descendants will not develop it". But schizophrenia may have other causes. Denying the antecedent does not deny the consequent .

(iv) Fourth, the naïve induction, where too few examples are used to support an induction or the induction is expressed in absolute terms rather than more appropriate probabilistic terms.

This is not an exhaustive list. There may be other fallacies such as: A. "the ecological fallacy" – transferring relationships that occur in populations to individuals; B. "the bundled fallacy" – that multiple pieces of evidence, each independently being suspect or weak, provide strong evidence when bundled together; C. "the risk fallacy" – relative risk, measuring association between a risk marker and disease, has little bearing on the absolute risk or probability that an individual will acquire that disease.[85]

(h) Further, for a particular technique underpinned by a theory, other relevant issues may relate to:

(i) whether there has been experimental verification of the instrument or technique;

(ii) the known or potential rate of error (*Daubert* at [33]);

[85] See Skrabanek, P. and McCormick, J. (1990): "*Follies and Fallacies in Medicine*": Prometheus, cited in Foster and Huber (1997) (supra) p142.

(iii) whether there are generally accepted standards for the application of the technique and whether such standards were set by an appropriate industry or other body (regulatory);

(iv) whether such standards allow for a certain rate of error and the "justification" for that rate;

(v) whether the standards were applied in the particular case in terms of the implementation of the technique.

Any scientific technique is necessarily imperfect and it is important to be aware of the relevant qualitative or quantitative appraisal of such imperfections made by the relevant scientific community.

In terms of the carrying out of tests, issues can arise, as enumerated by Imwinkelried,[86] as to whether:

(a) before the test, the scientist took proper preparatory action and whether the scientist adequately prepared:

- (i) the person to be tested;
- (ii) the sample to be tested;
- (iii) the instrument to be used in the test; or
- (iv) other materials used in the test;

(b) during the test, there was any substitution for or contamination of the sample, instrument, or other test material;

(c) the scientist used a proper environment for the test;

(d) the scientist used the appropriate combination of test procedures including whether the scientist carried out:

- (i) confirmatory tests (i.e. two or more tests to confirm the result);
- (ii) positive control tests (i.e. if testing for a drug in an unknown sample, you would test a known sample of the drug for a positive result to validate the test, i.e. designed to exclude false negatives); and
- (iii) negative control tests (i.e. adopted a test designed to eliminate other competing causes for the positive result – i.e. designed to exclude false positives);

(e) the test design was defective because the circumstances at the time of the test differed substantially from those at the time of the relevant event due to:

- (i) physical differences; or
- (ii) differences in the state of mind of the test participants;

(f) the test design was defective because the test situation omitted a factor necessary for a reliable test;

---

[86] Imwinkelried, E.J. (2004) (supra) at Chapter 12. I put to one side any statutory or other presumptions that may apply as to the functioning or accuracy of measuring instruments.

(g) the scientist reached a premature opinion because he:

  (i) conducted the test hurriedly; or

  (ii) reached a conclusion before examining most of the data;

(h) the scientist reached a conclusion or diagnosis although criteria needed for that diagnosis were missing;

(i) the scientist overlooked data inconsistent with his opinion.

Now it may be that the trier of fact is not in a position to pursue any one or more of these questions. But if they have not been pursued by counsel or are not disclosed in the evidence, the trier of fact may be able to take into account any relevant omission in assessing the weight to be given to the theory or technique and its application in the particular case.

C Conclusion

I have endeavoured to set out various themes that should be borne in mind when assessing any form of scientific evidence. The epistemic lens suggested is unremarkable. There are no novel elements. Its themes resonate with the scientific method. Its application in the specific case should assist in filtering out what the scientists themselves would in any event ultimately discard. No doubt the time horizon for judicial assessment of a theory or its application is narrower than that of the relevant scientific community. No doubt judicial decision making necessarily has to give a definitive ruling on a theory or its application at a particular point (even where the theory or its application is in a state of uncertainty), whereas the scientific community can tolerate shades of grey for a longer period. No doubt judicial decision making is individualized, whereas the judgment of the scientific community is in a broader setting, including its incremental value to a broader set of theories or knowledge. No doubt the sources of scientific knowledge that the judge can consider are limited by the time and adversarial context, whereas the sources of knowledge for the scientific community upon which a theory or its application can be assessed is ever expanding and has no similar artificial constraints. But all of this is just to point out that the judicial epistemic lens has a narrow time frame, a narrower context and a different focus. But even accounting for these differences does not obviate the need to include in that lens the epistemic themes that I have discussed.

Questionable science has not only been admitted but been accepted in many different legal contexts, for example:

(a) breast implant litigation;[87]

(b) criminal proceedings where fingerprint evidence was adduced;[88]

(c) the Bendectin litigation.[89]

---

[87] Angell, M.: "*Science on Trial – The Clash of Medical Evidence and the Law in the Breast Implant Case*", Chapters 5 and 6: W.W. Norton & Co; New York, 1997.

[88] Cole, S.A.: "*Suspect Identities – A History of Fingerprinting and Criminal Identification*", Chapter 11: Harvard University Press, 2001.

Application of the epistemic themes suggested above may reduce the risk of such acceptance. At the least, a circumspect epistemic lens can better contextualize the derivation of a particular scientific theory or its application and the weight to be given to it.

JBRB

[89] Edmond, G. and Mercer, D. (2000) (supra).