The Paradox Of Proof And Scientific Expertise

Carlo Martini [†]
uni.c.martini@gmail.com

ABSTRACT

In this paper I criticize the current standards for the acceptability of expert testimony in current US legislation. The standards have been the subject of much academic literature after the *Frye* and *Daubert* cases. I expose what I call the Paradox of Proof, and argue that the historical and current standards have sidestepped the problem of determining who is an expert and who is not in a court of law. I then investigate the problem of recognizing expertise from the layperson's standpoint, and suggest what courses of action the future research ought to take on the problem of identifying expertise.

Keywords: expertise, experts, laypeople, proxies, legal proof, demarcation, Harry Collins, Robert Evans.

Introduction

Living in societies, we rely constantly on the work of others for our needs: we rely on the baker to provide us bread in exchange for money, or on the construction worker to build our house. But reliance on others is not only material, it is also epistemic. Contemporary philosophers and social epistemologists, inspired by what Adam Smith called "division of labor" – he probably had in mind mostly material labor (see Smith, 1976) – have investigated the division of "epistemic labor" among the members of our epistemic communities. Kitcher (1990), focusing on scientific communities, has called it "division of cognitive labor"; it takes place, for instance, when modelers rely on experimenters to parameterize their models. But division of cognitive or epistemic labor can be found among the members of all epistemic communities, not only among scientists: We rely on doctors to know what our

[†] Academy of Finland Centre of Excellence in the Philosophy of the Social Sciences. Department of Political and Economic Studies, University of Helsinki, Finland.

symptoms indicate, and we rely on bankers, sometimes incorrectly, to know what retirement or mortgage plan suits our personal and professional needs. One can safely assume that the division of epistemic labor is what allows us to get around our daily needs with the minimum amount of knowledge, or epistemic effort, required to survive and perform our daily tasks. We don't need to know what kinds of food are poisonous, or how to build a safe water system or a shelter, because others have learned how to do all those things for us.

The focus of this paper is the legal sector, where epistemic reliance on others is ubiquitous. In particular, I will be mostly concerned with how judges rely on experts to disentangle technical and scientific issues in cases where common knowledge of facts is not sufficient for a verdict. Such cases are more and more common in liability litigations (e.g., pharmaceutical or medical cases) and criminal cases where scientific techniques are employed (e.g., DNA testing, psychological evaluation). A much-studied 1993 U.S. Supreme Court case, "Daubert v. Merrell Dow Pharmaceuticals", determined that in order to admit an expert's testimony to a trial the court must "look not to an expert's conclusions, but to his "methodology", to determine whether proffered evidence is really "scientific ... knowledge" and hence reliable." (Haack 2005, S66). In practice, the ruling asked the court to determine whether the expert's testimony could be admitted on grounds that it was "scientific", that is, that it derived from the correct application of the scientific method. Decisions on whether to admit an expert's testimony or throw it out are calls that courts and judges must often make, even when they do not have sufficient knowledge to judge on the complicated issues for which reliance on experts was required. Courts and judges are, in this sense, laypeople, in relation to the experts whose testimonies they have to evaluate.

The problem just illustrated is the problem of how a layperson – that is, someone who has no expertise in a certain field – can adjudicate who is and who is not an expert in that field. In many instances in which we rely on others we are in an easier position to assess whether the people, or groups, we rely on have the necessary expertise for our needs. For example, as long as our standards of consumption are similar, we might be able to recognize a good carpenter by the quality of their crafts, or a good baker by the tastiness of their bread (see Collins and Evans 2007, 57-60). But there are harder cases, like legal litigations, which bring to the fore an apparent Paradox of Proof: How can someone who doesn't know the subject matter know who is an expert on that subject matter? As a simple example we could ask "how can someone who does

not have specific knowledge of mathematics know which mathematician has the proof to a certain theorem?" The Paradox of Proof exists in the mathematical case because there do not seem to be external standards, by which we can evaluate the goodness of a proof, that are independent of the mathematical standards that make a proof a good one. There are many other cases like this one in science, and they have consequences for legal adjudication.

In the legal scenario, for instance, how can a judge, who does not have domain knowledge of the complexities of DNA testing, know which experts' opinions are more qualified to weigh on the verdict? The problem is a complicated one from the point of view of legal theory (see Haack 2002), but for the purposes of this paper I will focus on the narrower philosophical issue of recognizing experts while standing in the shoes of the layperson. As it should become clear in the following sections, judges are equivalent to laypeople when it comes to evaluating expertise.

In the next section I will look more in detail into two landmark answers that the U.S. legal system has given to the problem of adjudicating expertise. I will argue that those answers are sidestepping the very problem. Next, I will introduce and explore one prominent stance on how to recognize expertise, suggested by Collins and Evans (2007). I will highlight some of the limitations in Collins and Evans's proposal. I will then consider a different proposal: Shanteau's contribution (1992) to the problem of expertise. I will argue that recognizing expertise should be done through *proxies* and *indicators*, which I define in the last section; but the task of finding such proxies and indicators is not a simple one, or one that can be done with theory alone. I will suggest a number of sub-problems that research on expertise will have to undertake to try to solve both the paradox of proof and the current lack of criteria for evaluating expertise in legal cases.

1. *Frye* and *Daubert* on expertise

Why is it important to find criteria of expertise, even if all we can formulate is only a tentative and probably imperfect list of them? The two major decisions on what counts as expertise in court litigations should be enough proof that criteria for identifying expertise are badly needed: these landmark court decisions about the involvement of experts testimony as evidence in court were the Frye test (see Frye v. United States 1923) and the ruling over Daubert vs.

Merrell Dow Pharmaceuticals, Inc. (Daubert v. Merrell Dow Pharmaceuticals 1993). In this section I will present the two cases, and argue that in both of them the decisions on the admissibility of expert evidence were an attempt to evade the question "who is an expert?"

In Frye v. Unites States (1923) the court decided to reject "the results of a then-new blood-pressure deception test on grounds that novel scientific testimony "crosses the line between the experimental and the demonstrable," and so is admissible, only if it is "sufficiently established to have gained general acceptance in the particular field to which it belongs."" (Haack 2005, S66). The Frye test established that the criterion for the admissibility of expert testimony is the "general acceptance" of the testimony – i.e., the science invoked by the testimony – in the relevant scientific community. At the time, since the science of lie-detector tests had not reached a consensus, a court could not accept the test.

The standard of general acceptance, applied by the court in Frye v. Unites States, seems to be founded on the idea that good science tends to generate consensus around established facts and method; so the consensus criterion is, prima facie, a reasonable standard for the admissibility of scientific evidence. Much of the epistemic work in science is done by consensual processes: scientists formulate hypotheses, they gather evidence, they present their evidence to their peers through conferences and journals, their peers evaluate the evidence, respond, criticize, reject what is inadmissible, and, slowly, a consensus may form as to what can be accepted on a more or less definitive basis, or at least until new evidence is brought to the fore. This is no doubt a rather idealized characterization of how science works; in practice, the process of accepting scientific facts and theories is much less linear. But we can still claim that under normal and slightly idealized circumstances science progresses by consensus (see also Kuhn, 1970, on the role that consensus has in the achievement of "normality" in science).

The problem with the use of consensus criteria in law is that too often consensus is only a byproduct, not a cause of good, and therefore, court-admissible, science. Scientific consensus forms because the science in question is grounded on good evidence, but sometimes it forms around bogus science as well, possibly caused by extra-scientific reasons like biases and political and economic interests. Therefore, we can accept the thesis that consensus is necessary for science without being committed to the thesis that it is also sufficient, and if that is the case a court should not accept scientific

claims based on the presence of consensus (i.e., the general acceptance criterion). Even more importantly though, there are additional reasons for rejecting the general acceptance criterion established in *Frye v. United States*, and they rest on the fact that consensus, as a byproduct of good science, typically forms too slowly for the need of courts to ascertain the truth – or, at least, the "provable" – in legal trials.

The Frye test and the general acceptance criterion remained the standard for expert testimony in United States courts for several decades until *Daubert v. Merrell Dow Pharmaceuticals*, in 1993, set a new standard for the admissibility of expert testimony. What happened in the meantime was that, in 1975, Congress had adopted the Federal Rules of Evidence, on the basis of which the consensus standard could no longer be upheld as the only standard for the admissibility of expert testimony. In light of *Daubert*, the Federal Rules of Evidence were further revised, and rule 702 now states:

> If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case. (Saks and Faigan 2005, 109)

Clauses 2 and 3 of the rule are important here, as they call upon scientific methodology. Judges are called to judge on the reliability of principles and method of the testimony given, and on the correct application of principles and methods to the facts under investigation. The concept of reliability in science is a technical one, and refers to a method's ability to give consistent results (see Buekens and Truyen, 2014); it does not apply to the ability of a method to give true results. The ability of a method to give true results is called "accuracy" and it is a harder requirement to meet for a scientific method. Reliability, however, provides enough evidence that the method is at least not random – i.e., that it provides results based on underlying facts – and this is an important requirement of any objective method, and of scientific methods in particular. The applicability of a given method to a particular case is also important in science, since a method demonstrated to be reliable on a certain domain may not be reliable under different circumstances (e.g., a method for DNA testing applied on samples that have not been properly handled or collected). In standard procedures like DNA testing, the rules of applicability may be

relatively straightforward, but, in general, in science such rules are not easy to formulate, especially in novel science. Yet despite the difficulties of establishing reliability and applicability, rule 702, above, requires the triers of fact to do just that.

Both the Frye test and the new rule 702 in the Federal Rules of Evidence are in fact ways to avoid the question "who is a legitimate expert in a court of law?" The Frye test and rule 702 do not try to establish who has legitimate expertise, and can thus provide reliable testimonial evidence of technical or scientific facts that neither the court nor the general public could assess. Instead, they ask the trier to evaluate the evidence that is brought forth by the experts. That does not seem to make much sense though, since the need to call upon experts to provide their judgments is exactly the inability of courts and judges to evaluate and weigh complex technical and scientific issues that bear on the matter under trial. Both the Frye test and rule 702, then, shift the problem from evaluating the validity of expertise to evaluating the validity of the evidence presented. Of course, in principle, this more is sensible: We would rather accept valid arguments and good evidence than just trust that our experts are giving us valid arguments and good evidence. But the move misses the point: It is because we are not in a position to assess the evidence directly that we resort to rely on expert testimony.

We may imagine a counterfactual scenario in which we were asked to judge whom we trust the most to be able to heal us from an ailment, and whose opinion we would rather not listen to. Among Western-educated people, it may be safe to assume that we trust medical doctors, rather than karma healers; but, if pressed, would we claim that we trust a doctor's judgment because it is agreed upon by most of the medical community? Against this, one must note that many of the cures that doctors provide are far from being accepted as the medical consensus. More importantly, whether there is or is not such consensus is hard to adjudicate from the standpoint of the layperson. Would we claim, instead, that we trust a doctor's judgment because we have assessed their method as reliable, as well as the applicability of their method to the specific case at hand? This answer is not a sensible answer either, because people untrained in science and its method cannot easily make judgments about reliability and applicability.

One could still claim that judges stand to experts in a different relation than patients stand to doctors. Perhaps judges are more capable of recognizing the trappings of expertise by looking at the science itself. This hypothesis is also

untenable. It is hard to see how judges could have the kind of training required to assess scientific evidence and method, especially since even scientists themselves often do not have the capacity to assess evidence and method outside of their field of specialization, and judges are often required to assess the acceptability of expertise in many different fields, each with different domain-knowledge and methods. If that was not enough prior evidence, Gatowski et al. ran an empirical survey to ascertain whether the judges could be relied upon to have enough scientific expertise to apply rule 702 of the Federal Rules of Evidence, and concluded that that was not the case: "The survey findings strongly suggest that judges have difficulty operationalizing the *Daubert* criteria and applying them, especially with respect to falsifiability and error rate." (2001, 452)

Good alternatives to either the *Frye* criterion or the current rule 702 of the Federal Rules of Evidence are needed, if we trust the findings from Gatowski et al. (2001) that "judges overwhelmingly support the "gatekeeping" role as defined by *Daubert*, irrespective of the admissibility standard followed in their state. However, many of the judges surveyed lacked the scientific literacy seemingly necessitated by Daubert." (2001, 433). An alternative, however, should deal directly with the question of who is a legitimate expert in a court of law on a given subject matter. Any attempt to ask the judge or jurors to evaluate the evidence the experts bring forth, their methods, and similar aspects of an expert's testimony will fall into the trap of requiring the kind of knowledge that jurors and judges do not possess, which was the reason why experts were consulted in the first place. Of course, a judge or a court might still check for consistency of an expert's method; whether the expert is giving contradictory statements, and other basic checks that a layperson would be able to perform on an expert's testimony. But much more than logical consistency and similar requirements is needed.

The next sections will focus on how we can define expertise, and what alternative criteria for the admissibility of an expert's testimony we can hope to develop. Any such criterion will have to make it possible for a layperson to recognize an expert, keeping in mind that the paradox of proof casts doubt over the entire enterprise: "How can someone, ignorant in a certain domain, know who is an expert in that domain?"

2. Collins and Evans on Expertise and Experience

Collins and Evans (2007) have written extensively on the problem of "ways to separate those who fall into the envelope of potential judges in respect of various expertises from those who fall outside that envelope" (2007, 67) They recognize that in most cases where we lack domain-specific knowledge needed to assess expertise – that is, whenever we are not experts ourselves, trying to identify other experts – we rely on "externally measurable criteria" (2007, 67). I call these criteria "proxies" of expertise: i.e., factors that indicate the presence of substantial expertise, when such presence cannot be detected directly.

Collins and Evans review a number of proxies of expertise, the first one being credentials and the second one being track-record. According to the former criterion, we can allegedly identify expertise by means of "certificates attesting to past achievements of proficiency", while according to the latter we would identify expertise by looking at one's past success in solving problems related to the relevant field of expertise. They dismiss both criteria on the grounds that in both cases there can be significant expertise even in the absence of credentials or a track-record. On the one hand, accreditation is a social practice, but expertise is substantial, and there can then be expertise that goes unaccredited. On the other hand, a track-record is not always available, if not in principle, at least in practice because, like accreditation, it is a social practice, whereas expertise is substantial and personal.

After correctly discarding credentials and track-record, Collins and Evans turn to experience – i.e., experience within the relevant domain – as the preferred criterion for expertise: "We know from the outset that without experience within a technical domain, or experience at judging the products of a technical domain, there is no specialist expertise. Without experience of doing science, talking to scientists, playing or listening to violin-playing, or looking at and discussing bathroom tiling, the minimal standards for making judgments in these areas have not been met." (2007, 68) The concept of experience is helpful for understanding expertise, but it is not fully analyzed in Collins and Evans's book. To be fair, they do talk extensively about experience throughout the book; for example, they describe experience as embeddedness in the relevant epistemic community. But the discourse always falls short of a detailed analysis of the concept.

Perhaps the assumption is that we have a very good common-sense understanding of experience, and that it seems also obvious to think that

experience must be a hallmark of expertise. Given this common-sense understanding of experience, it may then seem unnecessary and pedantic to provide a philosophical analysis of the concept of experience before we dare use it. However, in the following I hope to show that unpacking the concept of experience will reveal problems that ought not be left unanswered and that relate to expertise and a layperson's ability to identify experts.

I will argue we cannot use experience as a proxy for expertise – that is, as a trait we look for, under the assumption that it is correlated with expertise. Instead, I will argue that experience is part of the substance of expertise. This will open two problems: 1) To identify experience one needs to be able to tell apart relevant from irrelevant experience and, in turn, this implies that only experts, as possessors of experience, can identify other experts; 2) Experience maybe a necessary but not sufficient condition for expertise and, in turn, even if the former problem could be resolved, we would still have no clear indication on how to detect expertise.

The first thesis in this section is that experience seems to be a substantial trait of expertise, not a proxy. To illustrate, let us imagine a scientist, working for several decades in a narrow and highly specialized field, and accomplishing great and substantial success in that field. With that in mind, we would certainly be confident in the fact that the scientist is a true expert in her field. That is because the experience of the scientist is relevant to its genuine expertise. But we cannot know, a priori, that the expertise is genuine, we infer it from the fact that we observe the imagined scientist's experience. But a scientist, through accomplishments achieved in her own field, may try to act as an expert in a much broader field, and on topics that are outside her own narrow field of specialization.

When that happens – when scientists speak outside their own field of genuine expertise – laypeople do not have a way to recognize which experience warrants which expertise; or, in other words, when a scientists speaks as an expert on a given matter, the public does not typically have the means to recognize, on the basis of an observation of that scientist's experience, whether the matter the scientist talks about is within her domain of genuine expertise, or whether she has overstepped the boundaries of that domain. This is to say that we typically do not have a direct way to assess the relevance or irrelevance of experience to one's putative expertise.

The scientist that was described in the preceding paragraphs is not only imaginary: Kitcher recounts the story of what Oreskes (2010) calls "merchants

of doubt"; namely, scientists who, while gaining the status of expertise in their own specialized field, where they had considerable experience, spoke publicly about climate change issues at large, despite their experience not being relevant to qualify them as experts on climate science.

> [...] a few scientists, with strong ties to particular industries and with conservative political connections, have played a disproportionate role in debates about controversial questions, influencing policy-makers and the general public alike. *Typically, these scientists have obtained their stature in fields other than those most pertinent to the debated question.* Yet they have been able to cast enough doubt on the consensus views arrived at by scientists within the relevant disciplines to delay, often for a substantial period, widespread public acceptance of consequential hypotheses.(Kitcher 2010, 3, my italics).

It seems evident that in order to be able to tell relevant from irrelevant expertise one needs to be at least an "interactional expert" on the field in question. Interactional expertise, in the terminology used by Collins and Evans, is "expertise in the language of a specialism in the absence of expertise in its practice." (2007, 28). Being able to discriminate experts from non-experts via the criterion of experience requires interactional expertise because one needs to be able to separate relevant from irrelevant experience. There may be cases where experience can easily be ruled out as irrelevant: For example, most laypeople would be able to say that experience in a bakery will never make one an expert in piloting airplanes. In general, however, laypeople would not be able to tell relevant from irrelevant experience in specialized sectors because of their lack of domain knowledge.

However, for the sake of the argument, let us imagine for a moment that we were able to bypass the problems just mentioned. We could imagine a nearly perfect organization of science (and its technological applications) where it is clear which kinds of experience are relevant for which domains of expertise. The problem that remains is whether experience is correlated with expertise; we may safely assume that it is at least a necessary condition, but is it also sufficient?

This is not just an appeal to the abstract philosophical requirement of providing "necessary and sufficient conditions for the occurrence of X". If Collins and Evans's concept of experience is indeed meant to stand as a proxy of expertise, then their account seems to overlook the fact that one may have considerable experience without having, in relative terms at least, much expertise. To be fair, the authors do acknowledge that experience may not be

sufficient for expertise, when they note that Harry Collins never acquired interactional expertise in the field of amorphous semiconductors, in spite of the several interviews with scientists on the physics of amorphous semiconductors (p. 33). It is possible then that Collins and Evans never intended experience to stand as a proxy for expertise, despite the fact that they list it after credentials and track-record. But the thesis defended here is simply that there can be experience without there being expertise, and because that does not hinge on the interpretation of Collins and Evans (2007) – who indeed make the same point – it should not affect the following considerations.

Shanteau et al. (2002) report on the relation between experience and expertise: If by experience we mean something that can be measurable (for example number of years on the field), then one can provide evidence of the fact that "there are many examples of professionals with considerable experience who never become experts. Such individuals may even work with top experts, but they seldom rise to the performance levels required for true expertise." (2002, 254) Shanteau and his coauthors conclude that while we should expect instances where more experience correlates positively with expertise, we cannot generalize for all instances.

To conclude, in this section I have argued that experience is not a proxy of expertise, regardless whether that may or may not have been the suggestion in Collins and Evans (2007). Collins and Evans have provided much valuable work on the concepts of expertise and experience. But for the purposes of this paper, the problem of identifying experts still seems elusive: We would like to be able to tell experts from non-experts, and we would like to be able to do so, even at the cost of some imprecision, while standing in the shoes of a layperson. For that, the criteria we have seen so far (credentials, track-record, and experience) do not seem to work.

The trappings of expertise

There are at least two substantial components of expertise: experience and competence. They are the past and future components of the concept of expertise; or equivalently, the backward- and forward-looking components. Experience is the amount of practice experts have put into solving problems relevant the their field. Experience must be relevant; that is, it must be focused on a specific domain of knowledge or practice. That experience be relevant is a requirement because there

can be putative experts with considerable experience who are not truly competent experts, though it seems unlikely that an expert with considerable relevant experience would lack the competence necessary to solve future problems. Competence, here, means the ability to solve problems in the relevant field of expertise. It seems obvious to think that a person's ability to solve problems depends, albeit not exclusively, on the amount of past attempts at, and successes in, solving problems. Solving problems can of course be real or virtual; much textbook training into a profession involves "solving" virtual problems, or at the very least the provision of tools for solving problems.

To repeat, while experience is the backward-looking component of expertise, competence is the forward-looking one. Both, however, are substantial traits of expertise, and it was shown in the previous section that we cannot take layman-perceived experience directly as a proxy for relevant experience and, thus, for expertise. Moreover, the two components are related: In human experts, it is unlikely to have the former without the latter. The fact that experience and competence are substantial traits of expertise is both accurately descriptive – we certainly observe considerable relevant experience and competence in experts – and normatively compelling, because we want experts to be able to solve problems in their field of expertise, and that is more likely to happen if said experts have solved problems in their field of expertise in the past.

Experience[1] and competence, however, as substantial traits, cannot be detected directly; the former for reasons explained in the previous section (a track-record, for instance, is a good proxy, but it can fail to detect experience if the field of experience differs from the field the track-record is taken from), and the latter for the obvious reason that competence is a forward-looking property: We cannot detect competence until it has been applied, namely, until it has become a past event. Of course there are proxies for competence, as well as for experience, but the question is how to identify, measure, and weigh the contributions of these proxies to true expertise.

The psychological literature has analyzed proxies for expertise at great length, both theoretically and experimentally. Shanteau lists a number of those proxies and evaluates them normatively: (1) experience; (2) certification; (3) social acclamation; (4) consistency within reliability; (5) consensus; (6) discrimination ability; (7) behavioral characteristics; (8) knowledge tests (see Shanteau et al. 2002). It is relevant to note what the behavioral and psychological characteristics

[1] In the following, I will use 'experience' to mean relevant experience, for brevity, except where otherwise stated.

that can be attributed to experts are: (a) possession of content knowledge; (b) perceptual/attention abilities; (c) ability to discriminate relevant from irrelevant information; (d) ability to simplify complex problems; (e) ability to communicate their expertise; (f) ability to handle adversity and difficult situations; (g) ability to follow established practices and make exceptions when appropriate; (h) self-confidence; (i) ability to adapt; (j) sense of responsibility (see Shanteau 1992). In the end, Shanteau et al. (2002) propose a ratio between discrimination and inconsistency and test it against available data. They conclude that the ratio (named *CWS*), if successful, "may provide an answer to the long-standing question of how to identify expertise in the absence of external criteria." The difference between a proxy and the solution Shanteau et al. propose is that the CWS ratio is a function of several proxies – that is, a derived measure – whereas each proxy is a direct measure: i.e., it could be measured directly with empirical data, for example, experience can be measured in years, and specific abilities can be tested in experimental conditions. This is an important distinction: On the one hand, direct measures are proxies, in the sense that they are candidates for something that we can detect directly instead of expertise, which cannot be detected directly. On the other hand, indicators are mere numerical values; i.e., combinations of proxies, which, ideally, point us in the direction of expertise. The better the indicator is, the more likely we will find genuine expertise there.

The approach by Shanteau and his collaborators is reductionist – that is, it reduces expertise to identifiable traits that we take to be proxies for expertise – and it seems to be a valuable step in the correct way to pose the problem of expertise. In other words, while the Federal Rules of Evidence mentioned in the previous sections seemed to do away with the problem by asking the judges not to evaluate the experts, but the content and form of what they say (i.e., method, relevance, etc.), what one should do, instead, is to find ways to evaluate the experts' expertise itself by means of proxies and indicators.

But is the ratio suggested in Shanteau et al. (2002) the ultimate word on what proxies we can use to identify expertise? Probably not. More empirical research is needed, as well as more development of theory, for better understanding expertise along the lines opened by Collins and Evans (2002, 2007) and Collins (2013). In concluding this section I make a number of observations on the issues that future research on expertise will have to tackle in order to make progress on the problem of expertise; that is, the problem of identifying experts while standing in the shoes of laypeople, both in the legal sector and in other sectors where expertise is needed.

The first observation concerns the analogy between proxies/indicators of expertise and evidence in science. There can clearly be many proxies of expertise, and even more indicators, given our definition of indicators as functions that combine more than one proxy. When evaluating which combination of proxies will likely give a good indicator, we should take into account the literature on the method of combining (i.e., amalgamating) scientific evidence. In other words, considering different proxies as bundled indicators of expertise is equivalent to combining evidence in science. If we are allowed to combine different sources of evidence in support of a thesis, there are methodological rules that we should follow: For instance, two sources of evidence, *a* and *b*, may not support a thesis more than another source, *c*, if there is a correlation between *a* and *b*. More concretely, if expert *X* has, say, both *accreditation* and *acclamation*, that may not give us evidence that *X* is more expert than *Y*, where *Y* only has *accreditation*, if *acclamation* and *accreditation* are correlated so that *acclamation* implies *accreditation*. Possible dependencies between proxies should be taken into account when developing indicators.

The vast literature on amalgamating evidence, then, ought to be considered when we look for evidence of expertise: Stegenga (2013) and Lehtinen (2013) present different results on the possibility (or impossibility) of amalgamating evidence coming from diverse sources. Bovens and Hartmann (2002) and Claveau (2013) also offer different conclusions in discussing the "variety of evidence thesis": i.e., the thesis according to which the warrant given to a hypothesis increases, ceteris paribus, when the body of evidence is more varied. Both the problem of amalgamating evidence and the problem of the variety of evidence need to be considered in the study and development of indicators of expertise. The theoretical results by Stegenga, Lehtinen, and the other authors mentioned in this section, could inform the empirical application of indicators to real data, which Shanteau et al. (2002) conduct in their paper.

A further issue with the study of expertise is that the definition of expertise is likely to be a moving target: "being an expert on X might change with the X". That is because there are likely very different types of skills and abilities (competences) involved with different domains of expertise, and these variations are likely to be reflected in the development and study of a certain indicator of expertise. Notwithstanding the problem of the moving target, some categorization might help here: are there kinds of expertise that may involve different sets of skills and demands? Collins and Evans (2007) have provided very useful categorizations for

expertise. Martini (forthcoming) also offers some suggestions on how to categorize some types of expertise, but more work on that needs to be done.

In the end, it is likely that a good understanding of expertise will come from the sociological and philosophical literature, but only in combination with the empirical and psychological literature on recognizing experts by means of proxies and indicators (see Shanteau 1992; Shanteau et al. 2002). Such collaboration is still in its infancy, but it might help in at least two ways: On the one hand, to move past the current standards for the acceptance of expert testimony in court, by giving criteria for separating true experts from just putative experts; instead of focusing on issues of method and relevance that laypeople, courts, or judges, are unlikely to be able to evaluate. On the other hand, it might help to sidestep what I called, at the beginning of this paper, the "paradox of proof": how can a layperson recognize expertise. A layperson can recognize expertise by looking at proxies and indicators of expertise, all of which need to be carefully developed and tested against evidence: This is both an empirical and theoretical task.

REFERENCES

Bovens, L., Hartmann, S. (2002). Bayesian Networks and the Problem of Unreliable Instruments. *Philosophy of Science*, 69:29–72.

Buekens, F., Truyen, F. (2014). The Truth About Accuracy. In *Experts and Consensus in Social Science*. Carlo Martini and Marcel Boumans (Eds.). Berlin, Springer (2014): 213-229.

Claveau, F. (2013). The Independence Condition in the Variety-of-Evidence Thesis. *Philosophy of Science*, 80(1):94-118

Collins, H., Evans, R. (2002). The Third Wave of Science Studies: Studies of Expertise and Experience. *Social Studies of Science*, 32(2):235-296.

Collins, H., Evans, R. (2007). *Rethinking expertise*. Chicago: University of Chicago Press.

Collins, H. (2013). The Core of Expertise. *Phenomenology and the Cognitive Sciences*, 12:399–416.

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

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

Gatowski, S.I., et al. (2001). Asking the gatekeepers: A national survey of judges on judging expert evidence in a post-Daubert world. *Law and Human Behavior* 25(5):433–458.

Haack, S. (2005). Trial and Error: The Supreme Court's Philosophy of Science. *American Journal of Public Health,* 95(S1):S66-S73. Reprinted in Susan Haack, *Evidence Matters. Science, Proof and Truth in the Law*. Cambridge: Cambridge University Press (2014).

Kitcher, P. (1990). The division of cognitive labor. *The Journal of Philosophy*, 87(1): 5-22.

Kitcher, P. (2010). The climate change debates. *Science*, 328:1230–1234.

Kuhn, T. (1970). *The Structure of Scientific Revolutions. Second Edition, Enlarged*. Chicago: Chicago University Press.

Lehtinen, A. (2013). On the Impossibility of Amalgamating Evidence. Journal for General Philosophy of Science, 44(1):101-110.

Martini, C. (forthcoming). Expertise and Institutional Design in Economics. Forthcoming in the *Journal of Economic Methodology*.

Oreskes, N., Conway, E.M. (2010). *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues From Tobacco Smoke to Global Warming*. Bloomsbury Publishing USA.

Saks M.J., Faigman, D.L. (2005). Expert evidence after Daubert. *Annual Review of Law and Social Science*, 1:105–30.

Shanteau, J., et al. (2002). Performance-Based Assessment of Expertise: How to Decide If Someone Is an Expert or Not. *European Journal of Operational Research* 136.2 (2002): 253-263.

Shanteau, J. (1992). The psychology of experts an alternative view. In: G. Wright and F. Bolger (Eds.), *Expertise and Decision Support*. NY: Plenum Press (1992): 11-23.

Smith, A.1976 (2007). *An Inquiry into the Nature and Causes of The Wealth of Nations*. Jonathan B. Wight (Ed.). Harriman House.

Stegenga, J. (2013). An Impossibility Theorem for Amalgamating Evidence. *Synthese*, 190(12):2391-2411.