

The gatekeeper’s dilemma: expert testimony, scientific knowledge and judicial reasoning

Edoardo Peruzzi*¹ and Gustavo Cevolani^{†2}

¹Institute of Philosophy, Leibniz University Hannover

²IMT School for Advanced Studies, Lucca, Italy

March 2024

Abstract

We examine the relationship between scientific knowledge and the legal system with a focus on the exclusion of expert testimony from trial as ruled by the *Daubert* standard in the US. We introduce a simple framework to understand and assess the role of judges as “gatekeepers”, monitoring the admission of science in the courtroom. We show how judges face a crucial choice, namely, whether to limit *Daubert* assessment to the abstract reliability of the methods used by the expert witness or also to check whether the application of those methods was correct. Undesirable outcomes result from both choices, thereby giving rise to the “gatekeeper’s dilemma.” We illustrate the dilemma by analyzing in some detail two well-known cases of *Daubert* challenges to economic experts. Finally, we present reasons for the absence of straightforward solutions to the dilemma and for its likely endurance.

Keywords. Expert Testimony; *Daubert*; Reliability; Scientific Expertise.

Contents

1	Introduction	3
2	The <i>Daubert</i> standard and Rule 702	5
3	Reliability and applicability questions	7

*edoardo.peruzzi@philos.uni-hannover.de

†gustavo.cevolani@imtlucca.it

4	The gatekeeper’s dilemma	10
5	What ways out of the gatekeeper’s dilemma?	13
5.1	Objection 1: Stick to the rules	13
5.2	Objection 2: Employ neutral experts	15
5.3	Objection 3: Amendment to Rule 702	16
6	Conclusion	18

Should the Daubert test apply only to the principles upon which the expert bases her testimony, or should Rule 702 also require that the application of the principles must be reliable as well? For example, with DNA tests, is it only necessary to show that the technique of DNA identification is reliable, or must it also be shown that the test was reliably conducted in the specific case?

— Daniel J. Capra¹

1 Introduction

In December 2023, the United States Congress and Supreme Court ratified an amendment to the Federal Rule of Evidence 702, marking the latest act in the century-long history of the relationship between science and law in US jurisprudence. The amendment is meant to clarify the correct standard for the admission of scientific experts in light of frequent misapplication by courts. Among the self-proclaimed goals of the Advisory Committee is to emphasize that, in evaluating expert testimony, “the court must focus on the expert’s opinion, and must find that the opinion actually proceeds from a reliable application of the methodology.”²

The recent amendment comes as a result of a long-standing debate over courts’ gatekeeping role vis-à-vis expert testimony. In short, the controversy has concerned whether courts should evaluate the application of the expert’s scientific knowledge to the case at hand or limit themselves to the assessment of the abstract reliability of such knowledge. While legal scholars have written extensively on the subject, among philosophers the issue has taken a back seat.³ This paper aims to fill the lacuna by providing a rational reconstruction of judicial reasoning in evaluating scientific testimony and showing the existence of a dilemma that threatens such evaluation.

Expert testimony based on scientific knowledge is becoming increasingly dominant in the adjudication of many civil and criminal cases. Just to mention a couple of examples: DNA testing often serves as a key piece of evidence in criminal cases; pharmacological studies come into play in toxic tort litigation, where the central question revolves around whether exposure to certain chemicals causes specific diseases; psychologists and psychiatrists routinely testify about whether a murderer suffers from a mental disorder; economists provide expert testimony in various legal areas, ranging from antitrust enforcement to employment discrimination cases; and the list goes on.

In the majority of jurisdictions worldwide, science enters legal proceedings through the testimony of expert witnesses employed by litigants. Although the primary purpose is to assist courts in getting the facts right, litigants often find it advantageous to portray non-scientific claims as robust scientific evidence to secure a favorable case outcome. Consequently, an increased reliance on scientific testi-

1. Memorandum from Daniel J. Capra, Reporter to Advisory Comm. on Evidence Rules, to Advisory Comm. on Evidence Rules (Sept. 11, 1997), cited in Bernstein and Lasker (2015, p. 14).

2. Memorandum of Advisory Committee on Evidence Rules, p. 870. Retrieved from: https://www.uscourts.gov/sites/default/files/evidence_rules_report_-_may_2022_0.pdf.

3. A noteworthy exception is found in the extensive scholarship of Susan Haack, as exemplified in Haack (2014).

mony may increase the risk of deception by experts. After all, individuals highly proficient in a particular discipline wield considerable influence over judges or jurors who may lack literacy in the relevant scientific field, thus making them susceptible to persuasion or misinformation.⁴ Recognizing this, legal systems worldwide have long sought remedies to prevent the infiltration of pseudoscience or bad science into the courtroom.

In the US, the guiding rule for addressing scientific evidence in legal proceedings was provided by the 1993 Supreme Court’s *Daubert* decision and the ensuing Rule of Evidence 702. In essence, the *Daubert* standard allows courts to ask questions both about the reliability of the expert and the scientific theory or empirical method she employed (reliability questions) and about the application of that theory or method to the particular case (applicability questions). In this role of “gatekeepers”, judges must monitor the admission of scientific arguments in the courtroom.

While some courts are happy to conclude a *Daubert* hearing after being convinced of the reliability of certain theories and methods employed by the expert witness, others find it necessary to control whether such theories and methods are correctly applied given the particular characteristics of the case at hand. Although we observe such a variety of attitudes, our focus centers on the normative dimension of this deliberation: after a positive answer to reliability questions, should courts conclude their *Daubert* hearings or should they proceed to ask applicability questions?

While such a question has been long debated by legal scholars, we approach it from the perspective of the philosophy of science. Given what we know about scientific knowledge, we demonstrate that both choices have undesirable outcomes – hence, what we call the “gatekeeper’s dilemma.” If judges merely ask reliability questions, then they are likely to admit an expert testimony that relies on a misapplication of scientific theories and empirical methods. On the contrary, if courts also raise applicability questions, then the lack of clear and widely accepted rules of applicability in many scientific domains makes it impossible to answer them.

To illustrate the gatekeeper’s dilemma, we discuss two prominent examples where economic experts faced challenges to admissibility under *Daubert*. These two legal cases will prove useful in explaining how a *Daubert* hearing works and what the gatekeeper’s dilemma consists of in practice. We chose economics for two main reasons. First, economics ranks among the most frequently challenged disciplines under the *Daubert* standard, placing fourth after medicine, psychology, and engineering.⁵ Second, some methodological aspects of economics help in clarifying the gatekeeper’s dilemma, making it especially hard to solve. However, as we will argue, most of our conclusions aim to be more general and not limited to economics.

We proceed as follows. In Section 2, we explain how *Daubert* hearings work, and briefly introduce our two case-studies. In Section 3, we present what we call the “*Daubert* decision tree” as a rational reconstruction of judicial reasoning during *Daubert* hearings. This leads us to the gatekeeper’s dilemma, which we discuss in Section 4 with reference to our case-studies. Finally, in Section 5 we raise and

4. A consistent body of evidence exists showing that judges and jurors have difficulty understanding scientific evidence. See, e.g., Lopatka (2016), Hans and Saks (2018), and Koehler (2018) and references therein. Deception by scientific experts can also be unconscious, possibly stemming from cognitive biases that affect the impartiality and reliability of expert decision-making (e.g., Dror 2016; Dror and Hampikian 2011)

5. Source: www.dauberttracker.com, accessed January 17, 2024.

answer three possible objections pointing to three quick ways out of the dilemma. Section 6 concludes the paper.

2 The *Daubert* standard and Rule 702

Federal Rule of Evidence 702 (henceforth, Rule 702) provides the legal foundation for federal courts to determine the admissibility of expert testimony in legal proceedings. Since 2000, for about twenty years, Rule 702 has read as follows:

A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if:

- (a) the expert’s scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;
- (b) the testimony is based on sufficient facts or data;
- (c) the testimony is the product of reliable principles and methods; and
- (d) the expert has reliably applied the principles and methods to the facts of the case.⁶

Rule 702 was drafted in response to the Supreme Court decision in *Daubert v. Merrell Dow Pharm. Inc.* (1993) and to the many cases applying *Daubert* to expert testimony in the following years.⁷ While the original *Daubert* case focused on the potential link between Mrs. Daubert’s use of Bendectin and her children’s significant birth defects, the Supreme Court Justices faced a broader decision: whether to depart from the seventy-year-old existing standard for admitting expert testimony and scientific knowledge in litigation, known as the *Frye* standard.⁸ The *Frye* standard originated from a 1923 murder trial in which the court refused to admit a lie detector test as evidence of innocence. This standard mandated that courts evaluate the admissibility of scientific knowledge based on its general acceptance among experts and practitioners in the relevant field. For this reason, the *Frye* standard became to be known as the “general-acceptance” test for admitting scientific evidence.

The *Daubert* Court superseded *Frye* in two important aspects. First, in contrast to the *Frye* ruling, the *Daubert* Court held that it is not sufficient for expert testimony to be generally accepted by the scientific community. Instead, federal courts had to determine the admissibility of expert testimony by ensuring that “any and all scientific testimony or evidence admitted is not only relevant, but reliable”

6. Retrieved from: https://www.law.cornell.edu/rules/fre/rule_702, accessed November 15, 2023.

7. *Daubert* was refined in a series of Supreme Court’s decisions referred to as the *Daubert* trilogy. Such a trilogy of cases, which occurred between 1993 and 1999, includes *Daubert v. Merrell Dow Pharm. Inc.* 509 U.S. 579, *General Electric Co. v. Joiner* 522 U.S. 136 (1997), and *Kumho Tire Co. v. Carmichael* 526 U.S. 137 (1999). The legal scholarship on the *Daubert* trilogy is extensive. See, among others, Bernstein (2007, 2013), for a history of rules for the admissibility of expert testimony in the US before and after *Daubert*.

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

(*Daubert*, p. 589).⁹ Second, the *Daubert* Court firmly established federal courts as “gatekeepers” of expert opinion, with the ability (regularly exercised) to exclude the testimony of scientific experts when the judge does not believe the individual or methods employed is reliable enough to be admitted at trial.

How are courts to screen expert testimony for reliability in the practice? Federal courts normally decide on the admissibility of expert testimony during the so-called *Daubert* hearings. *Daubert* hearings usually follow the filing of a *Daubert* challenge – a legal motion presented by a party to the judge before the start of litigation proceedings, with the purpose of excluding the other party’s expert witness. After issuing the challenge, the court takes the lead and assumes its gatekeeper’s role by applying *Daubert*/Rule 702 to evaluate expert testimony.

To see how *Daubert* hearings work, let us consider two examples, which will return several times in the rest of the article. The first case – *Concord Boat v. Brunswick* (2000) – is perhaps the most renowned example of a *Daubert* exclusion to an economic expert.¹⁰

In 1998, a number of boat builders brought an antitrust action against the stern drive engine manufacturers Brunswick Corp. Plaintiffs claimed that Brunswick engaged in various anticompetitive schemes, including the use of exclusionary discount programs where boat builders could agree to purchase a certain percentage of Brunswick’s engines in exchange for a discount off the list price of the engine. In support of Concord Boat’s damage claim, plaintiffs hired Stanford economist Robert Hall. Professor Hall relied on the Cournot duopoly model to represent the counterfactual market of boat engines, that is, the market that one would observe without Brunswick’s anticompetitive conduct. A standard workhorse model of industrial economics, the Cournot model describes a simple market containing just two firms, both producing exactly the same product at constant marginal costs and competing on the amount of product they produce.¹¹ Applying the Cournot model, Hall contended that in the absence of anticompetitive behavior, both Brunswick and another firm would have each held a 50% market share.

Defendants raised a motion to exclude Hall’s expert testimony before the Eight Circuit Court of Appeals. The appellate judges agreed with the defendant’s claim that Hall’s model attributed to the challenged conduct all sales made by Brunswick above half of the total. By doing this, Hall’s model ignored the fact that Brunswick already had about a 75% market share even before undertaking the challenged conduct. Therefore, Hall’s expert opinion was excluded due to the incompatibility of his model with the economic market it aimed to represent.

The second case, *United States v. Am Express Co.* (“AmEx” for short), saw the US Department of Justice contest several anticompetitive trade restrictions imposed by American Express Company in the credit card market.¹² To bolster the defendant’s strategy, New York University Professor Janusz Ordover employed instrumental variables regression to show a positive causal relationship between the share of merchants accepting AmEx cards and the share of total spending by cardholders. Instrumental variables regression provides a way to obtain consistent

9. See Haack (2005) for a critical assessment of the concept of reliability in the *Daubert* ruling.

10. 207 F.3d 1039, (8th Cir, 2000). See, among others, Hovenkamp (2005, ch. 4) and Giocoli (2020) for an assessment of the case.

11. For a textbook treatment of the Cournot model, see Tirole (1988).

12. 10-CV-4496 (NGG) (RER) (E.D.N.Y. June 24, 2014).

parameter estimates even when an explanatory variable of interest is correlated with the error term, in which case ordinary least squares yields biased results.

Despite being a standard device in the toolbox of applied economists, plaintiffs urged the court to exclude Ordover’s expert testimony “on the grounds that his analysis was based on inappropriate variables, or ‘weak instruments,’ in econometric terms” (*AmEx*, p. 6). In contrast to *Concord Boat*, however, the *AmEx* court denied the motion to exclude because the expert opinion “is not plagued by the type of overt flaws that would render the regression so unsound as to be irrelevant and unhelpful to the court” (*AmEx*, p. 9).

The two cases just discussed clearly illustrate how *Daubert* hearings work in practice, as well as how different their outcomes can be. In the next section, we offer a simple model of the underlying reasoning, highlighting some of its crucial issues.

3 Reliability and applicability questions

Daubert hearings exhibit significant diversity: some are short and superficial, while others are remarkably lengthy and detailed. Many factors contribute to such variability, including the personality of the presiding judge, the legal domain relevant to the case, and even the nature of the scientific disciplines under scrutiny.¹³ Nonetheless, we believe it is feasible and useful to rationally reconstruct judicial reasoning during *Daubert* hearings at a slightly more abstract level.

In our reconstruction, courts dealing with a *Daubert* challenge have to consider two different kinds of questions. The first concerns the reliability, in the abstract, of the scientific theories and empirical methods employed by the expert witness. The second kind of questions focuses on the possibility and opportunity of applying such theories and methods to the present case. We shall refer to the first kind of question as a “reliability question”, and to the second as an “applicability question”. We can schematically represent them as follows:

Reliability question Are the scientific theories and empirical methods as employed by the expert a reliable source of knowledge in general?

Applicability question Are the scientific theories and empirical methods employed by the expert correctly applied to the particular case?

Reliability and applicability questions may take different forms depending on the particular nature of the litigation and the expert testimony considered. For example, when confronted with a multivariate regression analysis, the court may investigate both the expert’s credentials as an economist and the prior use of the regression in comparable lawsuits (reliability questions). Within the same *Daubert* hearing, the court might also assess whether the regression model includes all relevant explanatory variables, thereby mitigating the risk of omitted variable bias (applicability questions).¹⁴

13. Helland (2019) provides some evidence that the ideology of judges (as measured by the party of the appointing president) influences the decision to grant a motion to exclude scientific experts

14. A third kind of questions that arises in legal proceedings pertains to what we could call “credibility questions.” These questions seek to determine whether the scientific theories and empirical methods are capable of addressing the specific issues for which they are

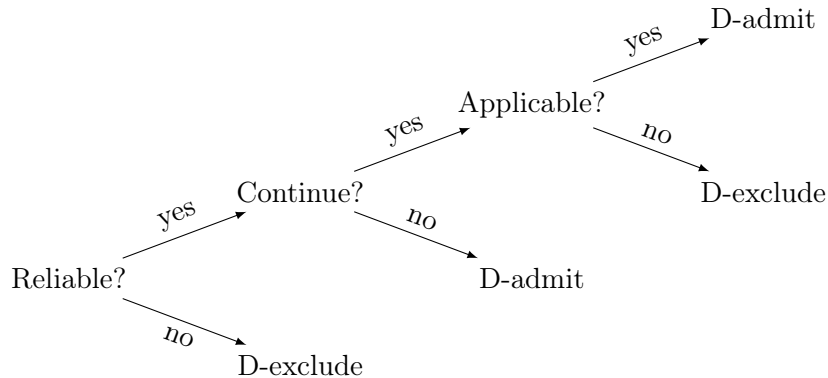


Figure 1: The *Daubert* decision tree. Both reliability questions (“Reliable?”) and applicability questions (“Applicable?”) may lead to both admitting (“D-admit”) and excluding (“D-exclude”) an expert testimony under *Daubert*/Rule 702. The node “Continue?” represents the choice to halt *Daubert* scrutiny after a positive answer to the initial reliability question or proceed with applicability questions.

Reliability and applicability questions are asked sequentially, and federal courts are represented as making YES/NO decisions along a tree, to which we refer as to the “*Daubert* decision tree” (Figure 1). A *Daubert* Decision Tree initiates with a reliability question. If the response to the reliability question is negative, the expert testimony is excluded under *Daubert*/Rule 702 (D-exclude, for short). A negative reply to reliability question stems from two main factors: (i) either the court believes that the expert witness is not a genuine expert or, more likely, (ii) the court believes the expert, even if genuine, employed an unreliable theory or method. However, negative responses to reliability questions are rare, as parties usually seek to engage highly reputable experts, such as university professors or professional scientific consultants, who consistently utilize methods endorsed by the scientific community. This usually ensures positive answers to reliability-related questions.

If the initial reliability question receives a positive response, the court has to make a crucial choice: deciding whether to conclude its *Daubert* scrutiny and admit the expert testimony or to additionally raise applicability questions. We represent this choice with the “Continue?” node in the *Daubert* Decision Tree.

Some courts decide to halt the *Daubert* examination after a positive answer to reliability questions, as in the *AmEx* case, where the point of contention was whether the defendant’s economic expert applied the instrumental variables method correctly. District judge Nicholas Garaufis dismissed the *Daubert* motion arguing

employed and aligning with the other evidence presented during the trial. There is little controversy that credibility questions lie beyond the purview of the court during a *Daubert* hearing. Assessing the extent to which the scientific arguments of an expert support a conclusion about the case or how they cohere with the broader body of available evidence is, particularly in common law countries, the responsibility of the jury (see Bartholomew 2014; Haw Allensworth 2012).

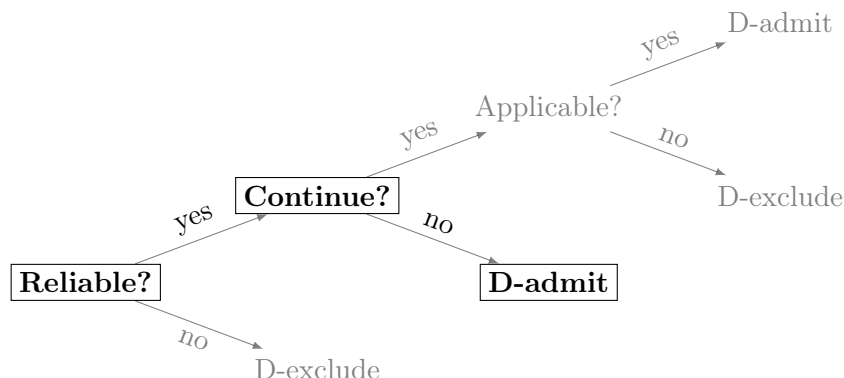


Figure 2: A *Daubert* decision tree leading to admission after a positive answer to reliability questions. The bold path represents the one followed by the court during the *Daubert* hearing in the *AmEx* case.

that for what concerns the admissibility of expert testimony, the only important criterion was that regression analysis is an accepted tool in economics research and has passed muster in previous *Daubert* hearings (see Figure 2). “Plaintiffs,” the judge explained, “do not and indeed cannot, contend that multiple regression analysis is not itself a well-established and reliable econometric methodology frequently relied upon by federal courts under Rule 702.” (*AmEx*, p. 8)

Other cases exhibit a different attitude by federal courts: after an (implicit or explicit) positive answer to reliability questions, they go on and raise applicability questions. For instance, in *Concord Boat* the swing of the *Daubert* challenge was about the application of Cournot’s model to represent the counterfactual market (see Figure 3). The appellate judges explicitly raised an applicability question by wondering whether the assumptions of the Cournot model were realistic enough for the model to be applicable to the target market. The Court of Appeal’s own words are very clear:

Even a theory that might meet certain *Daubert* factors, such as peer review and publication, testing, known or potential error rate, and general acceptance, should not be admitted if it does not apply to the specific facts of the case. (*Concord Boat*, pp. 1056–1057)

Concord Boat best exemplifies one of the many cases where *Daubert* challenges are not about the reliability of a model or an empirical method *per se*, but rather about their application to a particular circumstance.¹⁵ In these instances, a positive response (implicit or explicit) is given to reliability questions, while the bulk of the *Daubert* hearing takes place on applicability.

So far we have seen that in many *Daubert* hearings, federal courts must choose whether a positive answer to reliability questions suffices for admitting the expert testimony under *Daubert*/Rule 702, or if it is necessary to also address applicability questions. Empirically, we observe that courts exercise a considerable degree of discretion: while some courts are comfortable concluding the *Daubert* assessment

¹⁵. Further examples are provided in Peruzzi (2023).

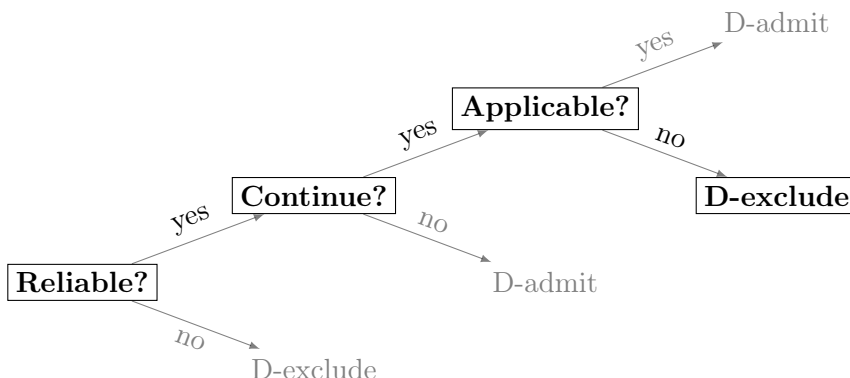


Figure 3: A *Daubert* decision tree leading to exclusion after a negative answer to applicability questions. The bold path represents the one followed by the court during the *Daubert* hearing in the *Concord Boat* case.

upon affirming reliability questions, others find it necessary to also consider applicability questions. However, our focus here is on the normative aspect of such a decision: after receiving a positive answer to reliability questions, should the court conclude its *Daubert* hearing, or should it proceed to raise applicability questions?

What we wish to emphasize is that the question of what courts should do when engaged in *Daubert* hearings is not only of legal interest, but also holds significance from a philosophical standpoint. Specifically, we find it promising to scrutinize such a question through the lens of philosophy of science. Roughly put, we will answer the following question: considering our understanding of how science operates, what should courts do after a positive answer to reliability questions?

4 The gatekeeper’s dilemma

In a *Daubert* decision tree, a positive answer to reliability questions give federal courts the responsibility of making a choice – specifically, deciding whether to stop their *Daubert* assessment or to raise applicability questions. What is the most advisable course of action for courts, given what we know about how science works?

As we argue, the answer to this apparently simple question leads to what we call “the gatekeeper’s dilemma:”

The gatekeeper’s dilemma. Suppose that reliability questions are positively answered. If courts stop their *Daubert* scrutiny, then they are likely to admit to the trial expert testimonies based on misapplications of scientific theories and empirical methods. If courts raise applicability questions, then the lack of clear and widely accepted rules of applicability will make it basically impossible to answer such questions.

A dilemma typically involves a situation in which a hard choice must be made between two or more undesirable alternatives. The gatekeeper’s dilemma is such in the sense that both choices – either restricting *Daubert* scrutiny to reliability questions or also introducing applicability questions – lead us to undesirable outcomes.

In this section, we show how the dilemma works in real cases; in the next section, we analyze it from a more abstract point of view.

Were *Daubert* scrutiny always halted after positive answers to reliability questions, the risk of admitting expert testimony based on faulty applications of scientific methods would greatly increase for sure. Scientific knowledge is full of theories and empirical methods that are reliable in the abstract but whose application to particular cases must be carefully controlled. By foregoing such control, courts expose themselves to a significant risk of admitting testimony founded on erroneous applications of generally reliable theories and methods. As legal scholar Hovenkamp (2005, p. 89) puts it

At a high enough level of generality virtually any methodology seems to pass muster under the *Daubert* criteria. But statistical methodologies can be misused, and often are grossly misused. Suppose, for example, that the plaintiff’s expert uses statistics in a highly idiosyncratic way, perhaps making serious errors by failing to control for obvious outliers in the data, or drawing conclusions much stronger than the data permit. It is hardly useful for the judge to proclaim that “statistics” is a widely accepted and reliable methodology of scientific investigation. Even an astrologer might use a telescope properly to observe the motions of celestial bodies, but that would not serve to validate his testimony that the alignment of the planets explained why his client murdered the victim.

This observation holds true for several areas of science. That regression analysis and other econometric techniques are generally accepted in the academic community and subjected to peer review, does not tell much about the validity of a particular regression model employed by an economic expert. In a similar vein, Haack (2014, p. 102) argues that “in the case of DNA identification, where there is solid underlying science, the most important thing is to ensure that these techniques aren’t misapplied through haste, sloppiness, mismanagement, or dishonesty, conscious or unconscious.” Even reliable scientific theories or empirical methods – whether it be statistical analysis, telescope-based observation or DNA testing – can be grossly misapplied. Therefore, confining a court’s *Daubert* hearing solely to reliability questions seems in general unreasonable.

The court’s reasoning in *AmEx* epitomizes the risk of halting the *Daubert* assessment after reliability questions. The core of the *Daubert* challenge in *AmEx* revolved around the allegation that the expert had used a weak instrument, implying biased estimates of parameters and standard errors. Applied economists employ instrumental variables (“instruments”) to estimate causal relationships and rely on specific conditions whose occurrence must be checked in the particular case. Such a method allows for consistent estimation if and only if, among other things, the instrument is highly correlated with the endogenous explanatory variables. When the instrument is highly correlated with the endogenous explanatory variable, it is referred to as a “strong instrument.” Conversely, if the correlation is low, it is termed a “weak instrument.”¹⁶ When these conditions are not satisfied, the instrumental variables method provides misleading inferences about the parameter

16. In general, a valid instrument induces changes in the explanatory variable but has no independent effect on the dependent variable, allowing a researcher to uncover the causal effect of the explanatory variable on the dependent variable. The main conditions for

estimates and standard errors. Hence, by stopping the *Daubert* evaluation at reliability questions, the *AmEx* court run the risk of admitting expert testimony based on the misapplication of a seemingly reliable empirical method.

Given that halting the *Daubert* scrutiny at reliability questions yields undesirable consequences, one might think it would be better to always proceed to raise applicability questions. Here, however, a more fundamental problem arises. Indeed, federal judges are likely unable to assess whether scientific theories and empirical methods have been correctly applied to particular cases. This is because many scientific disciplines lack clear and agreed-upon rules of applicability that, once followed by scientific experts, ensure that scientific theories and empirical methods can be successfully applied to explain or predict a particular case. In the absence of such rules, courts found themselves in the nearly impossibility of answering the majority of applicability questions.

Take again *AmEx*. Before dismissing the *Daubert* challenge based on a vague appeal to the reliability of regression analysis, Judge Garaufis examined the plaintiff's arguments. In our terms, the judge proceeded to check whether the instrumental variables method was correctly applied by the defendant's expert. As a matter of fact, he quickly realized that there was no rule such that below a certain threshold the correlation between the instrument and endogenous variable is too weak for the estimation to generate unbiased results. "Plaintiffs", he claimed, "acknowledge that there exists no econometric rule dictating what particular F-statistic values demonstrate the use of weak instruments or what values necessarily render an analysis unreliable." Moreover, he added, "plaintiff's own expert has acknowledged that there exists no bright-line standard to differentiate strong from weak instruments." (*AmEx*, p. 8)

The *AmEx* court faced a dilemma: Judge Garaufis attempted to pose an applicability question, but upon realizing that it was impossible to answer due to the absence of clear-cut rules of applicability, he admitted the expert testimony based on the positive response to reliability questions. In this way, the *AmEx* court chose to bear the risk of admitting into the trial an expert testimony founded on an incorrect application of the instrumental variables method.

The Court of Appeals faced the same dilemma in its *Daubert* decision in *Concord Boat*. In response to the plaintiff's expert testimony, the defendant raised questions about the applicability of the Cournot duopoly model, stating that it predicts an evenly divided market only under specific assumptions – namely, when products are homogeneous and costs are uniform. "There is no dispute", the defendant argued, "that marine engines are differentiated, with manufacturers offering diverse product features, promotions, support, and [...] that engine manufacturers had different costs."¹⁷ Against these allegations, the plaintiff vigorously defended Dr. Hall's choice of the Cournot model:

Dr. Hall testified [...] that the Cournot model has been shown to work quite well in predicting performance in many markets, including markets where the products are somewhat differentiated. Many eminent economists have noted the usefulness of the Cournot model. [...] Moreover, Dr. Hall extensively testified that he chose the Cournot

consistent estimation using instrumental variables are known as the "exclusion restriction" and the "relevance condition." See Cunningham (2021) for a textbook treatment.

17. Reply Brief: Appellant-Petitioner. 1999 U.S. 8th Cir. Briefs LEXIS 20. Retrieved from: CONCORDBOATCORP.v.BRUNSWICKCORP.,1999U.S.8thCir.BriefsLEXIS20.

model to use in estimating damages only *after* he conducted a comprehensive analysis of the engine market, and after having looked closely at the relationship of price to cost in this industry. (Initial Brief, p. 132, original emphasis)¹⁸

The court grappled with two compelling arguments. On the one hand, the defendant contended that the unrealistic assumptions underpinning Cournot’s model rendered it unsuitable for accurately determining damages. On the other hand, the plaintiff maintained that economists routinely employ the Cournot model to represent markets that do not align with its foundational assumptions.¹⁹

Faced with the problem of assessing the applicability of the Cournot duopoly model to the boat engine market, the court could not lean on specific rules governing the level of realism necessary for applying the model to a market. This is due to the absence of such rules: the assessment of the similarity of assumptions between a model and a target lacks precise guidelines or a checklist, and disagreement among economic experts is commonplace.

In conclusion, we think we have identified a dilemma that rests on the shoulders of judges during *Daubert* hearings. The following section will tackle potential objections to the gatekeeper’s dilemma.

5 What ways out of the gatekeeper’s dilemma?

The gatekeeper’s dilemma, as we defined it, has two horns: either restricting *Daubert* scrutiny to reliability questions or also raising applicability question. We believe that there is no need to convince the reader that stopping at reliability questions has undesirable consequences. To support the second horn of the dilemma, however, we must confront three main objections.

5.1 Objection 1: Stick to the rules

The first objection argues that science actually possesses clear and widely agreed-upon rules for the applicability of scientific theories and empirical methods. While these rules may evolve over time, scientists, at any given moment, are aware of them and can adhere to them if genuinely motivated. At most, the gatekeeper’s dilemma affects only those scientific disciplines that have not yet developed such rules, but it certainly does not apply to science as a whole.

We find such an objection to be reasonable as we think that, while rules of applicability plays an important role in all scientific fields, it is undeniable that their clarity and consensus comes in various degree. In some cases, a well-established and widely accepted set of rules, commonly known as “best practices,” governs the application of theories and empirical techniques to specific cases. In these favorable instances, scientists possess a checklist to guide them when applying a certain theory or method to a particular case. For example, scientists engaged

18. Initial Brief: Appellee-Respondent. 1999 U.S. 8th Cir. Briefs LEXIS 53. Retrieved from: CONCORDBOATCORP.v.BRUNSWICKCORP.,1999U.S.8thCir.BriefsLEXIS53.

19. Under some circumstances, economists may increase the range of applicability of their models by removing unrealistic assumptions (for relevant examples, see Peruzzi and Cevolani 2022) However, it is widely acknowledged in the methodological literature that de-idealization faces both pragmatic and conceptual limitations.

in DNA identification in legal proceedings must adhere to stringent guidelines to ensure accuracy, including factors such as storage temperature and light conditions. Following this checklist ensures that the determination of a DNA profile is as fast and accurate as possible.²⁰

Although some brilliant cases, such as DNA analysis, exist in modern science, we believe that the scope of the second horn of the gatekeeper’s dilemma remains intact. While some areas of science may have developed precise guidelines for applying their specialized knowledge to specific cases, we contend that many disciplines lack clear and universally accepted rules. What is even more troubling is that these very disciplines – particularly medicine, psychology, and economics, which typically lack such guidelines – are the ones most frequently targeted by *Daubert* challenges.

In this connection, the case of economics is paradigmatic. For instance, there is no rule about how realistic a model’s assumptions must be with respect to the target in order for it to be used to explain or predict in particular circumstances. In his analysis of economic modeling, Harvard macroeconomist Dani Rodrik (2015) advocated for the importance of considering the realism of critical assumptions when applying economic models for explanation and prediction from a practitioner’s perspective. While he proposed heuristic guidelines such as “model assumptions must be approximately true” or “critical assumptions must be realistic”, these suggestions fall short of constituting a universally agreed-upon set of rules of applicability. To be fair, Rodrik and his followers seem to agree that judging which model assumptions are critical in a given application (and, therefore, must be realistic) involves an undeniable subjective judgement on the part of the economist. Indeed, as Rodrik himself reiterates in his book, “there is an unavoidable craft element involved in rendering models useful.” Rodrik (2015, 64).²¹

Likewise, economists do not have a rule for deciding which covariates should be included in a regression model as explanatory variables. It is well known by applied economists that the validity of our econometric models depends on the *a priori* confidence that we are considering all relevant factors in the given circumstance. As vividly pointed out in a classic article by the applied econometrician Leamer (1983, p. 32), “as every beginning econometrics student knows, if you omit from a model a variable which is correlated with included variables, bad things happen.” However, there are no golden rules to understand whether a model is well specified (that is, important variables have not been left out of the model equations) in non-experimental settings studied by economists.²²

Those are only two simple examples, but the general lesson should be clear: in economics (and social sciences in general), rules of applicability are not easy to formulate or may not even exist. This point is not a novel one; in fact, it has a long

20. See, for example, the 2022 US Department of Justice’s report on best practices for DNA identification: <https://www.ojp.gov/pdffiles1/nij/304051.pdf>.

21. Rodrik’s book has garnered considerable interest from philosophers of economics. For a thorough examination of Rodrik’s methodological stance, see the 2018 symposium of the *Journal of Economic Methodology* (Vol. 25, No. 3).

22. It is worth noting that the credibility revolution popularized by Angrist and Pischke (2010) has not been a game changer. First, the common practice in empirical research in economics is still based on taking the empirical model specification as known and rely on arbitrary decisions about the choice of explanatory variables (Moral-Benito 2015; Steel 2020). Second, in the context of legal proceedings quasi-experimental findings are hard to come by (Einav and Levin 2010).

tradition in the history of the methodology of economics. Milton Friedman, for example, makes a similar consideration about the challenge of applying economic models to particular cases:

The rules for using the model [...] cannot possibly be abstract and complete [...]. Each occurrence has some features peculiarly its own, not covered by explicit rules. The capacity to judge that these are or are not to be disregarded, that they should or should not affect what observable phenomena are to identified with what entities in the model, is something that cannot be taught; it can be learned but only by experience and exposure in the “right” scientific atmosphere, not by rote. (Friedman 1953, p. 25)

5.2 Objection 2: Employ neutral experts

The second objection contends that the gatekeeper’s dilemma is, at its core, a manifestation of a problem of expertise (or, more precisely, a lack thereof), and resolving the latter would naturally tackle the former. Courts and judges are laypeople in comparison to the experts whose testimonies they have to evaluate; why would anyone think that they are capable of making judgments about whether an expert witness has correctly applied scientific knowledge?²³ However, the objection goes, if judges were experts in a specific discipline, or more realistically, if they were assisted by a neutral expert, then they would be better equipped to assess the validity of the application of scientific knowledge to particular cases. In practical terms, this objection posits that judges, by appointing their own court experts, would make more informed decisions regarding the admissibility of expert testimony.²⁴

Unfortunately, given the adversarial nature of the US legal system, hiring a neutral expert does not provide a viable solution to the gatekeeper’s dilemma, for various reasons. First, practical problems exist, such as determining the criteria for selecting neutral experts and establishing responsibility for their payment, that pose significant challenges to appointing neutral experts. Second, a compelling argument arises regarding the existence of neutral experts in certain fields like psychology and economics. “There is no such thing as a neutral antitrust economics expert”, as Lopatka (2016, p. 456) pointed out. “Antitrust economics,” he continued, “is intensely ideological. Economists begin with different preconceptions about the functioning of the markets and the likelihood of anticompetitive conduct, and those differences may have a profound impact on their analysis of the case.”

Yet, in the case of economic expertise we believe that court-appointed experts suffer from a more profound problem. Courts are likely to seek the help of neutral experts to understand the validity of a particular application of a theoretical model or empirical technique. However, we have seen that the rules of applicability in economics are themselves open to expert disagreement. So, even a perfectly neutral expert will not be able to give truly decisive tips to the court. Let us recall the words of the district court in *AmEx*:

23. For a discussion of the problem of expertise in connection to the *Daubert* standard see, e.g., Haack (2014, 2015) and Martini (2015).

24. Posner (1999), Rubinfeld and Cecil (2018), and Sidak (2013) advocate for court-appointed experts, highlighting their potential role in *Daubert* hearings.

Plaintiffs acknowledge that there exists no econometric rule dictating what particular F-statistic values demonstrate the use of weak instruments or what values necessarily render an analysis unreliable. [...] Plaintiff’s own expert has acknowledged that “there exists no bright-line standard to differentiate strong from weak instruments.” (*AmEx*, p. 8)

Would resorting to the neutral expert have helped in such a case? We doubt it. The neutral expert, albeit not involved in party advocacy, would not have been able to provide the precise answers that Judge Garaufis was seeking. A comparable scenario arises in numerous *Daubert* disputes, especially when they concern the inclusion of relevant explanatory variables in regression models. While neutral experts could certainly offer their reasoned opinion, which might be well-founded, that the expert witness omitted crucial variables, the absence of a defined procedure for specifying the model leaves lingering doubts that the court aims to resolve.²⁵

Contrary to the objection, therefore, we believe that the gatekeeper’s dilemma is not solely rooted in the fact the judges lack scientific expertise to decide on the admissibility of expert testimony. Instead, we argued that the dilemma derives from methodological properties specific to economics and, arguably, other sciences. The dilemma would persist even if judges were fully versed in economics or had enlisted the best expert available on the market. The absence of clear and widely agreed-upon rules of applicability poses formidable methodological challenges for judges. This is because, when confronted with disagreement between the two experts representing the parties in dispute, the judge cannot rely on unequivocally recognized rules for applying economic models and econometric methods to specific cases. Consequently, courts engaged in *Daubert* hearings find themselves grappling with the near impossibility of addressing the majority of applicability questions.

5.3 Objection 3: Amendment to Rule 702

While the first two objections attack the gatekeeper’s dilemma – particularly its second horn – a third objection admits that the dilemma existed but was in fact resolved by the latest amendment to Rule 702. US lawmakers have long been aware of the variety of attitudes courts take in *Daubert* hearings, with increasing voices pushing for a reform of the language of Rule 702 to clarify the court’s gatekeeping role.²⁶ The whole debate can be rephrased again as follows: should the court limit itself to reliability questions or also raise applicability questions?

The Advisory Committee of Evidence Rules, which met in Washington, D.C. in May 2021, gave final approval for a rephrasing of the language of Rule 702 to clarify that courts can (or rather, should) ask applicability questions. “The Committee”, we read, “unanimously favored a slight change to existing Rule 702(d) that would emphasize that the court must focus on the expert’s opinion, and must find that

25. To be sure, we are not suggesting that court-appointed experts can have no role in a *Daubert* decision. On the contrary, the court-appointed expert surely sits in a better position than the judge to evaluate the application of economic theory to particular cases. However, we argue that neutral experts can hardly speak the clear words that courts seek in such cases.

26. See, for instance, Bernstein and Lasker (2015), Mangrum (2022), and Schroeder (2020).

the opinion actually proceeds from a reliable application of the methodology.”²⁷ The Supreme Court approved the change to Rule 702, among other amendments to various federal rules, and transmitted the proposed edits to Congress in April 2023. The new rule amendments took effect on December 1, 2023.

The amended clause (d) of Rule 702 now reads:

Original Rule 702(d): the expert *has reliably applied* the principles and methods to the facts of the case

Amended Rule 702(d): the *expert’s opinion reflects a reliable application* of the principles and methods to the facts of the case.

Although the change may appear insignificant, its goal seems very clear: to halt courts from taking a passive and overly liberal role in admitting expert testimony. In our terms, the Advisory Committee aims to avoid walking the first leg of the gatekeeper’s dilemma, that is, having courts asking only reliability questions. As we read in the Committee Note (*ivi.*, emphasis added):

The language of the amendment more clearly empowers the court to pass judgment on the conclusion that the expert has drawn from the methodology. Thus the amendment is consistent with *General Electric Co., v. Joiner*, 522 U.S. 136 (1997), in which the Court declared that a trial court must consider not only the expert’s methodology but also the expert’s conclusion; that is because *the methodology must not only be reliable, it must be reliably applied*.

While the impact of the amendment on expert testimony will need to be evaluated over time, on the surface it dissolves the gatekeeper’s dilemma by specifying that courts not only *can* but *must* check whether the scientific expert has properly applied scientific theories and empirical methods. In doing so, the amended Rule 702 prevents the possibility of terminating the *Daubert* investigation when a reliability question yields a positive answer.²⁸

In response to such an objection, a couple of comments are in order. First, we share the reasons behind the latest amendment to Rule 702 because we also believe that stopping *Daubert* hearings after positive answers to reliability questions poses significant risks from the perspective of philosophy of science. As we have argued, even the most reliable methodology can yield nonsensical results if applied incorrectly. At the same time, we do not see the amendment as a complete remedy to the gatekeeper’s dilemma. This is because the second horn of the dilemma remains intact. Admittedly, in cases where the rules of applicability are well-defined and widely accepted, the amendment shows promise, reinforcing the imperative for the court to delve into applicability questions that have clear and ascertainable answers. However, in other contexts involving, for example, the social sciences, the

27. Memorandum of Advisory Committee on Evidence Rules, p. 871. Retrieved from: https://www.uscourts.gov/sites/default/files/evidence_rules_report_-_may_2022_0.pdf.

28. “Under this amendment, the following statements, made by some courts in the past, are not supportable. These include: [...] ‘Whether the expert has properly applied the methodology is a question for the jury, not the court.’” Advisory Committee on Evidence Rules, Agenda for Committee Meeting (May 6, 2022), p. 148. Retrieved from: https://www.uscourts.gov/sites/default/files/evidence_agenda_book_may_6_2022.pdf

amendment may not provide a definitive solution. The persistent nature of the dilemma in these contexts stems from the absence of clear and widely accepted applicability rules governing the application of theories and empirical methods to specific circumstances.

6 Conclusion

Legal systems around the world seek to identify and exclude unreliable science from courtrooms, and the United States is no exception. Under Federal Rule of Evidence 702, US courts must act as gatekeepers of expert testimony, engaging in an analysis of its admissibility. During each *Daubert* hearing, federal courts must make a choice: decide whether to address only reliability questions or also applicability questions (as we have called them). Unfortunately, both choices have undesirable consequences, giving rise to the gatekeeper’s dilemma. Since we see no practical solutions on the horizon, the dilemma is likely here to stay.

References

- Angrist, Joshua D., and Jörn-Steffen Pischke. 2010. “The credibility revolution in empirical economics: How better research design is taking the con out of econometrics.” *Journal of economic perspectives* 24 (2): 3–30.
- Bartholomew, Christine P. 2014. “Death by *Daubert*: the continued attack on private antitrust.” *Cardozo Law Review* 35 (6): 2147–2198.
- Bernstein, David E. 2007. “Expert witnesses, adversarial bias, and the (partial) failure of the *Daubert* revolution.” *Iowa Law Rev.* 93:451–489.
- Bernstein, David E. 2013. “The misbegotten judicial resistance to the *Daubert* revolution.” *Notre Dame Law Rev.* 89 (1): 27–70.
- Bernstein, David E, and Eric G Lasker. 2015. “Defending *Daubert*: it’s time to amend federal rule of evidence 702.” *William & Mary Law Review* 57 (1): 1–48.
- Cunningham, Scott. 2021. “Causal inference: The mixtape.” New Haven & London: Yale University Press.
- Dror, Itiel E. 2016. “A hierarchy of expert performance.” *Journal of Applied Research in Memory and Cognition* 5 (2): 121–127.
- Dror, Itiel E., and Greg Hampikian. 2011. “Subjectivity and bias in forensic DNA mixture interpretation.” *Science & Justice* 51 (4): 204–208.
- Einav, Liran, and Jonathan Levin. 2010. “Empirical Industrial Organization: A Progress Report.” *Journal of Economic Perspectives* 24 (2): 145–62.
- Friedman, Milton. 1953. “Essays in positive economics.” Chap. The methodology of positive economics, 3–43. Chicago, IL: Chicago University Press.

- Giocoli, Nicola. 2020. "Rejected! Antitrust Economists as Expert Witnesses in the Post-Daubert World." *Journal of the History of Economic Thought* 42 (2): 203–228.
- Haack, Susan. 2005. "Trial and error: The Supreme Court's Philosophy of Science." *American Journal of Public Health* 95 (S1): S66–S73.
- Haack, Susan. 2014. *Evidence matters: science, proof, and the truth in the law*. New York: Cambridge University Press.
- Haack, Susan. 2015. "The expert witness: lessons from the US experience." *Humana Mente: Journal of Philosophical Studies* 28 (2015): 39–70.
- Hans, Valerie P, and Michael J Saks. 2018. "Improving judge & jury evaluation of scientific evidence." *Daedalus* 147 (4): 164–180.
- Haw Allensworth, Rebecca. 2012. "Adversarial economics in antitrust litigation: losing academic consensus in the battle of the experts." *Northwestern University Law Review* 106 (3): 1261–1306.
- Helland, Eric. 2019. "The role of ideology in judicial evaluations of experts." *The Journal of Law and Economics* 62 (4): 579–611.
- Hovenkamp, Herbert. 2005. *The antitrust enterprise: Principle and execution*. Cambridge, MA: Harvard University Press.
- Koehler, Jonathan J. 2018. "How trial judges should think about forensic science evidence." *Judicature* 102:28.
- Leamer, Edward E. 1983. "Let's take the con out of econometrics." *American Economic Review* 73 (1): 31–43.
- Lopatka, John E. 2016. "Economic Expert Evidence: The Understandable and the "Huh?"" *The Antitrust Bulletin* 61 (3): 434–460.
- Mangrum, Richard Collin. 2022. "Comment on the Proposed Revision of Federal Rule of Evidence 702: "Clarifying" the Court's Gatekeeping Responsibility over Expert Testimony." *Creighton L. Rev.* 56 (1): 97–108.
- Martini, Carlo. 2015. "The paradox of proof and scientific expertise." *Humana.Mente Journal of Philosophical Studies* 28:1–16.
- Moral-Benito, Enrique. 2015. "Model averaging in economics: An overview." *Journal of Economic Surveys* 29 (1): 46–75.
- Peruzzi, Edoardo. 2023. "Models on trial: antitrust experts face Daubert challenges." *Journal of Economic Methodology* 30 (4): 337–351.

- Peruzzi, Edoardo, and Gustavo Cevolani. 2022. "Defending De-idealization in Economic Modeling: A Case Study." *Philosophy of the Social Sciences* 52 (1-2): 25–52.
- Posner, Richard A. 1999. "The law and economics of the economic expert witness." *Journal of Economic Perspectives* 3 (2): 91–99.
- Rodrik, Dani. 2015. "Economics rules: the rights and wrongs of the dismal science." Oxford: Oxford University Press.
- Rubinfield, Daniel L., and Joe S. Cecil. 2018. "Scientists as experts serving the court." *Daedalus* 147 (4): 152–163.
- Schroeder, Thomas D. 2020. "Toward a More Apparent Approach to Considering the Admission of Expert Testimony." *Notre Dame Law Review* 95:2039–2062.
- Sidak, Gregory. 2013. "Court-appointed neutral economic experts." *Journal of Competition Law and Economics* 9 (2): 359–394.
- Steel, Mark F. J. 2020. "Model averaging and its use in economics." *Journal of Economic Literature* 58 (3): 644–719.
- Tirole, Jean. 1988. *The theory of industrial organization*. MIT press.