
C. Michael Holloway; NASA Langley Research Center, Hampton, Virginia, U.S.A.

Keywords: software system safety, software engineering, accident analysis, allegory

Abstract

This paper is a work of fiction, but it is fiction with a very real purpose: to stimulate careful thought and friendly discussion about some questions for which thought is often careless and discussion is often unfriendly. To accomplish this purpose, the paper creates a fictional legal case. The most important issue in this fictional case is whether certain proffered expert testimony about software engineering for safety-critical systems should be admitted. Resolving this issue requires deciding the extent to which current practices and research in software engineering, especially for safety-critical systems, can rightly be considered based on knowledge, rather than opinion.

Introduction

This paper takes the form of the fictitious legal case of Polly Ann Smith Co. v. Ned I. Ludd. The Polly Ann Smith Company does not exist. Ned I. Ludd does not exist. Equipment such as described here does not exist for small airplanes. The court system described in this paper does not exist. But the questions raised in, and by, this fictitious case do exist, and include the following:

- What is truly known about software system engineering, especially for safety-critical systems?
- Are there software engineering experts? If so, what are their qualifications?
- What constitutes proof of software engineering principles, tools, and techniques, especially when system safety is at risk?

This paper is not a legal commentary on how existing law applies, or should apply, to software. Although references to several real legal cases are used in the paper, any resemblance between the fictitious case presented here and particular cases actually adjudicated in the courts is coincidental.

To simplify the presentation, there are only two separate parties in this fictional case. Were something like what is described herein to happen in real life, it is highly likely that there would be more parties involved. This paper also ignores some issues, such as what would happen to the case after the decision, that would be important in a real case. Also, in the fictional legal system in which this case takes place, a company may be held liable for damages caused by one of its products if and only if the company is shown to be negligent in the design of the product. In most United States court systems today, a showing of negligence is not required; the existence of a defect is sufficient.

To avoid directly criticizing any existing approach to software system development or assessment, purposely nondescript approaches are invented. Any resemblance between these approaches and existing approaches is coincidental. Similarly, names used in this paper for people, courts, companies, and things are not generally intended to refer to real people, courts, companies, or things unless identified otherwise, either implicitly by context or explicitly in the references.

Finally, to avoid cluttering the text with many quotation marks, footnotes, or other reference indications, explanatory material, including the sources of excerpts and other material are not directly indicated in the main text. The Notes section at the end of the paper contains this information.

The remainder of this paper is divided into four main sections: a description of the accident that prompted the lawsuit, including background information about Ned I. Ludd and the Polly Ann Smith Company; a discussion of the lawsuit, including summaries of the decisions of the lower courts in the case; a summary of the opinion of the highest court, including a dissenting opinion; and closing remarks.
The Accident

The Players: Before the accident that gave rise to the lawsuit is described, some brief background information about both parties in the suit is given below.

Ned I. Ludd: Mr. Ludd was 56 years old at the time of the accident. He was once a highly successful corporate lawyer in Washington, D.C., but he abruptly quit his lucrative practice at the age of 45, and moved to a secluded cabin in the Blue Ridge Mountains near Roanoke, Virginia. He obtained his private pilot license, including an instrument rating, while he was still in law school. At the time of the accident, Mr. Ludd had logged over 2000 hours in a variety of small, private airplanes. Six months before the accident, he purchased one of the most advanced private airplanes available on the market. Mr. Ludd had logged 20 hours on his new plane before he set out on the journey that ended with the accident.

Polly Ann Smith Company: The company was founded in 1918 by Gregory A. Smith to make depth charges for use in the Great War. Polly Ann was Mr. Smith’s daughter. He named the company after her, because he hoped to help make the world safe, not only for democracy, but for Polly Ann, too. Over time the company expanded and became one of the largest privately held companies in the country. Although the company continued to produce military devices, its primary business for the past decade was supplying electronic equipment, including avionics, for small aircraft.

The President and Chief Executive Officer of the company at the time of Ludd’s accident was Anthony G. Hopson, Polly’s second child. Mr. Hopson was responsible for selecting the name ‘Amelia’ for the company’s newest autopilot. In selecting this name, which was the name of his elder daughter, he followed in the tradition of his grandfather. He also ignored the advice of many within the company, who suggested that the connotations of the name within the aviation community were not entirely positive. At the time of its certification, the Amelia system was the first autopilot certified on a private airplane to perform completely automated landings in extremely low visibility conditions.

The Event: After visiting his mother in Fredericksburg, Virginia, Ned I. Ludd flew his private plane back towards his home in southwestern Virginia. Ludd’s plane was equipped with the sophisticated autopilot system named ‘Amelia’ from the Polly Ann Smith Company. Ludd engaged the Amelia system shortly after taking off from Fredericksburg, and left the system engaged throughout the flight.

As Ludd approached his destination, the weather grew increasingly unpleasant. Nevertheless, all the relevant conditions remained within the parameters for which Amelia was certified to operate. Ludd opted to keep the system engaged throughout approach and landing. Conditions continued to deteriorate as the plane got nearer to the runway. The airplane impacted the ground 90 feet short of the runway threshold, bounced into the air slightly, and contacted the ground again about 20 feet from the runway. The plane remained on the ground after the second contact, slid about 500 feet on the ground and runway surface, and eventually came to a stop in an upright condition in the mud about 50 feet to the left of the runway.

Ludd survived the accident, but received multiple fractures, sprains, and lacerations. These injuries left him with partial disabilities of both legs and his left hand. The airplane was declared a complete loss.

The Investigation: After an investigation, the safety board determined that the probable cause of the accident was the pilot’s decision to continue an automated landing under the existing conditions. The board also noted that a significant contributing factor to the accident was erroneous output from the Amelia autopilot system. Detailed investigation by software engineers acting as consultants to the safety board discovered that the Amelia system contained erroneous software. This erroneous software, would, under the particular meteorological and geographic conditions present during the accident, cause inappropriate pitch commands to be sent to the aircraft control surfaces. These commands, in turn, if not overridden by the pilot, would usually cause the aircraft to impact the ground short of the runway. According to the board’s analysis, this is what happened in the accident. Ludd did not override Amelia during the landing, so the system flew the aircraft into the ground.
After learning of the safety board’s findings, Ned I. Ludd brought suit against the Polly Ann Smith Company (Smith for short), claiming that the company was negligent in the design and implementation of the Amelia system. In particular, Ludd claimed that Smith failed to apply certain software safety techniques to the design and verification of the Amelia software. Ludd alleged that these software safety principles represented the current state-of-the-practice for developing critical software, that Smith’s knowledge of these principles was deficient, and that records showed no application of these principles to the design of the Amelia system. Thus, according to Ludd, Smith was clearly negligent, and this negligence contributed to his injuries. That is, without Smith’s negligence in the design of the Amelia system, Ludd’s airplane would not have crashed, and Ludd would not have been injured.

For proof of his allegations, Ludd relied in significant part on the depositions of an internationally recognized software safety researcher, G. Clarke. At Ludd’s request, Clarke conducted an extensive review of the available documents pertaining to the development of Amelia. Based on this review, and on his knowledge of safety-critical software development, Clarke testified in his depositions that Smith showed no evidence of either understanding software safety practices, or of applying any of these practices to the Amelia software. Clarke also testified that, in his expert opinion, software safety practices were sufficiently well established as to constitute state-of-the-practice techniques. Clarke further testified that he believed that had Smith applied the appropriate software safety practices to Amelia, the flaws in the software that contributed to Ludd’s accident would have been discovered prior to deployment of the system.

In responding to the suit, Smith did not contest that its Amelia system contributed to Ludd’s accident, but it moved to exclude Clarke’s proffered testimony on two grounds. First, he did not qualify as an expert witness. Second, even if Clarke did qualify as an expert, his opinions on the deficiencies in the development of the Amelia software did not rise above “subjective belief or unsupported speculation.” Such opinions did not qualify for admission as expert testimony. Smith also alleged that Ludd himself was negligent in his piloting of the aircraft during the landing sequence, and that this negligence was primarily responsible for the accident. (Because this latter allegation played no part in the court decisions, it is not discussed any more.)

In support of its motion, Smith offered depositions from C. Vantile, an internationally recognized software researcher, and developer of widely used techniques for analyzing the correctness of software systems. In these depositions, Vantile disputed the claim that software safety practices were state-of-the-practice, and asserted that the true state-of-the-practice was represented by his own techniques, which Smith had used in developing the Amelia software. Vantile further asserted that the flaws present in the Amelia software were extremely difficult to detect; so difficult, in fact, that no one could have been reasonably expected to detect them.

The District Court granted Smith’s motion to exclude Clarke’s testimony, and entered summary judgment for Smith. In granting the motion, the Court ruled that although Clarke qualified as an expert witness, his proffered testimony was not allowable. In its published ruling, the Court cited Federal Rule of Evidence 702, and various decisions by the Supreme Court of the United States giving it the obligation to ensure that proposed expert testimony is both relevant and reliable. The Court said that Clarke’s testimony about the Amelia software was relevant, but not reliable; and thus, it must be excluded.

The Court based its finding of unreliability on three factors. First, Clarke cited no controlled experiments to support his contentions. Second, Clarke was unable to produce any logical proofs of the accuracy of any of his contentions. And third, a substantial number of software practitioners and researchers, including C. Vantile, disagreed with Clarke’s basic contentions, and favored different techniques and principles. According to the Court, the absence of empirical or logical proof, and the presence of conflicting theories rendered unreliable the basis of Clarke’s testimony.
Finally, the Court ruled that without Clarke’s testimony, Ludd had insufficient evidence to bring the suit. Thus it entered the summary judgment for Smith.

On appeal by Ludd, the Circuit Court reversed. It ruled that the District Court had abused its discretion in excluding Clarke’s testimony. According to the published opinion, Clarke’s international reputation, numerous published papers, and frequent consulting jobs firmly established his reputation as an expert in the field. By disallowing the testimony of a person with such credentials, the District Court went well beyond the authority allowed to it in such matters. Thus, the exclusion of Clarke’s testimony constituted a clear abuse of discretion.

Smith appealed to the Grand Court, which accepted the case.

The Ruling

This section provides a summary of the Grand Court’s ruling, excerpts from opinion of the Court, and excerpts from a dissenting opinion.

Summary: By a 6 to 3 majority, the Grand Court affirmed the judgment of the Circuit Court overruling the exclusion of G. Clarke’s testimony; however, the Court said that the Circuit Court erred in its rationale for reversal. Whereas the Circuit Court stated that Clarke’s stellar reputation alone was sufficient to warrant the inclusion of his testimony, the Grand Court said that the reputation of a proffered expert witness is simply one factor that a court may consider in determining whether the witness is qualified to testify as an expert. Qualifying a person as expert, however, is not the same thing as admitting particular testimony by that expert. In determining whether particular testimony may be admitted that a qualified expert witness proposes to offer, the court’s attention must be on the reliability of the data and methods on which the testimony is based, not on the reputation of the witness. Thus, the Circuit Court was looking at the wrong issue.

In considering the right issue --- whether Clarke’s particular testimony should have been admitted --- the Grand Court ruled that although the law assigns to the courts the role of ‘gatekeeper’ for expert testimony, but it does not assign to the courts the role of ‘arbiter’ between conflicting theories or methods in scientific or technical fields. Clarke and Vantile offer very different opinions about the proper ways to build safe software systems. Both may be wrong, but both cannot be right. In cases such as this, the only reasonable action a court can take is to admit the testimony from both sides, and allow the jury to decide on relative credibility.

For these reasons, the Court remanded the case to the District Court for further proceedings consistent with the opinion. Justice Watson delivered the opinion of the Court, in which Justices Harper, Hall, Mason, Brown, and Williams joined. Justice Blake filed a dissenting opinion, in which Justices Randle and Collins joined.

Opinion of the Court: This section contains excerpts from Justice Watson’s opinion for the court. All of the words in the remainder of the section are direct quotes from the opinion.

The Circuit court was correct in applying the ‘abuse of discretion’ standard of review to the rulings of the District Court. Both lower courts were correct in applying Federal Rule of Evidence 702 to the case, also.

For Rule 702 to apply there must exist “scientific, technical, or other specialized knowledge” that “will assist the trier of fact to understand the evidence or to determine a fact in issue.” Understanding Ludd’s allegations of negligence on the part of Smith in developing the Amelia system requires understanding Ludd’s detailed about software development practices, details that cannot be understood without the aid of “scientific, technical, or other specialized knowledge.” Understanding Smith’s defense against these allegations requires similar knowledge. This knowledge is well outside the normal body of knowledge that can be expected to be possessed by average citizens, or even average judges. Thus it comes as no surprise
to this court that both sides in the dispute offered expert witnesses: G. Clarke (for Ludd) and C. Vantile (for Smith).

Faced with the proffer of these two expert witnesses, the District Court was required to determine first whether each of the witnesses “qualified as an expert by knowledge, skill, experience, training, or education.” Ludd did not challenge the qualifications of Vantile, but Smith challenged the qualifications of Clarke. The District Court correctly rejected the challenge to Clarke, and qualified both individuals as expert witnesses about safety-critical software development practices. As that court noted, Clarke and Vantile possessed similar credentials: doctorate degrees from prestigious institutions, over 75 published papers in refereed journals and conferences, membership in appropriate professional societies, completion of numerous government contracts, invited participation in advisory boards, and extensive practical experience as consultants to several organizations that develop software, including safety-critical software. It is only slightly hyperbolic to say that, if Clarke and Vantile are not experts in the field, there are no experts in the field.

If the District Court failed to qualify either person as an expert, it would have clearly abused its discretion. But the Court did not exclude Clarke or Vantile as experts; instead, it excluded Clarke’s specific testimony as being neither “based upon sufficient facts or data,” nor “the product of reliable principles and methods.” We must determine whether that exclusion was an abuse of discretion.

We find that it was.

A trial court has wide discretion in determining how to test the reliability of the principles and methods upon which expert testimony is based; however, this “is not discretion to perform the function inadequately. Rather, it is discretion to choose among reasonable means of excluding expertise that is fitusse and science that is junky.” The means chosen by the District Court was not reasonable, however, as we now show.

The District Court cited three factors for finding that the principles and methods upon which Clarke’s testimony was based were unreliable: lack of controlled experiments, lack of logical proofs, and existence of conflicting principles and methods. The court was basically correct in stating that these three factors exist; it was incorrect in inferring that these three factors constitute sufficient proof of unreliability.

We can think of several different ways to show that these three factors cannot constitute sufficient grounds for excluding expert testimony. Perhaps the way most likely to persuade many judges (although not our dissenting colleagues) is to point out that these same three factors apply just as well to the law as they do to software development. Few, if any, controlled experiments support legal rulings. Few, if any, logical proofs are offered, either. And certainly, there exist conflicting principles and methods, as the lack of unanimity in this decision attests. Nevertheless, none of us (including our dissenting colleagues) are inclined to determine from these facts that there exists no legitimate expert testimony about legal matters.

Of course, the real problem with the District Court’s determination is not that it may be used against judges, but that it simply presumes too much. It presumes too much in asserting that there exist only two means of showing reliability: controlled experiments and logical proof. There are other means, such as case studies, quasi-experiments, and rigorous (although not strictly formally sound) reasoning. Clarke’s testimony cited examples of the use of these methods to support his contentions.

The ruling also presumes too much in assuming that a court is an appropriate arbiter between conflicting theories in technical fields. By citing the third factor --- existence of conflicting principles and methods --- the District Court implies, without quite ever exactly saying so, that these conflicting principles and methods (specifically, Vantile’s principles and methods) are more credible than Clarke’s. The court has determined what the professional community of which Vantile and Clarke are members has not been able to determine: namely, that one of these internationally acclaimed men is wrong (Clarke), and the other is right (Vantile). To make this determination, the court has become the ‘amateur scientists’ against which
we were warned. This is not only abuse of discretion; it is arrogance of the highest (or should that be, lowest) order.

The Circuit Court was right to overturn the District Court’s ruling, but it was wrong in its reasoning. We affirm the judgment, but remand the case to the District Court for further proceedings consistent with this opinion.

Dissenting Opinion: This section contains excerpts from Justice Blake’s dissent. All of the words in the remainder of the section are direct quotes from the dissenting opinion.

I agree with much of what is said in the opinion of the Court. Certainly, abuse of discretion is the right standard of review. If I agreed with the Court’s assertion that Rule 702 applies in this case, I probably would concur with the analysis presented by my distinguished colleagues. However, the most beautiful edifice will not long stand if it is built upon a foundation of quicksand; the Court’s analysis cannot stand, because it is built on nothing more substantial than quicksand.

The foundation is quicksand because Rule 702 simply does not apply. It does not apply because the precondition for its application is not satisfied. The Court correctly identifies this precondition: “For Rule 702 to apply there must exist ‘scientific, technical, or other specialized knowledge’ that ‘will assist the trier of fact to understand the evidence or to determine a fact in issue.’” The Court believes that such knowledge about software development practices exists in this case. I believe the Court has confused ‘knowledge’ with ‘opinion’. The two do not mean the same thing, as any dictionary will show.

The primary definition of ‘knowledge’ is “the fact or state of knowing,” where ‘knowing’ comes from ‘know’, which means primarily “to regard as true beyond doubt.” In contrast, ‘opinion’ means primarily “a belief or conclusion held with confidence but not substantiated by positive knowledge or proof.”

There is no question that a substantial body of literature exists about every aspect of software engineering, including the description of various methods for ensuring the safety of software used in critical applications. G. Clarke and C. Vantile have made numerous contributions to this literature. However, when reviewing representative publications from this body of literature, including papers by both Clarke and Vantile, one cannot help but be struck by the extent to which these publications contain little more than “belief[s] or conclusion[s] held with confidence but not substantiated by positive knowledge or proof” (that is, opinion).

Even more striking is the way in which quite a few people, Clarke and Vantile included, appear to start off their publishing careers acknowledging the basic lack of knowledge in the field. As time goes on, these people, Clarke and Vantile included, make increasingly dogmatic statements, without any increase in the quantity or quality of the evidence given to support these statements. The clear impression is that many people, Clarke and Vantile included, come to deceive themselves into believing they have knowledge, when all they really have is opinion.

To put it bluntly, to assert that there exists software engineering ‘knowledge’ is to strip the word ‘knowledge’ of any distinction from mere opinion. This I am not willing to do.

The Court writes, “It is only slightly hyperbolic to say that, if Clarke and Vantile are not experts in the field, there are no experts in the field.” Under the meaning of ‘expert’ in Rule 702, neither Clarke nor Vantile nor anyone else in the field is a software engineering expert, because there is no “scientific, technical, or other specialized knowledge” in software engineering in which to be an expert. It is only slightly hyperbolic to say that in software engineering, all expertise is fausse and all science is junky. Perhaps one day there will exist ‘knowledge’ in software engineering, but that day is not today.

The Court’s analogy between the software engineering and legal fields is intriguing, but not compelling. It is not compelling, because there exists a significant difference between the two fields. For legal matters, there is substantial agreement about certain principles that are “regarded as true beyond doubt.” For
example, much disagreement exists about what to do about murderers, but nearly universal agreement exists that murder is wrong (at least in this country). Thus, in the legal field, unlike in the software engineering field, there is an adequate basis for believing that knowledge exists. Exactly how much knowledge exists may be a matter of, well, opinion, but that some exists is not.

For the reasons given above, I respectfully dissent. The District Court did not abuse its discretion when it disallowed Clarke’s testimony. I would reverse the ruling of the Circuit Court. I would also note that the issue of whether to allow Vantile’s testimony became mute when the District Court ruled that, with Clarke’s testimony excluded, Ludd did not have a case. Had the issue not been mute, Vantile’s testimony should also have been excluded, for the same reason as Clarke’s testimony should be excluded.

Concluding Remarks

The most important issue in the fictional legal case described above was whether certain proffered expert testimony about software engineering for safety-critical systems should be admitted at trial. Resolving this issue required the courts to decide the extent to which practices and research in software engineering, especially for safety-critical systems, could rightly be considered based on knowledge, rather than mere opinion. Making this decision is just as important in the real world as it was in the fictional world. Everyone involved with safety-critical, software-intensive systems, whether as developers, researchers, or users, can benefit from thinking carefully about the issue, and the questions raised by it:

- What is truly known about software system engineering, especially for safety-critical systems?
- Are there software engineering experts? If so, what are their qualifications?
- What constitutes proof of software engineering principles, tools, and techniques, especially when system safety is at risk?

Perhaps, if we think carefully about these questions, we will all be a little less prone to put blind faith in automation, and a little less prone to jump on the bandwagon of the latest software engineering fad. Perhaps, too, we will be a little more prone to seek to show what is good about our favorite methods, instead of seeking to show what is bad about everyone else’s.

Notes

Ned I. Ludd does not exist: Most people are probably familiar with the term ‘luddites.’ The American Heritage Dictionary of the English Language (ref. 1) says concerning the origin of the term: “After Ned Ludd, an English laborer who was supposed to have destroyed weaving machinery around 1779.”

Equipment such as described here does not exist: There are no currently certified autopilot systems for private aircraft that are capable of landing an airplane in the conditions described here.

showing of negligence is not required: See, for example, “Proving Fault for Defective Product Injuries” at http://www.nolo.com/.

make the world safe, not only for democracy: When the United States entered the war in 1917, President Woodrow Wilson described the war as a war to “make the world safe for democracy.” See http://www.whitehouse.gov/history/presidents/ww28.html.

connotations … not entirely positive: Because Amelia Earhart and her navigator, Fred Noonan, went off their intended course and lost their way over the Pacific. Going off course is not usually something with which one wants to associate an autopilot.

did not rise above … speculation: The wording here is from the Syllabus in General Electric Co. v. Joiner (ref. 2).
Federal Rule of Evidence 702: The Federal Rule of Evidence 702 was amended in December 2000 (ref. 3), primarily to address directly the issues addressed in the “various decisions” below. It now includes specific tests to be applied in determining the admissibility of expert testimony.

various decisions of the Supreme Court: See references 2, 4, and 5 for the actual rulings, and reference 6 for a discussion about the first of these rulings (Daubert).

both relevant and reliable: In this context, ‘reliable’ means simply “that which can be relied upon” (ref. 6); no quantitative assessment (like what may be done for hardware reliability predictions) is implied.

the Grand Court: This court in our fictitious world is equivalent to the Supreme Court of the United States in the real world.

gatekeeper: The Supreme Court of the United States introduced this term in Daubert (ref. 5).

both ... be right: Inspired by Abraham Lincoln’s statement: “In great contests each party claims to act in accordance with the will of God. Both may be, and one must be, wrong.” (ref. 7)

direct quotes from the opinion: Omitted from these quotations are most of the citations of the legal authorities.

is not discretion to ... science that is junky: These words are from Justice Scalia’s concurring opinion in Kumho Tire (ref. 4).

correct in stating that these three factors exist: Many people have commented on the general lack of empirical or rational foundation for most of software engineering. Reference 8 is one such commentary.

the ‘amateur scientists’ against which we were warned: See reference 5, Chief Justice Rehnquist, concurring in part and dissenting in part.

direct quotes from the dissenting opinion: As with the opinion of the Court, the citations of the legal authorities are generally omitted.

any dictionary will show: All definitions that follow are from reference 1.

come to deceive themselves: See reference 9 for an interesting discussion of philosophical issues involved in self-deception.

Concluding Remarks: In this section, we leave the fictional world in which Polly Ann Smith Co. v. Ned I. Ludd takes place, and return to the real world.

Perhaps … Perhaps, too … : Then again, perhaps I am still living in a fictional world.

References


**Biography**

C. Michael Holloway, Senior Research Engineer, NASA Langley Research Center, MS 130 / 100 NASA Road / Hampton VA 23681-2199, USA, telephone - +1.757.864.1701, facsimile - +1.757.864.4234, e-mail - c.m.holloway@larc.nasa.gov, web – [http://shemesh.larc.nasa.gov/](http://shemesh.larc.nasa.gov/)

C. Michael Holloway has been at NASA Langley since 1983, and a member of the formal methods team since 1992. His professional interests include accident analysis, software system safety, and foundations for high-integrity software development techniques. He is currently leading two research projects. Causality Analysis Using Symbolic Expression (CAUSE) is investigating the use of formal notations in accident and incident analysis and reporting. Formal-Enough Notations for Computer-system Engineering (FENCE) is exploring the integration of natural, formal, and graphical languages for the elicitation and propagation of requirements for software-intensive systems. He was graduated from the School of Engineering and Applied Science at the University of Virginia with a B.S. in computer science in 1983, and completed all-but-dissertation towards a Ph.D. in computer science at the University of Illinois in Champaign-Urbana. He is a member of the IEEE, the IEEE Computer Society, and the System Safety Society. Mr. Holloway is married and has two children.