

Chapter 2

Credibility Assessment, Common Law Trials and Fuzzy Logic

Gerald T.G. Seniuk

Judges or juries make decisions about the credibility of witnesses, decisions that might send one person to prison for years, strip another of her fortune or deny a parent full access to his children. An on-going judicial research project has been studying how such questions of contested fact are determined in a trial (Seniuk, 1994). The project reached out to experts from outside the legal profession to assess what knowledge or insight these other disciplines might shed on this question. For example, knowledge of forensic psychology and what the discipline has learned of credibility assessment and lie detection has greatly assisted this project (see Seniuk & Yuille, 1996; ten Brinke & Porter, present volume).

The purpose of this chapter is to reflect on that knowledge exchange and, as such, the focus here is on that type of trial where key witnesses disagree under oath about the essential facts of the case. These are not cases deductively reasoned toward legal principles or public policy positions. What are examined here are those kinds of cases where the trier must determine if the factual elements of a case have been proven. Furthermore, these are cases where there is no independent evidence that determines that question of fact. Although other pieces of evidence are considered in making the ultimate decision, in such cases, the fact finder, either judge or jury, ultimately makes the decision by relying on one witness over the other.

Although the forensic psychologists and the judges in the on-going judicial research project (Seniuk, 1994) considered various aspects of this question, one example is sufficient to illustrate my conclusion, which is that, while much can be learned by such exchanges, the essential issues are not resolved. Instead, new insights raise new issues for the trial process. The example I would choose for this purpose is the use of demeanour evidence. There has been awareness within the legal profession of the frailties of relying on demeanour evidence but, in my

G.T.G. Seniuk (✉)
Provincial Court of Saskatchewan,
Visiting Scholar, College of Law, University of Saskatchewan.
Saskatoon, SK, Canada
e-mail: seniuk@sasktel.net

experience, that was a dim awareness that resulted primarily in an anxious, impotent wringing of hands. For example, it was argued many years ago that such fact finding amounted to nothing better than guesses (Frank, 1949), but the practice of fact finding continued as before. Generations ago, a leading Canadian jurist, Chief Justice O'Halloran, repeatedly warned about the dangers of making decisions on the basis of which witness was believed. Phrases from his decisions such as, "The judge is not given a divine insight into the hearts and minds of witnesses appearing before him. Justice does not descend automatically upon the best actor in the witness box." (*R. v. Presley*, 1948), are quoted in decisions to this day, but the call he made for general reform more than 50 years ago did not lead to that.

Up until recent years, judges generally continued to explain their reliance on a witness because of "the demeanour of the witnesses" or because "the testimony had the ring of truth to it." However, it now appears that is changing, and such explanations for findings of fact are becoming rarer. Although there were changes in the law that also led to this growing concern with demeanour evidence, I think the change is at least in part due to the work of the judicial research project and the kind of credible scientific information forensic psychologists have brought into that discussion (see Seniuk and Yuille, 1996). One example that leads me to this observation is the use within judicial circles of a quotation that I first encountered only through the input of psychologists. Although the following idea by Ekman (1992) may have been well known in psychology circles, it was new to many of us in the 1980s:

"It is amazing to many people when they learn that all of the other professional groups concerned with lying—judges, trial attorneys, police, polygraphers who work for the CIA, FBI or NSA (National Security Agency), the military services, and psychiatrists who do forensic work—did no better than chance. Equally astonishing, most of them didn't know they could not detect deceit from demeanour" (p. 285).

Presenting that quote and idea in judicial project workshops in those years did, in fact, bring amazement to many. It may have been saying essentially the same thing that Jerome Frank and Chief Justice O'Halloran had been cautioning everyone about decades before, but this warning had the rigor of science to back it up, and it commanded attention. Over the years, the quote began to surface in other workshops or presentations, and eventually has been joined to a list of other similar scientific reports, adding more emphasis to the message and gaining wider circulation within the legal community. The muted response to the jurisprudential cautions has now become conventional wisdom within the profession, thanks in part, to the work of the project.

Recent developments in the legal universe, such as changes to rules on corroboration (see Seniuk, 1992, for a summary of the history of these changes) and new theories in evidence scholarship (Allen, 1994a, 1994b) were also amplifying the previously muted concerns raised by legal writers such as Frank and O'Halloran. However, the ability to now point to scientific research, to be guided by quantified conclusions rather than to rhetorical warnings, strengthened the arguments of those, who for decades, previously might have been whistling in the jurisprudential wind.

Of course, psychologists did more than demonstrate the dangers of relying on demeanour evidence. They have also made advances in credibility assessment techniques (see Colwell, Hiscock-Anisman, & Fede, present volume; Griesel, Ternes, Schraml, Cooper, & Yuille, present volume; Hancock & Woodworth, present volume;

O'Sullivan, present volume; ten Brinke & Porter, present volume) which the North American courts have not been able to incorporate into their decision making on credulity matters. One problem is that these new techniques require specialized training (see Yarbrough, Hervé, & Harms, present volume). But even if a fact finder were trained in these new techniques, the new techniques are not recognized in law as something appropriate for application by a judge or a jury. The law looks to common human abilities as the primary intellectual tool fact finders should use to assess credibility. The law is wary of specialized knowledge (*R. v. Belland and Phillips*, 1987), especially newer scientific knowledge that has not been universally adopted by the particular discipline and proven over time. Thus, if a fact finder in a trial presumed to apply these techniques, lawyers on the losing side may very well use that as a ground of appeal. The law jealously guards this human function of the fact finder and is wary of allowing expert evidence that may supplant the fact finders role in determining the truthfulness of a witness.

In addition, the new techniques are geared more for investigative or interview processes (see Yarbrough et al., present volume) rather than the formalized and restricted courtroom environment. To the extent that they may have an application in the courtroom, it is the cross-examining lawyer who could use these skills in assessing which line of questioning to pursue with a witness. Judges and jurors are assigned a more passive, listening role. Finally, there is the question of the degree of reliability achievable by any of the current techniques. It may be that new technologies on credibility assessment might reflect the kind of certainty and reliability that we have experienced with DNA evidence, which is relied upon in the courts. But so far, advances in credibility assessment techniques do not make claims to that type of measurement of reliability. Although credibility assessment techniques do report improved accuracy above the 50% level (e.g., Ekman, 1992; Colwell et al., present volume; ten Brinke & Porter, present volume), the level achievable by chance or via the use of demeanour evidence, they do not claim to attempt to achieve levels of 90%, the threshold level many would assign to proof beyond reasonable doubt. However, even if such techniques were successful over 90% of the time, that kind of statistical, frequency measurement of accuracy is not useable by courts in assessing the credibility of a particular witness. Frequencies and statistical or subjective estimates of probabilities are not used in trials to establish the reliability of any particular witness.

The example of demeanour evidence is intended to demonstrate that, while the evolving knowledge in psychology can help myth-busting in the trial, those advances do not resolve the main issue in question. Although no longer using the lens of demeanour evidence alone, judges and jurors are still making decisions about whether a witness is truthful. We have made important gains in removing that error but that gain alone has not resolved the fundamental problem. If anything, it has raised more uncertainty. Removing that faulty lens helped us realize we were seeing things that were not there, namely reliable indicia of credibility, but our vision without the lens is not made more crisp or focused. If anything, everything is fuzzier now. As a result, we have brought a new issue into greater focus—the issue of indeterminism in the process generally. This can be seen as Appellate Courts review Trial Court decisions on credibility. In Canada, there has been a history of some Provincial Appellate Courts seeking more deterministic explanations by trial judges

of their findings on credibility assessment. Generally these attempts by Provincial Appellate Courts have been reversed by the Supreme Court of Canada (see Allen & Seniuk, 1997, for a review of some of these cases in the latter half of the last; see *R. v. J.H.S.*, 2008, for a current example).

As we continue to grapple with this issue of indeterminacy, there are likely other new developments that could assist in our understanding of the interplay of credibility assessment with other aspects of the trial process—for example, the relatively new lens of *fuzzy logic*. Dr. Lotfi Zadeh, from the University of California, is the originator and leader of fuzzy logic and, his colleague, Dr. Madan Gupta, from the University of Saskatchewan, is an international leader in the field of neural networks and fuzzy systems. In their explanations of fuzzy logic, two important points are made right off the start. Fuzzy logic is not fuzzy but rather is a precise way to deal with imprecision (e.g., Ding & Gupta, 2000; Kaufmann & Gupta, 1985; Zadeh, 2004). Fuzzy logic has been and still is to a lesser degree an object of controversy, but one that is gaining more attention.¹

Although it is precise, fuzzy logic is very different from traditional logic, and traditional logic is the basis of judicial reasoning. Traditional logic is bivalent, implying that every proposition is true or false with no degrees of truth allowed. This logic would seem to fit well in criminal law where you are either *guilty* or *not guilty*, and you are always *not guilty* unless the prosecution has proven beyond a reasonable doubt that you are guilty. In fuzzy logic everything is, or is allowed to be, partial (a matter of degree), imprecise (approximate), granular (linguistic) and perception based. Kaufmann and Gupta (1985) describe the distinction by using the example of illumination from a light bulb and a light switch. Bivalent logic is like having a light switch that you turn on or off. There is either zero illumination from the light, or full illumination. In law, you are either guilty or not guilty. Fuzzy logic is akin to using a dimmer switch to turn the illumination up or down, and there are degrees of illumination from zero to one. In between, there can be 0.2, or 0.4 or 0.7 or any degree of illumination between. In addition, we can use natural language to describe the degrees without requiring mathematical precision. The lighting may be very dim, somewhat dim, just right or too bright. In law, with the exception of the unique Scottish additional verdict of *not proven*, the switch is either on or off, *guilty* or *not guilty*, and there are no such degrees. Could there be? Should there be? By using this new lens of fuzzy logic as a framework for analysis, we can explore that question. Legal evidence scholarship has already begun looking at related questions.

In evidence scholarship, there has been recognition of insights similar to those of fuzzy logic, although legal scholarship has certainly not embraced fuzzy logic. However, there is recognition in evidence scholarship that fact finding and the logic of proof is not as crisp and precise as formalized legal reasoning makes it appear (Allen & Seniuk, 1997). From that perspective, the conventional view of legal proof

¹ Count of papers containing “fuzzy” in the title as compiled by Engineering Library, UC Berkley to October 2005 from INSPEC databases: 1970–1979=569; 1980–1989=2,404; 1990–1999=23,211; October 2000. 2005=17,785.

that focuses on the elemental structure of liability is replaced by a holistic view, from deciding the truth or falsity of particular elements to deciding the relative plausibility of opposing stories. This is analogous to the earlier shift in scientific thinking from the view that science is embarked on a march toward truth (i.e., just as trials are seen as a search for truth; *R. v. B. [K.G.]*, 1993), to the view that “progress” is measured by the articulation of better theories, where “better theories” means “better than the available alternatives” (Allen & Seniuk, 1997). The relative plausibility theory of evidence recognizes that evidence in a trial is not invested with only two probabilities—1.0 and 0.0—but rather views those as the end points of an infinite range of possibilities. One can hear evidence, not believe it to 1.0 probability, and still be influenced by it (Allen & Seniuk, 1997; see also *R. v. Mackenzie*, 1993, for a distinction between “facts” and “evidence of facts”). Thus, despite the conventional theories of legal adjudication which has a crisp framework, there is recognition that both the reality of the world and the trial process are fuzzy. While there is a growing recognition of this fuzziness, we lack an accepted method to reason in this uncertain mode, and that is where fuzzy logic might help.

Fuzzy logic is aimed at a formalization of modes of reasoning which are approximate rather than exact. As Zadeh (2004) explains, in the exact mode, we reason that “all men are mortal, Socrates is a man, and therefore Socrates is mortal.” In approximate mode, we reason that “most Swedes are tall, Magnus is a Swede, and therefore it is likely that Magnus is tall.” Criminal law forbids that kind of reasoning in the fact finding process, and instead insists that the question is whether Magnus is tall regardless of the incidence of tallness among other Swedes. The fundamental legal question that needs to be addressed is what use, if any, can be made of the generalization about most Swedes?

Related questions in law were the subject of debate 20–30 years ago (e.g., Anderson & Twining, 1991). For example, say your degree of proof was 97% (note, however, that proof beyond reasonable doubt is not a matter of probabilities) and you knew that 98 of a group of 100 people each stole an apple; could you just convict any one of those 100 people of theft since the probability of you being right is 98%? I have not seen a similar degree of interest about such questions in legal literature in recent years. The initial flurry of discussion generated insight into the complexity of the fact-finding process, but it never provided any practical applications. Judges were still reliant on the basic tests established by legal precedent over the decades; and there is little or no guidance in legal precedent, other than the standard of reasonable doubt, as to appropriate modes of reasoning where the law allows fact finders the discretion to convict under uncertain and ambiguous conditions. Because of the developments outlined above concerning demeanour evidence, there is now great pressure on trial judges to demonstrate the path of their reasoning to a conclusion about facts that are in dispute, facts that are often resolved by reliance on a witness telling the truth. However, while there is this growing expectation to demonstrate highly accurate findings of fact, there is precious little in the way of criteria or framework to guide a fact finder in the work of concretely demonstrating the correctness of such findings when they are based on the trust placed in a witness.

This is where I think fuzzy logic can help because it is developing concepts that offer an approximation to reality—the reality of pervasive imprecision, uncertainty and partiality of truth. Those are exactly the problems that fact finders encounter in trials. There would mostly be no need for a trial if there were precision, certainty and full truth. Fuzzy logic may not determine whether Witness A or Witness B, or either, is telling the truth, but it might provide a framework for discussion about the degrees of reliability of trust that a fact finder has in a witness. I suspect it may have much to say about the systemic application of a most fuzzy concept—reasonable doubt.

The reasonable doubt standard is a single, objective and exacting standard of proof. It is not the same as a proof of probability, and it is not like subjective standards of care that we apply in important everyday situations. It is not proof to an absolute certainty. It is not proof beyond *any* doubt, nor is it an imaginary or frivolous doubt. It is based on reason and common sense, and not on sympathy or prejudice (*R. v. Lifchus*, 1997). Proof beyond reasonable doubt falls much closer to absolute certainty than to proof on a balance of probabilities (*R. v. Starr*, 2001).

It is especially at this level, at the systems level, that fuzzy logic may enrich our dialogue and study. For example, given the fuzziness of the application of a standard like *reasonable doubt*, is it not highly likely that different fact finders will end up at different degree points along the range of 0.0–1.0? Given such vagaries among fact finders, the outcome of a trial may vary according to the particular fact finder. This systemic indeterminacy is recognized (Polya, 1988), but not fully confronted as an issue.

In this connection, the insights provided by psychologists in the judicial workshops (Seniuk & Yuille, 1996) become even more important than we realized at the beginning of our project. We can see now that credibility assessment of witnesses is not simply a matter of technique and training. Given the complexity of assessing credibility, and given the legal discretion afforded fact finders in those determinations, that exercise in itself will be fuzzy in many cases. Add to that the fuzziness of the application of the standard of proof of reasonable doubt, and the systemic fuzziness in the common law trial may be significant. This is where fuzzy logic analysis provides a new lens through which to consider the implications of this indeterminacy. The following charts are typical of fuzzy logic analysis. Figure 2.1 shows an example of student grades as set out on a crisp, traditional chart. We define the range of marks, 0–59, for a student we rate as *poor*, 60–84 for *good* and 85–100 for *excellent*. The lines are fixed. If you have a score of 58, you are given a *poor* rating, despite the fact that your colleague, who only got 2 more marks than you, is rated as *good*. That student gets the same *good* rating as the student who is almost *excellent* at the 84 level.

Greater precision can be provided through the exact grade, but when the rating of students is at the verbal level, such as poor or excellent, that precision is lost. The ratings in a criminal trial are almost all at the imprecise verbal level—you are either guilty or not guilty. Although criminal proof is not a matter of probabilities (*R. v. Lifchus*, 1997), commentators often define proof beyond a reasonable doubt as proof above 95%, but not necessarily absolute certainty (see Fig. 2.2). The civil standard of proof is proof on a balance of probabilities, which is taken to mean theasserter's case is proven to above 50% (see Fig. 2.3).

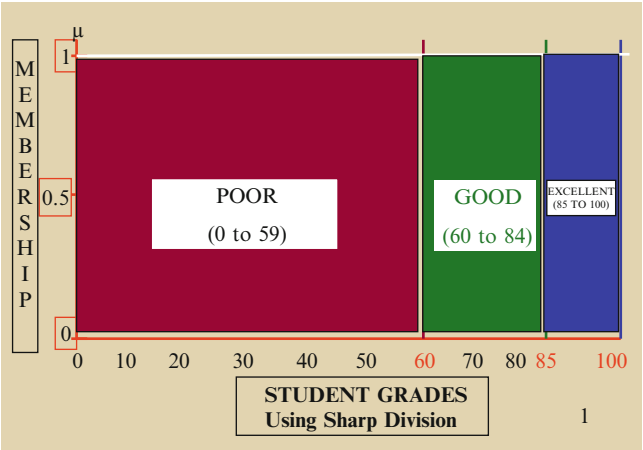


Fig. 2.1 Example of fuzzy logic analysis using sharp division for student grades

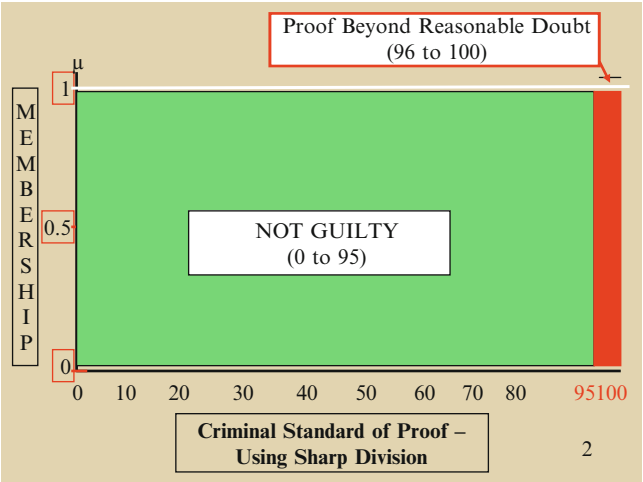


Fig. 2.2 Example of fuzzy logic analysis using sharp division for criminal standard of proof

Such comparisons to probabilities are misleading when discussing human cognitive actions such as determining guilt or liability. Probabilities have to do with physical events. For example, drawing one card randomly from a full deck allows one to state the probability of drawing any particular card. But assessments of evidence and the making of judgments are human, cognitive actions. So, proof beyond reasonable doubt is not proof between 95 and 100%—it is simply a very high degree of proof. That is a fuzzy concept.

By either standard of proof, once the determination of proof is made, a crisp line is drawn. You are either guilty or not guilty. You are either civilly liable or you are not liable. But such crispness must be illusory in some cases because there is a

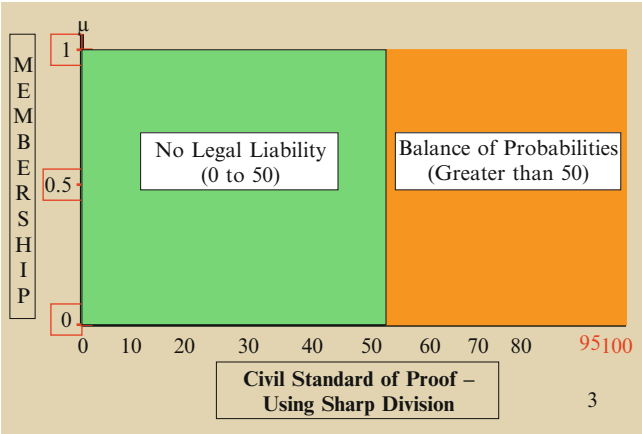


Fig. 2.3 Example of fuzzy logic analysis using sharp division for civil standard of proof

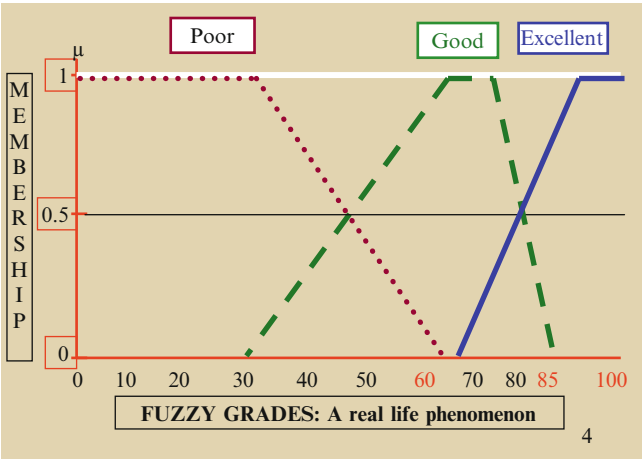


Fig. 2.4 Example of fuzzy logic analysis without sharp division for student grades

subjective element in making these determinations of criminal guilt or civil liability. In addition, such determinations are affected by the *context* of the case and the evidence. And as indicated earlier, modern evidence scholarship acknowledges the cognitive role played by the *experience and predilections* of a particular fact finder. Whenever we say that reasonable people might disagree with the determination, we are conceding that there is fuzziness about it.

Considerations such as these are all matters that fuzzy logic tries to come to grips with. In the example of the student grade, instead of drawing a crisp line, fuzzy logic would try to capture the slow, smooth and gradual progression within the sets of *poor*, *good* and *excellent*. In Fig. 2.4, we can see how one such graph might look, although the choice of graph style is itself subjective and fuzzy. The key concept in the fuzzy logic chart is that the categories of *poor*, *good* and *excellent* can overlap.

You can be both a *poor* and a *good* student at the same time. Instead of drawing crisp lines at “60” and “85,” fuzzy logic treats the three categories as sets with ranges of membership in each.

The way to interpret such a graph is to view the green line with long dashes for *good* students as indicating that anyone at a mark of 60 and over has full membership of 1.0 in the *good* category. But, in addition, it tells us that those on the upward slope who would have been excluded from the *good* category in a crisp line are seen to be members of the set to varying degrees. For example, someone with a mark of 40 is a member of the *good* set to a degree of about 0.25. At the same time, someone on the downward slope of the green-dashed line is still a full member of the *good* category to a degree of 1.0, but they are also now members of the *excellent* set to varying degrees. For example, someone on the green-dashed line with a mark of 80 is totally a *good* student, but that student is also a member of the set of *excellent* students to a lesser degree, by about 0.5 on this graph. And while someone with a mark of 25 on their exam is fully within the set of *poor* students to a degree of 1.0, someone with a mark of 58 would only be a member of the *poor* student set to a degree of about 0.1 and, more importantly, would now be considered a member of the set of *good* students to a degree of 0.9.

Judges are experienced in dealing with such fuzzy sets when it comes to matters such as sentencing. For the same charge, different offenders might receive significantly different sentences (see Fig. 2.5). That is because, as fuzzy logic recognizes, each case is contextual and unique to its circumstances and experience. The offenders and their circumstances may differ, and so too might the circumstances or the harm intended or caused of the particular offence. And not unlike the example of student grades, sentencing is also exact—18 months for example—but people might differ over whether such a sentence was in the *light* or *moderate* set.

Other areas of judicial adjudication, however, are fuzzy without the underlying crispness that specific sentences provide—for example, the more fundamental issue of whether someone committed the offence at all, an issue that relates to credibility. Although the legal definition of the offence will be a crisp definition, the necessary finding of facts, often based on the credibility of a witness, or the matching of particular facts to the definition can be fuzzy. In actual practice, we, of course, make crisp conclusions—either guilty or not guilty—and make no use of fuzzy logic’s insight of gradation and partial membership. Fuzzy logic reveals the uncertainty behind our picture of crispness and certainty, and it also provides a framework within which we can explore issues of uncertainty in a more precise manner. How far could one go in applying this framework?

It would be unsafe and unacceptable to recognize gradations or degrees of guilt, as in Fig. 2.6, for example, when considering whether someone was guilty of an offence beyond a reasonable doubt. That would be a slippery slope that is fraught with danger. Before we dare to allow the state to punish citizens, centuries of experience across different civilizations and continents has shown the necessity of maintaining the strict application of burdens of proof and high standards of crisp proof.

At the same time, real trial experience encounters the public frustration with such safe and crisp outcomes in the face of fuzzy and festering human and communal

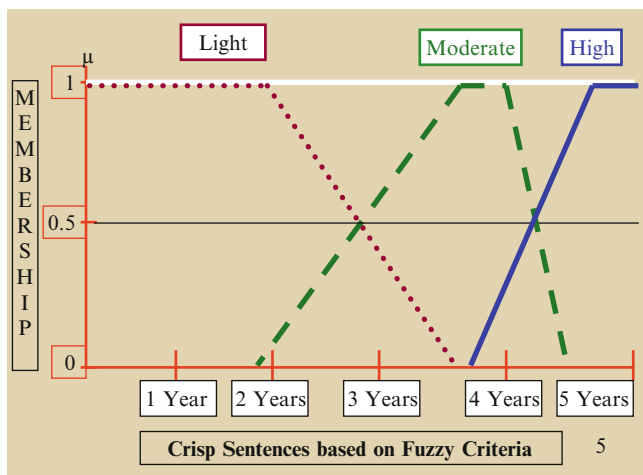


Fig. 2.5 Example of fuzzy logic analysis for criminal sentencing: crisp sentences for fuzzy criteria

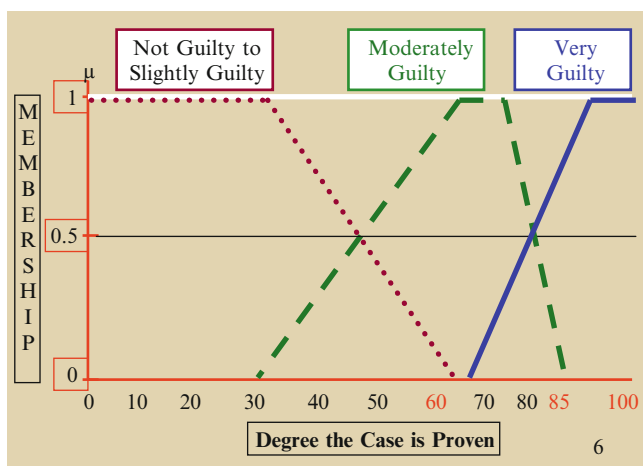


Fig. 2.6 Example of fuzzy logic analysis for degree of proof and guilt decision

relationship problems. Should there or could there be a finding other than “guilty” or “not guilty?” Could there be a category, for now call it “finding X”, that somehow revealed the fuzziness and the uncertainty of exactly who did what, but yet recognized a community’s ability, short of punishment, to respond to this “finding X,” as in Fig. 2.7?

Could the prosecution have proven its case to a degree of 0.7 and the accused raised a reasonable doubt to a degree 0.3? We are able to deal in such gradations once the threshold has been crossed and the accused proven guilty beyond a reasonable doubt. Are there not some steps, albeit not a criminal sentence, that we

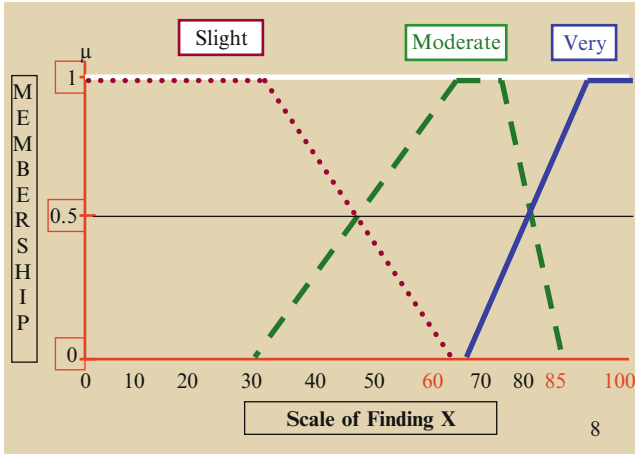


Fig. 2.7 Example of fuzzy logic analysis for “finding X”

could take when the proof does not cross the threshold of full membership of 1.0, but is definitely at membership level of 0.8? Another related question is whether we could apply the same gradation to the credibility of a witness, who might be quite believable, but how do you know for sure? In law, you apply the standard of reasonable doubt to questions of credibility, the same standard as applied to the test of criminal proof (*R. v. J.H.S.*, 2008). I know the answer in law would be to fear that such gradations will become a slippery slope with dangerous consequences. Hopefully, however, it can be safely discussed without leading us down any slippery slopes in the way demeanour evidence has apparently lead us for many, many years.

Acknowledgment Thanks to Dr. Madan M. Gupta, University of Saskatchewan, College of Engineering, for explaining fuzzy logic and for guiding the development of the fuzzy logic charts.

References

- Allen, R. J. (1994a). Burdens of proof, uncertainty and ambiguity in modern legal discourse. *Harvard Journal of Law and Public Policy*, 17, 627–646.
- Allen, R. J. (1994b). Factual ambiguity and a theory of evidence. *Northwestern Law Review*, 88, 604–640.
- Allen, R. J., & Seniuk, G. T. G. (1997). Two puzzles of juridical proof. *Canadian Bar Review*, 76, 65.
- Anderson, T., & Twining, W. (1991). *Analysis of evidence*. London: George Weidenfeld and Nicholson.
- Ding, H., & Gupta, M. M. (2000). Competitive and cooperative adaptive reasoning with fuzzy causal knowledge. *Journal of Intelligent and Fuzzy Systems*, 9, 191–196.
- Ekman, P. (1992). *Telling lies*. New York: W.W. Norton.

- Frank, J. (1949). *Courts on trial*. Princeton: Princeton University Press.
- Kaufmann, A., & Gupta, M. M. (1985). *Introduction to fuzzy arithmetic: Theory and applications*. New York: Van Nostrand Reinhold Company Inc.
- Polya, G. (1988). *Mathematics and plausible reasoning: Patterns of plausible inference* (Vol. 2). Princeton: Princeton University Press.
- R. v. B. (K.G.)* (1948) 7 C.R. 342.
- R. v. Belland and Phillips*, (1987) 2 S.C.R. 398.
- R. v. J.H.S.*, (2008) SCC 30.
- R. v. Lifchus* (1997) 5 C.R. (5th) 1 S.C.C.
- R. v. Mackenzie* (1993) 18 C.R. (4th) 133.
- R. v. Presley*, (1948) 7 C.R. 342.
- R. v. Starr* (2001) 36 C.R. (5th) 1.
- Seniuk, G. T. G. (1992). Judicial fact-finding and a theory of credit. *Saskatchewan Law Review*, 56, 79.
- Seniuk, G. T. G. (1994). Judicial fact-finding and contradictory witnesses. *Criminal Law Quarterly*, 37, 70.
- Seniuk, G. T. G., & Yuille, J. C. (1996). *Fact finding and the judiciary*. Saskatoon: Commonwealth of Learning.
- Zadeh, L. A. (2004). *Fuzzy logic systems: Origin, concepts, and trends*. Hong Kong: Paper presented at the Web Intelligence Consortium.

Applied Issues in Investigative Interviewing, Eyewitness
Memory, and Credibility Assessment

Cooper, B.S.; Griesel, D.; Ternes, M. (Eds.)

2013, XXII, 349 p. 19 illus., Hardcover

ISBN: 978-1-4614-5546-2