UNIVERSITY COLLEGE LONDON

STUDIES OF EVIDENCE SCIENCE

THOUGHTS ABOUT A SCIENCE OF EVIDENCE

DAVID A. SCHUM, PROFESSOR*

SCHOOL OF INFORMATION TECHNOLOGY & ENGINEERING
SCHOOL OF LAW
GEORGE MASON UNIVERSITY
HONORARY PROFESSOR OF EVIDENCE SCIENCE
UNIVERSITY COLLEGE LONDON

[December 29, 2005]

TABLE OF CONTENTS

THOUGHTS ABOUT A SCIENCE OF EVIDENCE

I have been pleased and honoured to be associated with a project at University College London [UCL] entitled: *Evidence, Inference and Inquiry: Towards an Integrated Science of Evidence*. This project, now enjoying the support of the *Leverhulme Foundation* and the *Economics and Social Research Council*, was initiated and is so capably directed by Professor A. Philip Dawid. This project currently involves the active collaboration of persons from a wide array of disciplines at UCL including probability and statistics [Phil Dawid's discipline], law, medicine, geography, education, philosophy, ancient history, economics, psychology, and computer science. Persons involved in the generation, analysis and application of ideas in any area of study have obvious interests in the study of evidence. Explanations and understanding of phenomena they encounter in these activities are grounded on evidence. There is no single discipline that forms the repository of all knowledge regarding the evidential foundations of reasoning. However, the field of law in our Anglo-American judicial system has supplied us with the oldest and perhaps the most extensive recorded legacy of experience and scholarship on evidence. We have so much to learn about evidence from each other and Phil Dawid's project is designed to stimulate and enhance this learning process. The word *integrated* in the title of this project acknowledges the importance of obtaining knowledge about evidential issues as they are encountered across disciplines and then combining and sharing this knowledge in useful ways.

As our work progressed it seemed obvious, from the title of our project, that we would eventually find it necessary to say what is meant by a *science of evidence*. As obvious as this requirement appears, there is nothing easy about satisfying it. What follows is my own attempt to identify what a science of evidence is or might become, what its proper domain and methods of study might be, and what it might contribute in the way of insights about evidence that are useful and helpful to all of us, regardless of our interests. I am under no illusion that my thoughts about such matters will escape criticism. I hasten to emphasize at the outset that my account of a science of evidence does not represent any consensus view reached by those of us involved in our UCL studies. In truth, as far as I can tell from our discussions, there has never been any attempt to reach consensus on this matter. All I can promise in my account of a science of evidence is that it will not exclude persons in any area having interests in the properties, uses and discovery of evidence, who might have valuable insights about evidence to contribute, or who might encounter evidential issues that can be addressed in at least potentially useful ways by such a science.

Defining a *science of evidence* is complicated by the necessity of defining the key words *evidence* and *science*. Defining the word *evidence* is not so easy, as I discovered some years ago[1]. My *Oxford English Dictionary* led me in a circle in its definition of the word *evidence*. I was forced to make a distinction psychologists make between *absolute* and *relative* judgements. We come with no judgmental mechanism that allows us, for example, to say absolutely or exactly how bright is a light. But we can easily say how bright this light is relative to another light with which it is

being compared. I ended up having to make a relative, not an absolute, judgment about what the word *evidence* means. I did so by comparing it to words often used synonymously with *evidence* such as *fact*, *data*, *information* and *knowledge*. Evidence does have three major credentials that can be reasonably defined in terms of questions they allow us to answer. *Relevance* answers the question: so what? Does an item of information bear directly or indirectly on alternative hypotheses or propositions we are attempting to prove or disprove? *Credibility* answers the question: can we believe the information obtained? *Inferential [or probative] force* answers the question: how strongly does relevant evidence point toward any hypothesis being considered? There is still much to be learned about these three credentials and how they are assessed in various contexts.

Defining the word *science* is not so easy either as we observe while examining shelf after shelf of books and papers on the topic of science, what it is or is not, its various alleged methodologies, and its accomplishments. Even talking about a "science" of evidence can arouse controversy. In fact, some colleagues have argued that we should drop the word "science" and simply say that our work involves the "study" of evidence. There are several reasons for discontent about use of the term "science". Many persons whose work requires consideration of the properties, uses and discovery of evidence would not wish to identify their work with science. In addition, we hear arguments that science alone, often made with reference to the physical sciences, can produce conclusions from evidence that can be taken seriously. Allegedly, conclusions reached from evidence by the rest of us are deficient in various ways. One of the ranking examples of an attempt to disparage the work of colleagues in other disciplines came from the physicist Lord Rutherford who said that all science was either physics or stamp collecting. Taken seriously, this leaves the work that most of us do as being equivalent to stamp collecting. But there is a footnote to this story about Rutherford. It happens that he received his Nobel Laureate in chemistry and not in physics[2]. Thus he joins the rest of us non-physicists in the field of stamp collecting.

But I also know that I cannot escape controversy in my attempt to be unrestrictive in a search for knowledge about evidence from whatever source it may come. It might be argued that we are all experts regarding evidence since each one of us makes use of it every day of our lives. However, there are many people who will accept as evidence, about events of concern to them, information that is provided in horoscopes, by psychics, or by televangelists. In America at present we often hear about governmental decisions that are "faith-based" rather then "evidence-based". In some instances a person might report that a conclusion or a decision was based on a revelation from God, or was based on information provided in ancient documents such as the Bible, the Koran or the Talmud. A fair question concerns the credentials, as evidence, of information provided from sources such as those just mentioned. An appropriately identified science of evidence might, at the very least, make us all better informed about the evidential basis for inferences and decisions of concern to us.

So, the above difficulties acknowledged, I will begin my attempt to describe a science of evidence. But I must first tell you briefly about the standpoint or frame of reference from which I have approached this task.

1.0 A STANDPOINT IN THIS ACCOUNT OF A SCIENCE OF EVIDENCE

I have no difficulty at all in recalling the many things I have already learned by my association with two persons also involved in our current UCL project on evidence. My contact with Professors William Twining [UCL] and Terence Anderson [University of Miami, Florida] now goes back almost twenty-five years. They are renowned scholars, educators and practitioners in the field of law. My training happens to be in the fields of psychology and mathematics. So, I have first-hand experience regarding the benefits of cross-disciplinary interactions. What initially brought the three of us together was our common interest in the work of John H. Wigmore, the American legal scholar who is arguably one of the most profound persons who ever wrote about evidence and proof. My discovery in 1970 of Wigmore's works led me to begin exploring the rich legacy of experience and scholarship on evidence to be found in the field of law. Shortly thereafter I began an attempt to bring this rich legacy to the attention of persons in many other disciplines who I believed would profit from their exposure to this legacy.

My contact with Twining and Anderson began in the early 1980s and has increased steadily since then. I was pleased and honoured more than I can say when they invited me to join them as a co-author in a revision of their very influential work on evidence[3]. I now come directly to one of the most important things I have learned from them; it concerns the importance of declaring one's *standpoint* in reporting any analysis based on or concerning evidence. Failure to do so can cause many problems for the person presenting the analysis as well as it can do for members of an audience who are trying to make sense of it. Following are the four essential elements in a declaration of my present standpoint.

The first standpoint element involves my telling you what role I am playing or what "hat" I am wearing in my present account of a science of evidence. In my entire academic career I have been a student of the evidential foundations of probabilistic reasoning. In my role as a student of evidence, studying as many of its subtleties or complexities I could discover, I quickly came to the conclusion that I would need all the help I could obtain, regardless of what disciplines or persons it came from. As I noted above, there is no single discipline known to me that provides all answers regarding the properties, uses and discovery of evidence. A bit later I will mention William Twining's assertion that a science of evidence must be interdisciplinary in nature[4]. I agree, and have already given evidence of my commitment to this view. In a previous work on evidence I drew heavily upon the insights of persons in law, philosophy, logic, probability, semiotics, history, psychology, and artificial intelligence[5]. Thanks to my present work with other admired colleagues on the UCL evidence project, I will now have even more valuable insights to draw upon.

The second element involves my specifying at what stage in what process am I. I have two responses here. The first concerns my career-long interest in the study of evidence. As far as this process is concerned, at age 73 I guess I will have to admit that I am in the latter stages of it. As far as the process of describing a science of evidence is concerned, I am in the early stages of it. Though I did make reference to a science of evidence in a previous work[6], I did not at the time dwell on any characteristics of such a science. My present interest in a science of evidence stems from Phil Dawid's proposal, written three years ago, to the Leverhulme Foundation. His focus on an emerging science of evidence was one of the major reasons for my enthusiasm for joining this effort. At the time this proposal was in its finished form, and was accepted for funding by the Leverhulme Foundation, I began to collect my thoughts about what a description of a science of evidence might entail. I expect to learn much more from the critical comments I expect to receive from readers of my first attempt to identify a science of evidence.

A most important standpoint element is a declaration of one's objectives in the analysis presented. As I have already announced, my objective in this present work is to identify what a science of evidence is or might become, what its proper domain and methods of study might be, and what it might contribute in the way of insights about evidence that are useful and helpful to all of us regardless of our interests. I add here the following related concern. Suppose you are convinced that there is such a thing as a science of evidence; should you care? Part of my burden in this present work is to provide reasons for a widespread interest in such a science that I believe will be of great assistance even to those who will continue to object to my use of the term "science".

Finally, I must also tell you about the materials to which I had access in my present account of a science of evidence. This question is easily answered; I have had a wealth of information to draw upon coming from a wide array of disciplines in which there has been interest in the properties, uses and discovery of evidence. Some of this information is very recent and comes from a meeting held at UCL on 7 June, 2005 during which some of my ideas about evidence were challenged. This meeting set off a vibrant subsequent exchange of views via e-mail that have been carefully recorded and distributed by Jason Davies and Steve Rowland[7]. I thank them for their efforts to make this discourse readily available to everyone interested. I guess my basic problem concerning available materials remains one of selectivity. There is so much relevant material to draw upon and I can't possibly make reference to all of it. You will surely find relevant materials that you believe I should have mentioned. My hope however is that you will not find fault with my interpretation of information I have received from any source. I have no interest in misrepresenting the views of colleagues, alive or dead, who have thought so carefully about matters concerning evidence.

So, these standpoint questions now put on the table, I can begin by telling you what I know about the first persons who entertained thoughts about a science of evidence. I first heard about a science of evidence from a rather unlikely source.

2.0 SOME BEGINNINGS

Israel Zangwill [1864 - 1926] was no scientist, philosopher or logician. Educated at the Jews' Free School and at the University of London, he became a novelist, playwright and Zionist leader. Among his many published works is a mystery story he wrote in 1891 entitled *The Big Bow Mystery*. This story, that Zangwill admits was inspired by Edgar Allen Poe's mystery stories, is the first in what are called *locked room stories*. In these stories a crime is committed in a room locked from the inside and so it becomes nearly impossible to tell who committed it. In this story we are introduced to a retired police investigator named Mr. George Grodman who has some reputation as an investigator who solves cases with logic and evidence. Of course Grodman brings Sherlock Holmes to mind. As Holmes frequently expressed his contempt for the abilities of Inspector Lestrade of Scotland Yard, so Grodman was also contemptuous of the abilities of a Scotland Yard detective named Edward Wimp.

At a late stage in Grodman's investigation of a murder committed in a locked room, he is asked to state his findings to the Home Secretary. Grodman begins[8]:

"Pray do not consider me impertinent, but have you ever given any attention to the science of evidence?"

"How do you mean", asked the Home Secretary, rather puzzled, adding with a melancholy smile, "I have had to do so, lately. Of course I've never been a criminal lawyer, like some of my predecessors. But I should hardly speak of it as a science. I look upon it as a question of common-sense". "Pardon me, sir. It is the most subtle and difficult of all the sciences. It is, indeed, rather the science of the sciences. What is the whole of inductive logic, as laid down, say, by Bacon and Mill, but an attempt to appraise the value of evidence, the said evidence being the trails left by the Creator, so to speak?. The Creator has - I say it in all reverence - drawn a myriad red herrings across the track, but the true scientist refuses to be baffled by superficial appearances in detecting the secrets of nature. The vulgar herd catches at the gross apparent fact, but the man of insight knows that what lies on the surface does lie."

My first reading of Mr. Grodman's comments about a science of evidence was not in the *Big Bow Mystery* itself; I only read this mystery story much later. I first saw Grodman's comment, that appears in slightly reduced form, as the frontispiece of Wigmore's *Science of Judicial Proof*[9]. Quite obviously, Wigmore was impressed by Grodman's assertion about a science of evidence. Wigmore begins his work by saying that he aspires to offer a *novum organum* for the study of judicial evidence[10]. He goes on to say that there are two parts to the study of evidence in law; one involves *proof*, the other *admissibility*. Issues of proof, he argued, take precedence over issues of admissibility. Even if there were no rules regarding the admissibility of evidence in our Anglo-American judicial system, we would still be concerned about the study of evidence as a vehicle of proof. This 1937 edition of Wigmore's work carries the title: "Science of Judicial Proof". But I have always thought that he could just as easily have titled this work:

"Science of Evidence", since most of this work is devoted to a study of the properties, uses and discovery of evidence.

Wigmore's *novum organum* for the study of evidence and proof has application in any area in which we have the task of drawing conclusions from masses of evidence. The title of his work announces that it is based on logic, psychology and general experience. I have no doubt that Wigmore believed his analytic and synthetic methods of making sense out of masses of evidence were applicable in any area; he just illustrated their use in judicial trials, the area in which he had the greatest interest. In tracing Wigmore's intellectual lineage we find a connection at UCL in the person of Jeremy Bentham [1748 - 1832]. In William Twining's analysis of Bentham's and Wigmore's theories of evidence, he argues that, as a theorist of evidence, Wigmore is a direct lineal descendent of Bentham[11]. As Twining states, comparing Bentham and Wigmore,[12]:

Their primary concern was to introduce system, clarity, simplicity, efficiency and above all rationality into the field of evidence. Both saw the study of evidence as a fit study for scientific treatment; even more important they saw factual inquiries in adjudication as being a standard example of factual inquiries in general, the peculiar conditions and constraints of litigation being but secondary. Their common starting point is a general theory of belief in the context of all factual inquiries.

I note here that Mr. Grodman, speaking with all due reverence, says that we all have the task of appraising the value of evidence in the trails left by the Creator who, unfortunately, has also drawn a myriad of red herrings across these trails. In our various disciplines, each of us follows many different trails and we also encounter the many red herrings that have been left across the trails we follow in the inferences and decisions of interest to us. It seems clear that Wigmore, and Bentham before him, were concerned about a science of evidence and proof that would allow us better means for following our trails while avoiding the red herrings we so often find in our searches for the secrets of nature. Wigmore noted that although the field of logic supplies us with many canons for the simplest cases of reasoning, they have not provided us with methods for reasoning based on masses of what he termed *mixed evidence*[13]. I have always interpreted Wigmore's term *mixed evidence* as being made with reference to the many recurrent forms and combinations of evidence that can be readily observed.

Wigmore's hope in his *novum organum* for the study of evidence was to advance the study of proof based on evidence in the complex situations we routinely encounter such as in the field of law. As he noted[14]:

For one thing, there is, and there *must* be, a probative science - the principles of proof - independent of the artificial rules of procedure; hence it can be and should be studied. This science, to be sure, may as yet be imperfectly formulated. But all the more need is there to begin in earnest to investigate and develop it.

It is fair to say that Wigmore made as many advances in a science of evidence and proof as any person who thought about such a science. But I believe that there have been a variety of advancements toward a science of

evidence since Wigmore's time. These advancements have come from persons in many disciplines.

I have often thought it interesting that a science of evidence, as the science of science, should have been proposed by the fictional literary character Mr. Grodman. Perhaps this possibility would have been more widely discussed if it had been proposed by the more well-known character Sherlock Holmes. The observational, interrogative, imaginative and inferential capabilities that Conan Doyle provided Holmes have been very often cited in discussions of the properties, uses and discovery of evidence by persons in many disciplines. There have been other thoughts about what would constitute a "science of science", a good example being those of the eminent biologist Ernst Mayr[15]. I will consider such alternative proposals as I proceed. My story about a science of evidence so far only goes back to Bentham. I must go back farther than this in my attempts to relate the concepts of evidence and science.

3.0 CONCEPTS OF EVIDENCE AND SCIENCE: THEIR EMERGENCE AND MUTATION

My analysis of a science of evidence would quite inadequate if I gave no attention to how the concepts of *evidence* and *science* have emerged and changed over the ages. I understand that appropriate study of the emergence and mutation of these concepts has occupied the attentions of a great many persons having interests in these matters. Here I come to the first of the selectivity problems I earlier said I faced. I have chosen to mention the works of others that I believed would be of assistance in my stated objectives concerning what a science of evidence might involve, what methods it might employ, and who might benefit from drawing upon this science. I have so far given only the briefest account of the meanings of the two terms *evidence* and *science* that I seek to relate. I hope that the following discussion, though embarrassingly brief, will be adequate in acknowledging the kinds of issues we ought to keep in mind in any discussion of a science of evidence.

3.1 On the Concept of Evidence.

We might take as a starting point the obvious fact that we have used information provided by our senses as evidence in drawing conclusions and making choices throughout our history as a species. Information provided by our visual, auditory, tactual, gustatory, olfactory, and proprioceptive or kinaesthetic senses has ensured our survival as a species. The same might of course be said about the role of sensory evidence in the survival of other species. The fact that we have employed evidence throughout our entire history is acknowledged by James Allen in his careful analysis of ancient debates about the nature of evidence[16]. Allen, focusing on works of the ancient Greeks and Romans, tells us that, although the term *evidence* is of ancient origin, its nature was discussed in antiquity using a different term *sign* [*semeia*] or often by the related word *token*. Allen says that the idea of inference from signs or tokens was accepted by the ancient Greeks in their efforts to discover or make clear what is unknown. An example is provided

by the term *semeiotikos*, which was used to identify persons, such as physicians, whose task it was to read and interpret the signs of nature. The physician Galen of Pergamum [139 - 199] understood medical diagnosis to be the process of *semeiosis*, or of sign interpretation[17]. Traditionally, the term *sign* [*signum*] was defined as: *something that stands for something else* [*aliquid stat pro aliqou*]. Thus a patient's reports of stomach pains, observable skin eruptions, or the smell of a patient's breath or urine are all examples of signs pointing to various possible illnesses. As I proceed, I will acknowledge the argument that there has already been a science of evidence for millennia, except that it has gone under another name: *semiology*, the *science of signs*. I have found it interesting that persons engaged in modern studies in the field of semiotics make very little, if any, reference to the work of the persons engaged in research on evidence. In the same way, there is very little reference made by scholars of evidence to the works of persons in the field of semiotics. This is most unfortunate since scholars in these areas have so much to learn from each other.

Allen goes on to tell us that it was Cicero [Marcus Tulliius, 106 - 43 BCE] who first introduced the term *evidentia*, as a Latin rendering of the Greek word ευαργεια, meaning the quality of being evident[18]. For something to serve as evidence for a conclusion, this something must be more evident than the conclusion itself. In any area of interest to us, we find it necessary to make inferences concerning past, present or future events that we can never observe directly. Thus, the historian attempting to make inferences about events in the past must rely upon whatever observable traces have been left behind that seem to bear upon these events. As the historian Simon Schama mentions[19]:

…historians are left forever chasing shadows, painfully aware of their inability ever to reconstruct a dead world in its completeness, however revealing their documentation. … We are doomed to be forever hailing someone who has just gone around the corner and out of earshot.

In the same way, the intelligence analyst attempting to predict possible terrorist activities at some time in the future cannot observe these activities since they have not happened yet. This person can only use as evidence in such predictions observable indicators of the capabilities and intentions of groups or individuals who might be involved in such activities.

Leaving the ancient Greeks and Romans, it is frequently said that civilization in the West entered upon what is commonly called "the dark ages". These ages may have been dark in the West, but they were anything but dark in cultures in the Middle East, in India and in China. Though specific scholarship on evidence is hard to find in these early cultures, the advancements they made in science and technology are indeed impressive, though they are often slighted in Western accounts of the history of science. The advancements made in these cultures certainly points toward their great sophistication in understanding the role of evidence in the advancements they made.

Returning to " the dark ages" in civilization in the West, I first draw upon the insights of Ian Hacking who has provided an account of forms of evidence sanctioned in this period, which he said lasted for a very long

time[20]. In his discussion of evidence [Chapter 4] Hacking tells us that the only form of evidence then taken seriously was testimonial evidence in the form of recorded assertions made by authoritative persons in the past, such as Aristotle, Archimedes, Galen and Hippocrates, or authoritative religious figures in the Roman church such as St. Augustine or St. Thomas Aquinas. Hacking says that what retarded progress in inductive reasoning till the time of Sir Francis Bacon was a concept of evidence that would sanction the incorporation of *tangible* things that *would point to other things*. In other words, reliance upon the recorded testimony of authorities is not a method for learning about any secrets of nature. We must attend to the signs we can observe in nature. Thus Hacking's argument brings him into contact with the field of semiotics. In his work Hacking makes abundant reference to signs, but never mentions the field of semiotics.

I have found a bit of fault with Hacking's account of how rejection of tangible evidence, in the form of observable signs of nature, retarded progress in inductive reasoning. He fails to mention the emergence of activities that were taking place in the early years of Oxford University that helped set the stage for methods by which we put questions to nature in the form of experiments we design and carry out. To set the stage for a discussion of the heroism it took to begin the process of relying on tangible evidence from nature, I draw upon a comment in a work by William and Martha Kneale, who said[21]:

The chief obstacle to steady scientific progress was not the influence of Aristotelian logic or anything else derived from Greece, but the lack of sustained curiosity about things which were not mentioned by ancient authors and did not appear to contribute in any way to salvation. It was easier to get the support of the Pope for an *inquisitio haereticae pravitatis* [an inquiry into the depravity of heretics] than for an *inquisitio naturae* [an inquiry of nature]. (Translations my own).

The penalties for going against the accepted teachings of the church were indeed severe, as several of the persons I now mention discovered for themselves. The four persons I mention were all at Oxford, and some were at various times in Paris: Robert Grosseteste [1168 - 1253], Roger Bacon [1214 - 1292], John Duns Scotus [1265 - 1308}, and William of Ockham [1280 - 1349]. Their activities are an important part of any story about the emergence of the concept of evidence since they all either developed or were aware of empirical evidential methods for isolating causes for the phenomena we observe in nature. Six centuries later, John Stuart Mill was credited for their discovery. We all hear about "Mill's Methods" but rarely do we hear about the persons who appear to have been the first to have discovered them.

The next part of my story about the emergence of the concept of evidence takes us to the field of law in England at the same time period as lived the four Oxford scholars just mentioned. What I have claimed as the rich legacy of experience and scholarship on evidence in the field of law certainly had an unpromising beginning. In these times a defendant charged with a crime had the burden of *proving his innocence*. The following sequence of events occurred: judgment, trial and sentence. At the judgment

11

stage the defendant was given the option of which of three methods would be employed to prove his/her innocence. There was a noticeable absence of evidence in these three methods of proof: trial by combat, various ordeals such as carrying a hot iron, and oaths taken on behalf of the accused by various numbers of people. Proof by all of these methods was left to the judgment of God [*judicium Dei*]. The argument was that God would not side with the guilty party in a trial by combat; God would not allow the burnt hand of a guilty defendant to heal in a certain short period of time; and God would strike dead any person who gave a false oath on behalf of the defendant.

The history of how these evidence-free methods of proof were replaced, gradually, by a jury system is recorded in a number of valuable and interesting works[22]. In their first forms juries bore no resemblance to modern-day juries. They consisted of accusers, witnesses and others with a vested interest in the case. Not surprisingly, many accused persons strongly preferred the older methods of "proof". Interesting accounts of the widespread resistance to these early jury trials are the ones given by Wells[23]. He tells us that until 1728 persons could be mistreated in various ways if they refused to plead guilty or not guilty; and it was not until 1827 that courts would automatically enter a plea of not guilty on behalf of a defendant who refused to plead. A not guilty plea would result in a jury trial.

Just when jurors ceased to be witnesses and began to take on the role of disinterested or unbiased persons who assess the credibility and force of evidence given by external witnesses is a controversial issue. In any case the process was gradual. Wells notes the importance of a statute of Edward III in 1352 that allowed the accused to challenge the suitability of any juror who had joined in his/her indictment[24]. This seems to have been an important step in the transition of the role of juries from being witnesses to being disinterested or unbiased evaluators of evidence introduced by persons not involved in judgments concerning the accused. Another important step was taken in a statute during the reign of Elizabeth I [1563] that compelled the attendance of witnesses at trial and made witness perjury a crime[25].

But there were many bumps along the road to the development of the jury system as we know it today. Before 1670 jurors could be attainted, and mistreated in various ways, if they rendered a verdict the crown or the courts said was "against manifest evidence". A landmark case in this year involved the trail of a jury foreman named Edward Bushel who was accused of encouraging a wrongful verdict in the case of the Quakers William Penn and William Mead, who had been charged with inciting to riot in London. The Chief Justice in this case, John Vaughn, ruled that Bushel could not be punished because state control over jury verdicts was as unfair to the jurors as it was to the defendant. Vaughn ruled that courts could not rule on judgments of fact, but jurors could not rule on questions of law. Thus, we have the division of labour among judges and jurors that, for the most part, exists today. Jurors rule on issues concerning the credibility and probative

force of evidence, although judges rule on the relevance and admissibility of evidence.

As the rights of an accused to representation by an advocate became more extensive, the process became truly adversarial in nature. In this adversarial climate it is to be expected that one side of the matter in dispute will meticulously, and often ruthlessly, examine the evidence and arguments provided by the other. It was here that concern about so many evidential issues arose in our Anglo-American legal system. The adversarial nature of our system led to an often-cited quotation by Sir Matthew Hale who said that questioning by the parties in contention, by the advocates, judges and juries is a better process for "beating and boulting the truth" than any other system lacking this adversarial quality[26]. The philosopher/logician Stephen Toulmin appears to have been equally impressed by this adversarial quality. He argued that logic is concerned with the soundness of claims we can make or the nature of the case we can make for these claims. He then argued that legal analogies are very helpful saying[27]: "Logic (we may say) is generalized jurisprudence".

I now leave the field of law by noting William Twining's discussion of what he terms the *rationalist tradition* of evidence scholarship in our Anglo - American legal system[28]. Twining mentions several variations of this rationalist tradition, which he says are rooted in the English empirical philosophy as reflected by the writings of Bacon, Locke and John Stuart Mill. This rationalist tradition involves[29]: "…facts in issue proved to specified standards of probability, on the basis of the careful and rational weighing of evidence which is both relevant and credible...". This "rational" method may be contrasted with the earlier "irrational" methods involving the judgment of God in trials by combat, ordeals and oaths. This rationalist tradition accounts for the concern in law about very difficult matters concerning evidential relevance and methods used to undermine or support the credibility of tangible or testimonial evidence given at trial. In my efforts to bring the rich legacy of scholarship and experience on evidence in law to persons in other disciplines, I noted that many evidential subtleties or complexities studied in legal contexts have been overlooked in so many other disciplines. William Twining gently chided me for overemphasizing the importance of legal scholarship in such matters, saying that these subtleties are only brought to the surface, in real life cases, in the crucible of adversarial argument[30]. In short, daily experience in trials counts the most.

Probability presents a paradox in the sense that it has a very long past but a very short history. Florence David tells us about objects resembling dice that were possibly used by paleolithic peoples in gambling games, but more likely used as devices for foretelling events in the future[31]. Hacking tells us that these objects were the first randomizers[32]. In any case, the short history of attempts to calculate probabilities is usually said to have begun in the 1600s and are associated with Blaise Pascal [1623 - 1662]. In fact, the historian of probability Irving Todhunter gives us a precise date. He says that the beginnings of mathematical probability began on July 29, 1654 in a letter Pascal wrote to a reputed gamester named Chevalier de Mere concerning a problem in gambling[33]. It is quite common to associate

probability calculations with games of chance, but it did not take long for persons with other interests to become interested in probability calculations. Historians became interested in determining probabilities associated events in the past; merchants became interested in the probability of the safe arrival of cargoes they would ship; some persons in law became interested in probability calculations concerning matters at issue in trials; and even theologians became interested in probabilities associated with miracles. Another event in the 1600s was the founding of statistics and is usually associated with the *bills of mortality* compiled by John Graunt [1620 - 1674] to record vital statistics associated with births, deaths, and the causes of death.

Emerging interest in probability calculations brought an emergence of interest in calculations associated with certain evidential situations, a good example being what were called *credibility-testimony problems*. One such problem involves what was called *simultaneous testimony* or what we would today call *corroborative testimony*. Suppose n independent witnesses who all tell us that event E has occurred where each of these witnesses has probability p of "speaking the truth". Our interest concerns P(E) following each of the testimonies of these n witnesses who all say the same thing. Another problem addressed in these early works was what was called *successive testimony* or what we would refer to today as *second-hand* or *hearsay* evidence. In this case we have a chain of human sources through which a report is passed where it is supposed that each source in this chain has probability p of "faithfully transferring this report". Our interest is in determining the probability that this final report is the same as the one described by the initiator of this report. As Keynes notes, early probabilists recognized that the rareness or improbability of the event reported in testimony, in addition to the credibility of sources, has a bearing on the value of testimony[34].

As Daston[35] and Zabell[36] both report, interest in credibility-testimony problems lapsed among probabilists over the years. In part this was due to their sole interest in enumerative conceptions of probability; i.e. probabilities determined by counting such as *aleatory* probabilities in games of chance or estimates of probabilities by statistical *relative frequencies* for replicable processes. Fortunately, or unfortunately, no one keeps any statistical information on attributes of our credibility as sources of evidence. Speaking of attributes of the credibility of witnesses, it is clear that the early probabilists did not give much thought to the apparent multi-attribute nature of the credibility of human sources. As I noted in another work, I have always found works on the credibility-testimony problem among probabilists to be more entertaining than they are useful[37]. As an example, take their definition of p as the probability that a witness "speaks the truth". In other words, a witness is being truthful only the event he/she reports has actually occurred. But this confounds witness veracity with other attributes such as observational accuracy and objectivity. The witness may have been mistaken in an observation, or was not objective in forming a belief based on this observation. I will return to the attributes of the credibility of human

sources of evidence when I consider epistemological issues that are so important in our understanding of evidence.

Concerning philosophers interested in evidence I begin by giving special attention to the writings of John Locke [1632 - 1704]. In his work *An Essay Concerning Human Understanding*[38] Locke used the term *degrees of assent* to indicate the force of both tangible and testimonial evidence on alternative propositions. But he went much farther by considering a variety of special forms of evidence including *concurrent [corroborative] evidence*, *contradictory evidence*, *second hand or hearsay evidence* and what we would today call *accepted facts*, whose probability Locke said, rises near to certainty. In addition, he considered many of the matters involved in assessing the credibility of evidence. Locke's work has always been very important to me since it gave me additional hope that recurrent forms and combinations of evidence can be usefully categorized. I return to this matter in Section 4.

In all views known to me, the force, weight, or strength of evidence is graded in probabilistic terms in one way or another. Locke's degrees of assent represent one early attempt to relate the force of evidence and probability. In my early work on evidence I became interested in generating equations, following from Bayes' rule, that represent the process of assessing and grading the force of evidence. I had formed the idea that most inferences from evidence involve *chains of reasoning* of various lengths. The first links in such chains concern the *credibility* of the source from which the evidence comes. Later links in a chain involve logical steps in an argument establishing the *relevance* of the observed or reported event on hypotheses or propositions at issue. These relevance links are sources of doubt you believe to be interposed between your evidence and what you are trying to prove or disprove from it. At the time, we called these chains of reasoning *cascaded*, *multistage*, or *hierarchical* inferences. When I began to read Wigmore, I quickly observed that he had already recognized the fact that inferences from evidence involve arguments or chains of reasoning, often having many links. He used the term *catenated inference* to refer to such chains of reasoning[39].

There is an interesting connection between Wigmore and the philosopher Stephen Toulmin. In a very influential work Toulmin described the essential ingredients of arguments based on evidence and how they are related[40]. But Wigmore had already noted these ingredients and their relations years before Toulmin did[41], a fact Toulmin does not acknowledge in his work. Wigmore went much farther the Toulmin in his concern about catenated chains of reasoning and his concern about inference based on masses of evidence. Today we refer to these complex argument structures as *inference networks*. Wigmore was the very first person to study what is involved in the generation of inference networks[42]. Only much later did Toulmin consider chains of reasoning, again without mentioning Wigmore's much earlier work[43].

When I first started my studies of evidence my experience was much the same as the one Wigmore recorded. I could find nothing in the literature in philosophy, or probability, on chains of reasoning, or inference networks,

particularly those based on various forms and combinations of evidence. As Wigmore noted[44]:

The logicians have furnished us in plenty with canons of reasoning for specific single inferences; but for a mass of contentious evidence in judicial trials, they have offered no system.

But I believe Wigmore would certainly have great interest in works on evidence by philosophers that have appeared since his time. I refer especially to the works on evidence of Peter Achinstein. In an edited collection of works by several other philosophers Achinstein provides a variety of thoughts on such issues as what qualifies as evidence, what constitutes relevant evidence, what is the role of confirming and conflicting evidence, and what is the role of evidence in the process of discovery[45].

It seems that Achintsein's edited collection of papers in 1983 provided just an introduction to his recent and more extensive work on evidence[46]. In this work, written essentially for philosophers, Achinstein provides extensive analyses of various concepts of evidence in which he is especially concerned about probabilistic issues that arise. He provides a wealth of examples, mainly drawn from the physical sciences. Especially interesting is his concern about beliefs based on evidence and how they may be justified. In the process, he provides analyses of beliefs formed on the basis of statistical relative frequencies as well as those formed in other situations when we encounter events that are singular or unique. In such situations, we have what are usually termed *epistemic probabilities* representing the degree of our beliefs based upon whatever knowledge we have.

It is to be expected that there are standpoint differences between philosophers and persons in other areas, such as law, in their studies of the complexities of evidence. As a result, different people will ask different questions about the properties, uses and discovery of evidence. If I had had access in the 1960s to Achinstein's 2001 book it is very likely that I would not have appreciated it. The main reason is that I was asking questions about evidence that Achinstein does not answer. I did find answers to these questions in works describing the centuries-old record of experience and scholarship in the field of law. Further, I came to believe that the answers to these questions apply in virtually every other context, I will return to Achinstein's most valuable thoughts about evidence when I seek to defend the idea of a science of evidence.

At our meeting at UCL on 7 June, 2005 someone mentioned that a science of evidence would only be just a part of epistemology. But any science you care to identify is also part of epistemology. The word *science* comes from the Latin word *scientia* meaning knowledge. As we know, epistemology is the study of the nature of knowledge and how we justify it. It is certainly true that most issues arising in the study of evidence have roots in epistemology. We may easily say that we have evidence about a certain event, but whether we have *knowledge* about this event is a more difficult matter. As an example, suppose we wonder whether or not event E happened. You say: "Let us ask person P, she will *know* whether E happened or not." So we ask P if event E occurred and she says that it did occur. Two basic questions we now have are: 1) How do we tell whether P

knows that E occurred?; and 2) Do we ourselves *know* that event E occurred as a result of P telling us that it did occur?

Epistemological issues are inescapable in studies of the credibility or believability of evidence. In the case of testimonial evidence from human sources, we have their *competence* as well as their *credibility* to consider. A competent source is one who had access to the event she/he reports or who made a relevant observation of this event, and one who also had some understanding of what was being observed. Human source credibility rests on other attributes such as veracity, objectivity and observational accuracy. The trouble is that competence does not entail credibility, and credibility does not entail competence.

Unfortunately, some persons confuse competence with credibility often with disastrous results. Here is a human source who did in fact make a relevant observation but who is untruthful in reporting the results of this observation. I will return to these matters in Section 4.2.2 when I consider epistemological issues that arise when we attempt to identify attributes of the credibility or believability of evidence.

I leave my account of how the concept of evidence emerged and changed over the ages by returning briefly to the field of semiotics. I mentioned earlier that semioticians could argue that there has already been a science of evidence in existence for millennia, except that they have called it the *science of signs*. For persons troubled by whether a science of evidence would be relevant to their inferential activities or disciplines, semioticians have an easy answer. They will say: "Our work on a science of signs applies to you, regardless of who you are and what work you are doing". Semiotics involves the study of any kind of communicative process and thus includes study of signs, signals, symbols, and codes of any sort as well as the means by which they are produced and understood. Following is one account of how semiotics claims to be the study of everything that might be construed as evidence in the form of any of the signs given to us by nature, as well as study of the process by which we establish the meaning of this evidence.

In addition to his world-wide reputation as a novelist, Umberto Eco is a prominent semiotician and scholar of medieval history. In one of his works[47], Eco suggests that semiotics studies the whole of culture. But he notes that this grandiose claim invites the criticism that semioticians are arrogant imperialists. He says: "When a discipline defines 'everything' as its proper object, and therefore declares itself as concerned with the whole universe (and nothing else) it's playing a risky game". We might also recall Mr. Grodman's statement in the *Big Bow Mystery*, that "the science of evidence is the science of science"; this seems to be a similarly arrogant claim. But Eco goes on to list nineteen categories of contemporary research in semiotics that indeed seem to cover a very wide array of the signs we receive from nature and our use of these signs as evidence in inferences of concern to us.

Finally, there is one interest in which semioticians and scholars of evidence do have in common and it concerns the exploits of Sherlock Holmes. For example, in historical scholarship, in probability, in law, and in so many other areas researchers have used examples drawn from one of

Holmes' cases to illustrate their concern about some evidential issue. In semiotics there is a very interesting book on the investigative abilities of Holmes and the thoughts of Charles S. Peirce regarding the abductive or imaginative reasoning involved in investigation or discovery[48]. This edited work includes chapters written by a sociologist, a historian, and three philosophers in edition to several chapters written by semiologists. This shows that semioticians are not averse to ideas about signs and evidence generated by persons in other disciplines. I will return to Peirce and Sherlock Holmes later on when I consider matters concerning the discovery of evidence.

3.2 On the Concept of Science and Its Methods

It is said that necessity is the mother of invention but curiosity is the mother of science. Something we all have in common, regardless of our disciplinary interests, is our curiosity or wonder. If we had no curiosity about past, present or future events and phenomena, whatever they are, we would not be engaged in the research we are doing. Whether any of us will say that this curiosity mother has made us scientists is left for each of us to decide. In defining the word evidence, I said that the OED led me in a circle. This is not the case with the word *science*, but there is considerable variation in the definition of this word; here are some alternatives the OED provides[49]:

Knowledge obtained by study; acquaintance with or mastery of a department of learning.

A particular branch of knowledge or study; a recognized department of learning.

A branch of study that deals either with a connected body of demonstrated truths or with observed facts systematically classified and more or less comprehended by general laws, and which includes reliable methods for the discovery of new truths in its own domain.

The kind of organized knowledge or intellectual activity of which various branches of learning are examples.

The intellectual and practical activity encompassing those branches of study that apply objective scientific method to the phenomena of the physical universe (the natural sciences), and the knowledge so gained.

The OED also includes various definitions of the word science that have been used in the past. For example, the word *science* was used in the 1500s to refer to a craft or trade; in the 1600s it was used with reference to activities concerned with theory rather then method. Lord Rutherford would be dismayed to learn that the OED has never defined science to be the study of physics.

Our studies of evidence would easily seem to qualify as a science under definitions 1, 2 and 4. But these definitions are quite unrestrictive and might be said to be weaker definitions. Definition 3 is stronger because it supposes a "connected body of demonstrated truths or facts systematically classified" that are "comprehended by general laws". We might easily argue about whether a science of evidence can have these attributes. However, I will later show how evidence can be classified in systematic ways. I am also

less troubled by the requirement in definition 3 concerning reliable methods for the "discovery of new truths". In Section 3 I will attempt to show various ways in which we are able to discover new truths about the properties and uses of evidence. Evidence has many subtleties or complexities that can be exposed by various methods I will discuss. Once exposed, they can be exploited in our inferential and decisional activities. There is a troublesome feature of definition 5. It seems to suppose that there is a unique objective method that applies only in the physical sciences that apparently excludes the rest of us. As I proceed in this section I will draw upon several very well informed persons who will argue that there is no special or unique method in science.

I return briefly to the saying that necessity is the mother of invention and curiosity is the mother of science. I have never put much faith in this saying since it seems very unlikely that people can invent things without having a strong level of curiosity about how some problem might be solved or how something could be made more efficient, easier, or safer. My belief is that curiosity or wonder has been the engine driving the imaginative reasoning underlying both science and invention. In any case, it seems that we were inventors long before we were scientists. A timeline of science constructed by Ochoa and Corey begins around 2,500,000 years ago, with the hominid species *homo habilis* developing the first stone tools, and ends in the year 1995, the year this timeline was published[50]. A very interesting feature of this timeline is that the authors record 59 human accomplishments between 2,500,000 BCE and 3,000 BCE [99.8% of the temporal distance to 1995] virtually all of which concern technological inventions and not scientific discoveries. I mention technology here because there has been a technology emerging concerning the use of evidence in complex reasoning tasks. In any of our disciplines we have been far better at collecting, transmitting, storing and retrieving information than we have been at using this information in drawing defensible and persuasive conclusions from it. Efforts are now well under way to close this important technology gap.

I go no farther in my account of concepts in science without dwelling upon the crucial role played by mathematics in scientific discoveries and explanations. We have all heard it said that the amazing thing about mathematics is that it works so well for so many purposes. I have found the following account of mathematics given by the philosopher Carl Hempel to be especially helpful in illustrating the value of mathematics in a science of evidence[51]:

But while mathematics in no case contributes anything to the content of our knowledge of empirical matters, it is entirely indispensable as an instrument for the validation and even for the linguistic expression of such knowledge: The majority of the more far-reaching theories in empirical science - including those which lend themselves most eminently to prediction or to practical application - are stated with the help of mathematics; … Furthermore, the scientific test of these theories, the establishment of predictions by means of them, and finally their practical application, all require the deduction, from the general theory, of certain specific consequences; and such deduction would be entirely impossible

19

without the techniques of mathematics which reveal what the given general theory implicitly asserts about a certain special case.

On some accounts I have read, it is said that one criterion for an area being called a "science" is the extent to which this area makes use of mathematics. Mathematics plays several crucial roles in a science of evidence; whether we choose to make use of it or not is another matter. All conclusions reached from evidence are necessarily probabilistic for five reasons. Our evidence is never complete, is usually inconclusive, frequently ambiguous, commonly dissonant to some degree, and comes to us from sources having any gradation of credibility shy of perfection. Probability theories, there are several of concern to us, offer us guidance about how we ought to assess and grade the inferential or probative force of evidence, offer us alternative ways in which we might combine these gradations, and offer us the means for expressing the probabilistic strength of the conclusions we reach. I will return again to the idea that probability is more about arguments than it is about numbers. In the construction of arguments as complex inference networks we are guided by the mathematical theory of graph structures. Both the Wigmorean methods I have mentioned and the newer Bayesian network analyses are consistent with the mathematical requirements of graph structures applied to the analyses of complex inferences. I mention one matter Hempel overlooked in the quote given above that occurs in the study of evidence; it concerns the heuristic merit of mathematics. Certain probabilistic analyses of forms and combinations of evidence we all encounter can prompt us to ask questions about the matters of concern to us that we might never have thought about asking in the absence of such analyses.

I pause here to mention again that one of my major objectives is to consider what methods of study would be appropriate in a science of evidence. My concern of course is to be able to describe methods that all of us could employ, or do now employ, and that we find useful regardless of our inferential and decisional interests. Historical studies of the development of methods in the sciences are interesting, but also vexing in various ways. But so is the history of accounts given in the philosophy of science in which we are told how knowledge is acquired by scientists, how reliable such knowledge can be, and how scientists might do a better job at acquiring knowledge. I left off my historical account of evidence, as far as science is concerned, by mentioning how early Oxford scholars laid some of the foundations for the empirical methods associated with what is frequently referred to as the *Scientific Renaissance* in Western cultures. I can now bring this account more up to date thanks to the work of a philosopher of science named David Oldroyd [University of New South Wales].

In a seminar on the processes of discovery and invention I offer for engineers, I have found a work of Oldroyd's to be especially useful[52]. Oldroyd employs a metaphor he calls the *arch of knowledge* in illustrating how a long list of scientists through the ages have thought about how they generate and test hypotheses of interest to them. In my use of Oldroyd's arch of knowledge I have taken the liberty of adding a few names to this list

including Sherlock Holmes, Charles S. Peirce and the logicians Jaakko and Merrill Hintikka[53].

The upward arm of this arch is grounded on observations we make, and from them generating new thoughts, in the form of hypotheses, about how these observations are to be explained. I have always associated this upward arm with discovery-related processes in which we seek explanations for observations or signs we observe in nature, whatever they might be. It seems that Galileo, the Port Royal Logicians [Antoine Arnauld and Pierre Nicole], Issac Newton, John Locke, and William Whewell all believed this upward, discovery-related arm involves inductive reasoning. John Herschel, however, was not so sure that new ideas are generated by inductive reasoning. His works seem to have opened discussion on the distinction between the generation or discovery of a hypothesis and the justification of it. But it remained for Charles S. Peirce to suggest that a new form of reasoning, which he called *abduction* [sometimes *retroduction*], to account for the imaginative process of generating a new hypothesis. In the case *A Study in Scarlet*, Sherlock Holmes referred to this process as *reasoning backwards*[54]. At the same time, however, he said that his generation of hypotheses was deductive when the backward reasoning he mentioned seems patently abductive.

The downward arm of Oldroyd's arch concerns the generation of possible evidential tests of a generated hypothesis. This involves part of an attempt to justify this hypothesis. All persons I have mentioned view this process as being essentially deductive in nature. What is interesting is that for so many years philosophers concentrated their interests just on the downward arm involving the justification of a hypothesis. In so many works we hear about the *hypothectico-deductive method of science*. What was so long avoided was the upward discovery-associated arm of the arch of knowledge. Even Hempel in the quote cited above just talks about the deductive generation of tests or predictions from some hypothesis and makes no mention of where this hypothesis came from in the first place. In one of the most curiously titled works I have ever read, *The Logic of Scientific Discovery*, Karl Popper advocated what he termed the "elimination of psychologism" from scientific inference. He says nothing about a logic of discovery, relegating it to psychologists. As he stated[55]:

The initial stage. the act of conceiving or inventing a theory, seems to me neither to call for logical analysis nor be susceptible of it. The question how it happens that a new idea occurs to a man - whether it is a musical theme, a dramatic conflict, or a scientific theory - may be of interest to empirical psychology, but it is irrelevant to the analysis of scientific knowledge.

The many contemporary philosophers now interested in the process of discovery cannot have taken Popper very seriously.

I have two more names to mention concerning the arch of knowledge, the logicians Jaakko and Merrill Hintikka[56]. Their view is that the entire process of the discovery and justification of hypotheses is deductive in nature and is based entirely on questions we pose to nature and how they are answered. They call their method the *interrogative approach* to inferences based on evidence. We play a game against nature in our efforts to understand her

secrets. At each play of this game we have two moves we can make. We can either make a deduction about an explanation based on what we have so far, or we can ask nature another question provided that it is in a form that nature can answer. The Hintikka's sneer at the concept of Peirce's abduction and Holmes' alleged employment of it. Their claim is that Holmes was simply superb at asking pertinent questions and in making deductions from the answers he obtained.

Peter Achinstein writes on evidence as I have noted above. But he also writes on methods in science and has given us a valuable collection of excerpts from the works of eminent scientists and philosophers of science regarding methods[57]. In some works "the scientific method" is identified with the hypothetico-deductive method mentioned above. From hypotheses we generate experimental or otherwise empirical tests of these hypotheses. It is usually argued that the hypotheses entertained in science are worthless unless they can be *invalidated* by the evidence we obtain. But Achinstein shows that there have been different views about what the hypothetico-deductive method entails. Many works on the methods of science assume that these methods apply only in instances in which our evidence comes in the form of results obtained in well--controlled experiments frequently those involving the observation of physical phenomena. We are given countless examples of "the scientific method" applied in works in physics, chemistry, and biology. Psychologists and others in the behavioural and social sciences are also well tutored in the hypothetico-deductive method as grounding the methods of science.

But the hypothetico-deductive method has its share of critics, one of whom is the philosopher Derek Gjertsen. He argues that this method does not accurately describe the methodology of science and can readily lead to false conclusions. As he states[58]:

Further, the system allows little room for creativity, originality, or even luck.

It is all very like attempts sometimes made by successful novelists and directors of courses on creative writing to lay down rules for the writing of novels. Yet it happens that novels that scrupulously follow the perceived rules can be totally unreadable, and others which violate every maxim in sight can enthrall. Science, no less than painting, cannot be done by numbers.

You will find no shortage of numbered rules for anyone claiming to do science. Here is a current "science by numbers" approach I found recently on the internet. It comes from Edmund Wilson, a well-known marketer of scientific and technological apparatus. Wilson gives us an eleven-step method for the discovery and testing of hypotheses in science[59].

I now direct my attention to those of you who have read this far and are feeling left out in this account of science and its methods. You may be a scholar of ancient history, political thought, education, religion, or in other disciplines in the humanities or the social sciences that are rarely included in discussions of scientific methods. I will ask you to consider the works of several current scholars whose views should be of considerable interest to persons whose research does not involve experimentation of any kind or the

employment of other methods so commonly associated in the past with the physical sciences. The first person I will mention is a chemist named Henry H. Bauer. Bauer claims that "the scientific method" is a myth that has caused others, including scientists, educators and the general public, no end of trouble[60]. He says that "the scientific method" is useless as a guide to what scientists actually do and that it is worse than useless as a guide to what the public might think about science and technology. In particular, this myth encourages the view that scientists are somehow not like the rest of us, but always are objective, patient, careful, and good observers. As he states[61]:

Indeed, thinking of science as using the scientific method portrays science as an activity that is highly unnatural: human beings are not by nature objective, judicious, disinterested, skeptical, rather, human beings jump to conclusions on flimsy evidence and then defend their beliefs irrationally. The widely held myth\ of the scientific method is one reason why scientists are stereotyped as cold, even inhuman.

In her works Susan Haack does a marvellous job of restoring the humanity of scientists by arguing that their methods and thought processes are the same as those used by careful thinkers in any discipline. In short, she will argue that no readers should feel left behind in the preceding discussion. In a recent work, she seeks to defend science against a variety of extreme charges[62]. On the one hand she rejects what she calls *scientism*, the exaggerated showing of deference towards science and the acceptance of any claim made by science as being authoritative, as if scientists are epistemologically privileged. On the other hand, she rejects the many current cynical critics of science who have said that scientists' stated concerns for honest inquiry, respect for evidence, and a search for truth are illusions being used as a cover for their other agenda relating to power, politics, or rhetoric.

In fact, I will rely upon Susan Haack's work in my claim that a science of evidence excludes no one interested in honest inquiry, a respect for evidence, and a search for truth. As she says[63]:

The core standards of good evidence and well-conducted inquiry are not internal to the sciences, but common to empirical inquiry of every kind. … respect for evidence, care in weighing it, and persistence in seeking it out, so far from being exclusively scientific desiderata, are the standards by which we judge *all* inquirers, detectives, historians, investigative journalists, etc., as well as scientists. In short, the sciences are not epistemologically privileged.

In my account of the science of evidence as a study of the properties, uses, and discovery of evidence I will claim that it includes everyone having the characteristics Susan Haack has just described. Following is just one example of persons who come to mind.

Dr. M. A. Katritzky would almost certainly not claim to be a scientist. Her interests concern historical studies in an area called *theatre iconography*, an interdisciplinary field that involves study of events and materials associated with theatrical performances. I was first made aware of her work while writing an overview of eight studies included in a work by Twining and Hampsher-Monk[64]. The chapter she wrote for this collection

concerns interesting characters referred to as mountebanks, or quacksalvers and the roles they might have played in theatrical performances in the Middle Ages[65]. Applying Susan Haack's criteria of: (i) respect for evidence, (ii) care in weighing it, and (iii) persistence in seeking it out, I can give no better example than this work of Dr. Katritzky's. Her work reveals all three of these attributes in great measure and I would match the quality of her work and the reasoning she applies against any paper I have read in the sciences.

I have one final comment to make before I construct my case for a science of evidence and it concerns *pseudoscience*. There are many recent books and articles on the efforts of possibly unscrupulous persons to advertise their work as scientific when in fact this work does not merit such a label. How surprised I was recently to learn that persons like Phil Dawid, and I have been included in what was recently called *academic pseudoscience*. I have aways had the greatest respect for Professor Mario Bunge [Professor of Philosophy, McGill University, Montreal, Canada]. I read with great interest his work on causality in science[66]. However, in a recent work on what he calls charlatanism in academia, Bunge includes studies of subjective probability and subjective utility as examples of what he calls academic pseudoscience[67]. What he has done is to disparage the work of countless hundreds of faculty members all over the world in fields like probability and statistics, economics, psychology, philosophy, and law in which studies of subjective probability and utility are important, interesting and necessary. I add here that most of the persons on this list, far from being shown intolerance, as Bunge advocates, have been awarded tenured positions at their universities as well as international acclaim for the research they have performed. This list also includes the brightest, most imaginative, and dedicated persons I have ever known No area of academic research is free of criticism and studies of subjective probability and utility certainly receive their share. We are all appropriately sceptical of our own work as well as the work of others. I have participated in several debates concerning which view of probability captures best captures the concept of the weight or force of evidence. But there is an easy explanation for Bunge's denunciation of works on subjective probability that Bunge himself illustrates. It is apparent, from his own words, that Bunge is a frequentist who has the view that the only probabilities that can be of interest are those such as relative frequencies that can be determined by counting. As a result he disparages the use of Bayes' rule; as he states[68]:

This approach contrasts with science, where gut feelings and wild speculations may be confided over coffee breaks but are not included in scientific discourse, whereas (genuine) probabilities are measured (directly or indirectly), and probabilistic models are checked experimentally.

This is the same argument that has been taking place between frequentist and non-frequentist probabilists and statisticians for decades, as Phil Dawid will certainly agree. Frequentists disavow the use of subjective probabilities that Bayes' rule requires. Bunge bears the burden of showing just one instance in science that is free from human subjective judgment.

Bunge also bears the burden of showing how probabilities are to be determined in instances in which we have nothing to count, the events of concern being singular, unique, or on-of-a-kind. But Bunge does allow that studies of belief [he does not call it subjective probability] deserve scientific study. But, as he says[69]:

Such study is important; and, precisely for this reason, it belongs in experimental psychology and sociology, and it should be conducted scientifically. There is no reason to believe that that probability theory, a chapter of pure mathematics, is the ready-made (a priori) empirical theory of belief. In fact, there is reason to believe that credences are not probabilities, if only because we seldom know all the branches of any given decision tree.

But Bunge reserves a special place among academc chartalans for the very large number of distinguished scholars in law for their interest in what has been termed "the new evidence scholarship", a term coined a few years ago by Richard Lempert[70]. Among the matters Lempert suggested in the new evidence scholarship were efforts the determine the guidance that other disciplines, including probability theory, might offer to judicial scholars and practitioners. But here is what Bunge says[71]:

I submit that probability hardly belongs in legal argument because probability measures only the likelihood of random events, not the plausibility of a piece of evidence, the veracity of a witness, or the likelihood that a court of law will produce a just verdict. Consequently, talk of probability in law is pseudoscientific.

In these assorted quotes, Bunge certainly reveals the extent of his myopia concerning probability. Of course he does not even mention works on other views of probabilistic reasoning such as those provided by J. Jonathan Cohen[72] and Glenn Shafer[73].

I began this work by noting that a ranking example of the belittling of colleagues belongs to Lord Rutherford. However, all Rutherford said was that anyone not studying physics was in the field of stamp collecting. But here we have Mario Bunge referring to academic colleagues as "charlatans" engaged in "pseudoscience". This exceeds Lord Rutherford by a considerable margin, and so Bunge now tops my list. Bunge might consider issuing an apology to the many distinguished persons whose work he has disparaged simply because they do not agree with his own frequentistic view of probability.

4.0 ELEMENTS OF A SCIENCE OF EVIDENCE

The burden I now face concerns constructing defensible and persuasive arguments that systematic studies of evidence can appropriately be termed *scientific* in nature. My arguments will concern studies of the classification, properties, uses and discovery of evidence. I can tell you now what my conclusions will be. Earlier, in Section 3.2, I mentioned that our studies of evidence seem to qualify as being "scientific" at least under the three "weaker", or less restrictive, definitions of science that are given in the OED. I repeat these definitions here:

1) Knowledge obtained by study; acquaintance with or mastery of a department of   learning.

2) A particular branch of knowledge or study; a recognized department of learning.

4) The kind of organized knowledge or intellectual activity of which various branches   of learning are examples.

I am going to argue that studies of evidence share several elements of the two stronger or more restrictive definitions of science that are provided by the OED:

3) A branch of study that deals either with a connected body of demonstrated truths   or with observed facts systematically classified and more or less comprehended by   general laws, and which includes reliable methods for the discovery of new truths in its own domain.

5) The intellectual and practical activity encompassing those branches of study that   apply objective scientific method to the phenomena of the physical universe (the   natural sciences), and the knowledge so gained.

It is true of course that the standpoints of scientists and of the writers of dictionaries may be quite different. Persons who do identify their work with science will argue that even the three restrictive OED definitions are still vague and are incomplete in characterizing scientific activity. I will argue that evidence can be usefully categorized, that there are theories concerning the properties and uses of evidence, that mathematics can be profitably employed in both analytic and discovery-related activities, and that experiments of a certain sort can be performed in studies of the uses of evidence. These are all matters that many persons will argue are overlooked in the OED definitions of science.

4.1 Classification of Evidence

I begin my account of a science of evidence by reflecting on the thoughts about science provided by Jules Henri Poincaré [1854 - 1912], the celebrated French mathematician and physicist. Poincaré emphasized the importance of classification in all of science. As he said[74]:

Now what is science? … it is before all a classification, a manner of bringing together   facts which appearances separate, though they were bound together by some natural   and hidden kinship. Science, in other words, is a system of relations.

Years later, the philosopher of science, Rudolf Carnap [1891 - 1970] echoed Poincaré's emphasis on the importance of classification in science[75]. Carnap argued that there are three kinds of concepts in science, as well as in everyday life: *classificatory*, *comparative*, and *quantitative*. He said that by a "classificatory" concept he meant that an object could be placed within a certain class. As examples, he cited taxonomies in botany and zoology. I will come to Carnap's comparative and quantitative concepts in due course.

Poincaré went on to say that the most interesting facts are *recurrent* and can be used several times. He says further that we are fortunate to be born in a world where there are such facts[76]. As an illustration, he asks us to suppose that, instead of there being eighty chemical elements [the number of elements identified in Poincaré's time], there were eighty million elements, more or less uniformly distributed. In such instances nearly every pebble we

could pick up would be unique; nothing we could say about one pebble would tell us anything about any other pebble. This would make any science of geology impossible. Similarly, if every living organism were unique this would make fields such as biology impossible. He concluded that the inability to classify would make any science impossible.

4.1.1 Forms of Evidence. In my studies of evidence I have given a fair amount of thought to the manner in which evidence might be usefully classified. This would be an utterly impossible task if we just considered the substance or content of evidence. The variations in the content of evidence would be at least as great as the number of possible pebbles we could examine if we supposed, as did Poincaré, that there were eighty million chemical elements having equal frequency of occurrence. It seems safe to say that the substance or content of evidence is unlimited in its variety. If all evidence items are unique with respect to their content, how are we ever to say anything general about evidence? Being able to say general things about evidence is useful for many purposes, among which are the *comparative* purposes that Carnap mentioned. For example, there are at least three disciplines that come to mind in which persons drawing inference from evidence must be prepared to evaluate evidence having every conceivable substance; the fields are: law, history, and intelligence analysis. You may be able to add other fields to this list. In such instances, how are persons ever able to compare substantively different items of evidence they encounter in terms of their relevance, credibility and inferential force? Also impossible would be comparisons of evidence encountered across inference problems, perhaps in different substantive areas.

It occurred to me that evidence might be usefully classified on *inferential grounds* rather than upon any grounds relying on its substance or content. What I was looking for were relations among evidence items that, as Poincaré said: "are bound together by some natural and hidden kinship". The forms and combinations of evidence I will describe are recurrent, as Poincaré also required in science; in fact, they occur in various combinations in every substantive area known to me. Figure 1 is a classification of individual items of evidence based on two inferential grounds: relevance and credibility.

RELEVANCE
DIMENSION

CREDIBILITY DIMENSION	DIRECT RELEVANCE		INDIRECT RELEVANCE
	DIRECT	CIRCUMSTANTIAL	ANCILLARY
TANGIBLE [+ or -] [Objects, documents, sensor records, images, charts, measurements, diagrams]			
TESTIMONIAL [UNEQUIVOCAL] [+ or -] Direct Observation Secondhand Opinion			
TESTIMONIAL [EQUIVOCAL] Complete Probabilistic			
MISSING TANGIBLES OR TESTIMONY			
ACCEPTED FACTS			

Figure 1

In several works I have described the taxonomy in Figure 1 as a substance-blind classification of individual items of evidence[77]. My use of this term has caused a bit of trouble that I will explain in more detail a bit later. For now I will just note that this classification only concerns what kind of evidence you have and not the particular use you are making of it, or how you have discovered it. This taxonomy arises in response to two questions that you, the evaluator of evidence, can ask about an item of evidence:

How do you stand in relation to this evidence item?

Generally, how does this evidence item stand in relation to what you are trying to prove or disprove?

I will show how answers to the first question involve credibility-related matters that serve to identify the forms of evidence listed in the rows of Figure 1. Answers to the second question involve common descriptions that refer to basic relevance characteristics of evidence and identify the columns in this figure.

First consider the rows in Figure 1. Different forms of evidence arise when we consider the questions you ask concerning the extent to which you might believe the event(s) recorded in the evidence. These are credibility-related questions about which I will have more to say in Section 4.2 on the basic properties of evidence. Different forms of evidence require different credibility-related questions. Some forms of evidence are <u>tangible</u> in the sense that you have a direct sensory interface with evidence that may or may not reveal the occurrence of events of interest to you. The first row in Figure 1 lists a variety of common items of tangible evidence that are open to your own inspection and allow you to observe what events the evidence reveals. The plus and minus signs associated with tangible evidence refer to <u>positive evidence</u>, that which reveals the occurrence of an event of interest, or to <u>negative evidence</u>, revealing the non-occurrence of an event of interest. Questions asked of the credibility of tangible evidence concern its <u>authenticity</u>, <u>reliability</u> and <u>accuracy</u>. Briefly, authenticity questions ask whether the tangible item is what it is represented to be. Reliability questions concern the repeatability or consistency of the process used in producing the tangible item. And accuracy questions concern the quality of the process by which the tangible item was produced.

In so many other cases, however, you have no access to tangible evidence you can examine for yourself in order to observe whether it reveals some event of interest to you. But you can ask another person whether or not this event occurred. In some cases, of course, this person may voluntarily tell you about certain events without your inquiring whether the events have occurred. What this other person reports to you is called <u>testimonial</u> evidence. There are several important distinctions we have to make regarding the various forms testimonial evidence can take; the first of which occurs in the second row of Figure 1. In some cases a human source of evidence may provide <u>unequivocal testimony</u> that a certain event E has occurred. This source does not hedge his/her testimony about the occurrence of this event; he/she says it definitely occurred. This would be positive testimonial evidence. But the source might instead report unequivocally that event E did not occur; this would be negative testimonial evidence.

In either the positive or negative cases of unequivocal testimony, we now encounter the array of epistemological issues I mentioned earlier in Section 3.1. Suppose a person P reports to you unequivocally that event E occurred. A natural question you could ask person P is: How do you know that this event occurred"? It seems that P has three basic responses he/she could make. They are as follows:

P says: "I observed event E myself". Let us call this a case of <u>direct observation</u>.

Instead, P says: "Person Q told me that event E occurred". In this case we have instances of <u>second hand</u> or <u>hearsay</u> evidence.

Or, P might say: "I had information that events C and D occurred from which I inferred that event E occurred". We can refer to this testimony as <u>opinion</u> evidence.

First consider testimony based on direct observation. You have two basic questions to ask about person P; the first concerns his <u>competence</u>; the other

29

concerns his credibility. These are different questions and should not be confused [as they so often are]. The competence question you would ask of P comes in two parts: (i) Was P in a position where he could have observed event E, or otherwise had access to the event he reports?; and, (ii) Did P have any understanding concerning the reported event? So, access and understanding are matters concerning the competence of a source providing testimony based on direct observations.

But person P might have all the access and understanding in the world and so you believe P to be competent. But whether you should believe what P tells you is another matter. This raises questions about P's credibility. I will argue in Section 4.2 that the credibility of human sources of evidence involves consideration of a source's veracity, objectivity, and observational sensitivity [including the conditions of observation]. These attributes come from a particular epistemological theory that has been around for a long time but is the subject of controversy. This theory is called the *standard analysis of knowledge*. Though this analysis is the subject of controversy, I have found it to be a valuable heuristic in thinking about the credibility of human sources of evidence. I will be able to back my use of this analysis by drawing upon centuries of experience in law, work in sensory psychophysics in psychology, and common experience.

Now consider second hand or hearsay evidence from source P. Person P tells you that he found out about the occurrence of event E from person Q. You may or may not be told how Q found out about event E. Perhaps Q heard about event E from yet another human source R. So we have hearsay chains of any length, some links of which can involve tangible evidence. For example, P tells you that he heard about event E from Q who says he saw a document containing a testimonial assertion by R who claims to have observed event E. In this case we have the veracity, objectivity and observational sensitivity of persons P, Q and R to worry about in addition to the authenticity of the document mentioned by Q. In another place I have examined credibility issues in hearsay chains that quickly become of baroque complexity[78].

We might ordinarily associate hearsay evidence only with trials at law in which there are many rules restricting its admissibility. However, second hand or hearsay evidence is so common in inferences we make every day. In many cases we are not even told about who the sources are in a hearsay chain. In such instances a report is no better than rumour or gossip. How many times have you heard a news reporter say things like: "NBC has learned today that…"; or "BBC has leaned from a usually reliable source that…"? Hearsay evidence is also not uncommon in many disciplines including history, but its hazards are well recognized. The historian Marc Bloch noted that the historian relying upon hearsay evidence: "…is as if at the rear of a column, in which news travels from the head back through the ranks. It is not a good vantage point from which to gather correct information"[79].

In the case of opinion evidence from P we have his credibility to consider as far as his alleged acquisition of information about events C and D are concerned. But we must also examine the arguments P offers to justify his

conclusion that event E follows from events C and D. What we have in the case of opinion evidence resembles the testimony of expert witnesses in trials at law. Such testimony is usually contrasted with the testimony pf ordinary or lay witnesses who are required to speak from personal knowledge, as in the testimony I mentioned based on a direct observation. In many cases a person will testify about an event that may occur in the future. Such testimony can only be opinion evidence since no direct observations are possible about events that have not yet happened.

Human sources of evidence do not always state with certainty that some event of interest has occurred; they equivocate or hedge their testimony in various ways as I indicate in the third row of Figure 1. I cannot help noting that so many government, industrial and military witnesses, who appear at hearings in Washington DC, have raised testimonial equivocation to the level of an art form. This is especially true of what I have termed complete equivocation. Asked whether event E has occurred, common responses are: (i) "I don't remember"; (ii) "I don't know", or (iii) "I'm not your best witness in this matter". Unfortunately, two inferences are possible from such complete equivocation: (i) The witness is honestly impeaching his own credibility; he does not remember, he does not know, or he is not a suitable witness in the matter at issue; or, (ii) The witness does remember that event E occurred, knows that E occurred, or is a competent witness, but refuses to tell us. In instances such as in case (ii), we seem entitled to conclude that this person believes that testifying about event E would not be in his/her best interests.

Equivocation may not be complete but only probabilistic on occasion. In some cases a human source will equivocate using numbers such as: "I am 60% sure that event E occurred, and 40% sure that it did not occur". In other cases, a human source equivocates verbally, using such statements as: "I am fairly sure that event E occurred", or: "It is very unlikely that E occurred". In such instances we might construe these probabilistically hedged responses to be indications of the source's own assessments of his/her credibility. Whether we should believe these assessments rests on other information we have about this source.

The fourth row in Figure 1 lists tangible or testimonial evidence items that we say are missing. To say that evidence is missing means that we have expected to be able to gather it but are unable to do so. In some cases our failure to obtain expected evidence can have inferential force depending on various reasons why it is missing. As we have all heard, missing evidence is not the same as negative evidence; the absence of evidence is not the same as evidence of absence. First, take the case of expected but missing tangible evidence. Perhaps you are looking in the wrong place for it; it may have been lost or destroyed; or it never existed in the first place [your expectation was incorrect]. But another explanation is that someone is withholding this evidence from you. If someone refuses to produce tangible evidence when you request it, you are entitled to draw the adverse inference that the production of this evidence would not be in the best interests of the person who refuses to produce it. This adverse inference is licensed in our Anglo-American system of law.

Refusal to give testimony is not the same as complete equivocation as described above. In such cases a source, whom we believe to be competent, when asked whether or not event E occurred, does not even say he does not remember or does not know whether event E occurred; he says nothing. In cases at law a defendant's refusal to testify is a privilege guaranteed in our legal system, and no adverse inference is permitted from his refusal to testify. Whether this same privilege is also granted to witnesses who are not defendants is the subject of controversy. In other contexts it seems natural to suppose that a human source who refuses to tell you whether some event E happened has reasons for doing so. Inferences about what these reasons might be can be inferentially valuable.

The final row in Figure 1 contains evidence of the sort that John Locke said "rises to near certainty"[80]. In some cases you would never be required to prove that information about certain events is credible; we have what are termed *accepted facts*. You would not be required to count the number of people in London and in Dover to prove that London has a larger population than does Dover. Nor would you be required to journey to the Middle East in order to prove that Iran and Iraq share a common border. You would also not be required to prove that heroin is a narcotic and that arsenic is a toxic substance. The entries in tide tables and celestial tables supply examples of authoritative records as do tables of chemical compounds, physical constants, and mathematical formulae. The only thing you would be required to prove if you used such tabled entries in an inference is that you had extracted the correct information from the table you consulted.

One final comment is necessary about the forms of evidence shown in the rows of Figure 1; they can occur in various mixtures that give rise to some very difficult credibility assessments. Earlier I gave one example of how we may have a mixture of testimonial and tangible evidence items in the case of hearsay evidence. But there are many other situations in which we encounter chains of various forms of evidence. Historians, for example, certainly encounter various tangible document chains or trails leading back to an alleged testimonial assertion made by some historically interesting person. Your passport supplies another example of a document trail leading back to a birth certificate on which is recorded the time and place of your birth as witnessed by the obstetrician or other person who assisted in bringing you into the world.

Now consider the three columns in Figure 1 that involve terms commonly associated with some very general relevance relations. As I noted, these terms arise in response to your question: how does this evidence stand in relation to what I am trying to prove or disprove? The first distinction I have made is between directly and indirectly relevant evidence. This distinction is necessary since we have evidence and evidence about this evidence. Evidence is said to be directly relevant if you can link it by a defensible chain of reasoning to hypotheses you are considering. Even if you cannot establish this linkage, evidence can still be relevant, but only indirectly so, if it bears upon the strength or weakness of directly relevant evidence. It is common to refer to this evidence as being ancillary evidence, since it is evidence about other evidence. Wigmore understood the

distinction between directly relevant and ancillary evidence as did another prominent scholar of evidence. As the Swedish jurist Per Olof Ekelof noted: "What we call a 'piece of evidence' is strictly speaking an evidentiary fact together with auxiliary facts attached to it"[81]. So, we might refer to ancillary evidence as <u>auxiliary evidence</u> or, perhaps, as <u>meta-evidence</u> since it is evidence about other evidence.

I have listed two forms of directly relevant evidence: <u>direct</u> and <u>circumstantia</u>l evidence. Direct evidence, if credible, is said to go in one inferential step to a matter revealed in the evidence. In other words, credible direct evidence is conclusive on a matter revealed in the evidence. Circumstantial evidence, on the other hand, is always inconclusive on some matter, even if it is perfectly credible. I have preserved the distinction between direct and circumstantial evidence in Figure 1 on the off chance that you seen these two terms used on occasion. However, in our recent work on evidence we have dropped this distinction[82]. The basic reason is that a single reasoning stage can often be decomposed into several stages, each revealing a source of doubt. As an example, the term direct evidence is often associated with "eye witness testimony". Here is a person who asserts that she saw event E occur. According to the definition of direct evidence, if this person is credible, that settles it, event E has occurred. The trouble is that we can decompose the link between her testimony and whether or not event E occurred into additional links that capture doubts associated with her veracity, objectivity and observational sensitivity.

In summary, Figure 1 shows fifteen generic and recurrent forms of individual items of evidence. This number is reduced to ten if we drop the distinction between direct and circumstantial evidence. Recall that mixtures of these forms of evidence are possible in any context or situation, as I illustrated. The example I gave involved cases in which we have testimonial evidence contained in a tangible document.

4.1.2 Combinations of Evidence. I have gone one step further in my evidence taxonomy by considering various generic and recurrent <u>combinations</u> of two or more evidence items. In another work I have provided diagrammatic representations of these combinations and more extensive discussions of them[83]. Following is a brief account of these combinations that involve patterns of evidential harmony, dissonance and redundance.

Two or more items of evidence can be <u>harmonious</u> in two basic ways. The evidence can be <u>corroborative</u> when two or more sources [of whatever kind] report or reveal the same event. For example, two human witnesses and a photograph, all tell us that event E occurred. But harmonious evidence can also be <u>convergent</u> in the following way. We have evidence from two or more sources about <u>different</u> events, all of which favour the same conclusion. As an example we have evidence about events E and F where we believe that both of these events favour proposition or hypothesis H. Evidence combinations reveal many interesting and important subtleties or complexities that can be exploited if they are recognized. A very important subtlety associated with convergent evidence is their possible <u>synergism</u>: two or more convergent items of evidence can have increased inferential

force when they are considered together than they would have if they are considered separately or independently. As an example, we continue to believe that evidence about events E and F, if credible, would each individually favour H. But we also believe that, taken together, they interact inferentially in favouring H to a greater extent that they would do if we did not notice their synergistic effect when they are considered together.

Failure to recognize convergent evidential synergism can be inferentially hazardous as we witnessed following the September 11, 2001 terrorist incidents in New York City and Washington, DC. Our FBI and CIA each had evidence pointing toward immanent terrorist activities but such evidence was never considered jointly. In part this was due to statuary restrictions on the sharing of information between these two agencies. Had items of in formation been shared by these two agencies at least the attacks on the World Trade Center in New York, using airliners as weapons, could have been anticipated and perhaps avoided. Evidential synergisms cannot be recognized and exploited when information is not shared.

Two or more items of evidence can be <u>dissonant</u> in two basic ways that are often confused. Evidence is <u>contradictory</u> if it reports the existence of mutually exclusive events, those which cannot happen together. In the simplest of cases, one source reports to us that event E occurred [this would be positive evidence of E as noted in Figure 1] . But another source tells us that event E did not occur [this would be negative evidence of E]. A contradiction would not necessarily involve evidence about binary events such as E and not-E. For example, one source tells us that Osama Bin Laden was in Kabul, Afghanistan on the morning of March 4, 2003; but another source tells us that he was in Karachi, Pakistan at this same time. Osama cannot have been at both of these places at the same time. However, Kabul and Karachi clearly do not exhaust all the places where Osama might have been at this time.

But another form of dissonant evidence is <u>divergent</u> in character. Here we have evidence about two events that can happen together but simply point us in different inferential directions. A contradiction always involves events that cannot happen together. I once was asked to analyze evidence obtained from over 50,000 patients who had experienced cardiothoracic surgery. Many of these patients lived, but some died. Two of the many indicators that were gathered from those who lived and those who died were: (i) the patient's age at the time of surgery, and (ii) the number of previous episodes of cardiothoracic surgery the patient had experienced. As you might expect, the younger a patient is the better are the chances of surviving this kind of surgery. But the existence of prior episodes of cardiothoracic surgery reduce the chances of this person's surviving another episode of this kind of surgery. Now, here comes a patient who is young; he is just twenty years old. This seems to favour his surviving the heart surgery that is being considered in his case. But we also observe that he has had prior heart surgery several years ago. This seems to disfavour his survival from the contemplated heart surgery. Notice that these two events can happen together; they are divergent in their inferential implications but are not mutually exclusive events.

Resolving evidential contradictions rests upon credibility issues. Since the events in contradictory evidence cannot occur jointly, one or the other of the sources must have been untruthful, not objective, or mistaken. But resolving evidential divergence is a more complex matter. Credibility issues arise, but there are other things to consider. In some cases it may be said that we lack understanding of what the evidence is telling us and we may be mistaken in saying that the evidence diverges. In some situations we can resolve evidential divergence by considering the joint occurrence of the events reported. We encounter no logical difficulties here because the events reported in such evidence can occur simultaneously, as I illustrated in my cardiothoracic surgery example. We can have a patient who is young but who also has had prior episodes of this surgery. When we examine the joint occurrence of these two events, we find that prior episodes of cardiothoracic surgery actuarially swamp patient age and so the divergence disappears. Our initially saying the age and prior episodes of surgery are divergent was based on treating these two diagnostic indicators separately or independently.

Finally, two or more evidence items can be <u>inferentially redundant</u> in one of two basic ways. In such instances the inferential force of one item of evidence can be reduced when other evidence is considered. The first pattern of redundance is said to be <u>corroborative</u> in nature. When we hear about the <u>same event</u> from several different sources [as in the corroborative case of harmonious evidence mentioned above], successive accounts of this same event usually mean less to us. All depends on the credibility of our sources of information about this event. If we believed our first source of evidence about this event was perfectly credible, then additional accounts of this same event would be completely redundant and would add no inferential force since they tell us something we already believe. The second account of this event springs to life to the extent that we find the first source not credible. The third account of this event springs to life to the extent to which we believe the second source not to be credible, and so on.

But there is a pattern of redundance that can involve evidence of <u>different events</u>; I have called this pattern <u>cumulative redundance</u>. I have chosen the word "cumulative" that is used in the field of law to indicate evidence that adds nothing new. Suppose source P reports that event E occurred and source Q reports that event F occurred. We believe that if event E occurred this would reduce the force of knowing that F occurred. Here is an example that comes from the analysis Jay Kadane and I performed on the evidence in the celebrated American law case involving Nicola Sacco and Bartolomeo Vanzetti[84]. Briefly, Sacco and Vanzetti were tried, convicted and executed for killing a payroll guard during a robbery that occurred in 1920 in South Braintree, Massachusetts. Controversy exists to this day about whether they committed this crime. They were anarchists, which they freely acknowledged. It has been repeatedly argued that Sacco and Vanzetti were convicted and executed because of their anarchistic beliefs and not because they were guilty as charged.

The prosecution produced two witness whom the press labelled "star" witnesses. The first was a man named Lewis Pelser; the second was a man

named Lewis Wade. Pelser testified that he saw Sacco at the scene of the crime at the time it was committed. The prosecutor expected Wade to corroborate Pelser's testimony by saying that he also saw Sacco at the scene of the crime when it was committed. But this is not what Wade testified; he only said that he saw someone who looked like Sacco at the scene of the crime when it was committed. Thus, Pelser and Wade reported different events; Sacco's being there and someone looking like Sacco being there are not the same events. In any analysis of the testimony from these two witnesses credibility plays the same crucial role as it does in the case of corroborative redundance. If we believed Pelser to be perfectly credible, then Wade's testimony tells us nothing, since if Sacco were indeed at the scene and time of the crime, then someone who looked like Sacco was certainly there. But Wade's testimony springs to inferential life to the extent that we believe Pelser was not credible in his testimony. There was a variety of evidence introduced that undermined Pelser's credibility and a smaller amount that undermined Wade's credibility. Unfortunately, this unfavourable evidence had little if any effect on the twelve jurymen who voted for the conviction of Sacco and Vanzetti.

4.1.3 On "Substance-Blindness". I have now completed a description of my categorization of recurrent forms and combinations of evidence. There are two reasons why I have dwelt on this matter; the first concerns the importance both Poincaré and Carnap placed on classification in science. No science of evidence would be possible if we could not classify evidence in any meaningful way. I understand that there are various ways in which evidence might be classified. I will put my classification of forms and combinations of evidence to use in Section 4.3.1.

The second reason for my dwelling upon classification involves my use of the term "substance-blind" with reference to the evidence classification I have proposed[85]. I have argued that the forms and combinations of evidence I described are observable and recurrent in any context or discipline known to me regardless of the substance or content of the evidence. All my classification scheme does is to say <u>what kind</u> of evidence is being considered; it says nothing about the particular properties, uses and discovery of evidence in specified contexts or situations; these are not substance-blind matters. Unfortunately, my use of the term "substance-blindness" has been used in various ways that I have never intended. Some have viewed my work as a substance blind "theory of evidence", or even as a substance blind "theory of probabilistic reasoning"[86]. Such theories, in either case, raise issues of substance and involve so much more than just a consideration of what kind of evidence is at hand.

I'll close my account of subtance-blind evidence classification with two sets of examples that I hope will illustrate what Poincaré said about "…facts which appearances separate, but which are bound together by some natural and hidden kinship". The first involves tangible evidence. You would be hard-pressed to find a greater substantive disparity between the kinds of evidence Dr. Katritzky encounters in her studies of theatre iconography in past ages and the kinds of ballistics evidence encountered in a murder trial. If you look at her Plate 3 on page 240 of her work that I have cited[87], you

will see three figures representing mountebanks performing on a makeshift stage called a "trestle stage". These three figures are all wearing theatrical costumes of interest to Dr. Katritzky. The source of this figure is an album allegedly compiled by one M. A. Pribil sometime during the years 1587 - 1594.

Now consider the murder trial of Sacco and Vanzetti[88]. A forensic surgeon named Dr. George Magrath testified that he extracted four bullets from the body of Alessandro Berardelli, the payroll guard Sacco was accused of shooting. Dr. Magrath says he marked each of these bullets with a Roman numeral on its base. He testified that the bullet he had marked "III" was the one that had killed Berardelli. Bullet III was marked as Exhibit 18 and was shown at trial. This bullet was represented by the prosecution as being the bullet that killed Berardelli. The prosecution then introduced expert witnesses who argued that Bullet III had been fired from a 32 calibre Colt automatic pistol Sacco had in his possession when he was arrested. But this Bullet III, Exhibit 18, was never shown simultaneously with the other three bullets. If it had been, the jurors and the defence might well have noticed that the other three bullets cannot have been fired from the same weapon as Bullet III. Other witnesses all testified that the person who shot Berardelli used the same weapon repeatedly.

We have tangible objects in both of these situations; a drawing and a bullet. What is their "natural and hidden kinship"? This kinship come in the form of the identical credibility issues they both raise. The first is Dr. Katritzky's concern about the <u>authenticity</u> of the figure in her Plate 3. Was it really compiled, and perhaps drawn, by M. A. Pribil sometime during the years 1587 - 1594? How <u>accurate</u> was this drawing as a representation of mountebanks' costumes? Did M. A. Pribil actually attend the mountebank performance in which the characters were wearing the costumes depicted in the figure? How good an artist was Pribil? Sadly, her devoted concern about authenticity was nowhere evident concerning Bullet III in the Sacco and Vanzetti trial. No one asked the question: Was Exhibit 18, shown at trial, the same Bullet III that Dr. Magrath extracted from Berardelli's body? Both the prosecution and the defence were allowed to test-fire bullets through Sacco's Colt automatic. One distinct possibility, much discussed, is that the prosecution substituted one of the bullets test-fired through Sacco's weapon for the real Bullet III that had been extracted from the guard's body. In short, the Bullet III shown as Exhibit 18, may not have been authentic. There was concern, exhibited by the defence, about the accuracy, and perhaps reliability, of records concerning the chain of custody through which Exhibit 18 might have passed. There were many missing or weak links in this chain of custody that have been recognized by all commentators on this case, but they obviously had little if any impact on the jurors who tried the case.

My examples of testimonial evidence come from two disparate situations; a psychoanalytic interview and a terrorist investigation. In a psychoanalytic session patients naturally give testimonial accounts of their difficulties and their experiences. In the investigation of a terrorist incident reliance is placed on HUMINT [human intelligence] that often comes in the form of a person's report about what he/she has seen or heard. What is the

"natural and hidden kinship" among these two contexts in which people give testimonial evidence? The answer is the concern in both situations about attributes of the <u>competence</u> and <u>credibility</u> of the persons who provide testimony. As I noted above in my discussion testimonial evidence in Figure 1, attributes of competence involve access to the information reported as well as understanding of it. Credibility attributes involve the veracity, objectivity and observational sensitivity or accuracy of person providing testimony.

There is concern in the field of psychoanalysis about the credibility of patient reports of their difficulties and the experiences they encounter[89]. Of special concern is the truthfulness of a patient's report and the role of memory in recounting events in the past. In Section 4.2 I will show how memory-related factors arise when we consider the objectivity of a person concerning how he/she formed a belief about these past events. Unfortunately, in the field of intelligence analysis a human source's competence and credibility are often confused. I have seen many accounts saying that we can believe a source S because he/she had good access to the event's being reported. Unfortunately, this is a competence matter and not a credibility matter. Source S may have all the access the world but be untruthful, not objective, or mistaken about what he/she reports. Unfortunately, the "hidden kinship" existing among instances of testimonial evidence is not equally well recognized in every context in which reliance is placed on testimony from human sources. This is just one reason why an identified science of evidence would be so useful to persons in many contexts as I will mention again in Section 5.0.

4.2 Studies of the Properties of Evidence.

I have just shown how recurrent forms and combinations of evidence can be classified in what I regard as useful ways, regardless of the substance or content of the evidence. But there is so much more to a study of evidence than just identifying what types of evidence we may encounter in any situation. Evidence has several important inferential properties that I have referred to on occasion as "credentials". These properties, which I will now consider, are: <u>relevance</u>, <u>credibility</u> and <u>inferential [or probative] force. weight or strength</u>. A fourth property may be identified, <u>completeness</u> or <u>sufficiency</u>, but I will include this property in my discussion of inferential force, weight or strength. The reason is that one view of the weight of evidence rests upon the degree of its completeness.

My objective in this section is to continue to show how studies of evidence involve the three major concepts Carnap identified as being associated with science: classificatory, comparative, and quantitative[90]. In the preceding section I dwelt on evidential classification and gave just a few examples of the necessity for evidential comparisons. But in the discussion to follow I will provide additional examples of evidence comparisons and will consider quantitative concepts when I consider the inferential force, weight or strength of evidence.

I note that these properties or credentials of evidence immediately raise matters that do depend upon the substance or content of evidence and the

particular context or situation in which the evidence is being employed. In short, these credentials are not "substance-blind". The major reason is that evidence never comes to us with these credentials attached; they must be established by defensible arguments that rest upon the substance or content of the evidence as well as the nature of its linkage to the propositions or hypotheses at issue in the evidence-based inference task of concern in a particular context or situation.

Finally, I note that theories abound in all of the sciences as well as in many other contexts. In fact, a compilation of over five thousand theories occurring in a wide variety of disciplines is available[91]. In studies of evidence we also encounter theories concerning the relevance, credibility and force of evidence. In the sciences we normally think of theories as being testable by various empirical means. But the theories we encounter in evidence science, in common with theories in many other areas, require testing by other means that I will describe as I proceed.

4.2.1 On the Relevance of Evidence

At the most general level *relevance* answers the question: So what? You receive an item of information or a datum and ask how this information bears upon anything of interest to you. Charles Darwin once asserted that any observation must be for or against some view if it is to be of any service at all[92]. We use the term relevance with reference to other matters besides evidence; we often speak of relevant hypotheses, assumptions, arguments, variables, and so on. Common synonyms for the term relevance are: pertinent, applicable, germane, apposite, connected, related and linked. Thus, following Darwin, we can say that an item of information or a datum only becomes evidence if its relevance on some matter at issue is established by argument. But arguments in defence of the relevance of an item of evidence can often be a complex chain of reasoning consisting of many links. Such arguments arise as the result of both imaginative and critical reasoning processes. We have first to imagine what the links in a relevance chain of reasoning might be. There are few cases in which we can consult a reference source to tell us what these links should be in particular situations. Then, we must then subject our chain of reasoning to critical analysis to see whether it contains any disconnects or non sequiturs. What we wish to avoid are what logicians call a *fallacy of relevance* in which the premises of an argument are incapable of establishing its conclusion.

As I noted in discussing Figure 1, a distinction must be made between directly relevant and indirectly relevant [ancillary] evidence. Here I introduce several terms I have borrowed from Wigmore[93]. He employed the Latin term *probandum* [a matter to be proved] to identify stages in an argument from evidence, and he distinguished among several levels of probanda. The major proposition or hypothesis to be proved or disproved can be called an *ultimate probandum*. To sustain this ultimate probandum, various main lines of argument must also be established; these can be termed *penultimate probanda*. Lower level probanda that serve to link an item of directly relevant evidence to a penultimate probandum Wigmore termed *interim probanda*. So, directly relevant evidence on this view

concerns evidence that can be linked through a chain of interim probanda to some penultimate probandum. On this view, we might say that these penultimate probanda provide touchstones for determining the relevance of evidence.

Now, the "glue" that either holds, or fails to hold, our relevance arguments together consists of generalizations, that license reasoning from one probandum to another, together with ancillary evidence that either supports or undermines the applicability of a generalization in the particular case in which it is being applied. Thus, the applicability of a generalization rests upon the quality and completeness of ancillary evidence. I note here that Toulmin, discussing the ingredients of argument from evidence, has used the term *warrant* or *license* instead of the term *generalization*; and he used the term *backing evidence* instead of the term *ancillary evidence*[94]. I will later mention how the term *assumption* is often used instead of the term generalization.

Generalizations, warrants or assumptions are necessary ingredients for establishing the relevance of evidence. In an argument concerning relevance, particularly one having many links or inferential steps, generalizations are frequently not overtly expressed but often lurk in the background. Though they are necessary in all arguments from evidence, there are some distinct hazards associated with them. In addition, many kinds of generalizations have been identified along with various inferential hazards they present[95]. Generalizations and their evidential backing are always content or context-specific. Evidence relevant in one context, or at one time, may be irrelevant in other contexts or at other times.

Before I consider various theories concerning relevance, I must briefly mention two basic forms of argument structures that arise in complex inferences based on a mass of evidence. These structures have a name currently given to them; they are called *inference networks*. The first person known to me to provide a systematic study of complex inference networks was Wigmore[96]. Wigmore's basic objective was to develop an analytic and synthetic method for making sense out of masses of evidence so that conclusions could be drawn in defensible and persuasive ways. The analytic part of Wigmore's method was to develop an index, that he called a *key list*, of all the probanda, evidence items, and important generalization involved in an inference network. Then, by means of his *evidence charts*, comes the synthetic task of showing how all of the key list items are linked together.

An example of a Wigmore evidence chart appears below in Figure 2A. Wigmore argument structures tend to be hierarchical in nature with lines of inductive inference proceeding from the evidence to an ultimate probandum. The interim probanda in such argument structure represent sources of doubt that may be interposed between evidence and penultimate probanda. So, an inference network in a Wigmore chart has lines of reasoning, indicated by the arrows, go from the evidence to an ultimate probandum. Thus, the arrows or links in such a chart indicate avenues by which items of evidence are relevant on a penultimate probandum. Wigmore's basic use of an inference network was to chart stages of argument in the task of trying to draw conclusions from an emerging mass of evidence.

Constructing an argument in defence of the relevance of a single item of evidence is often a difficult task. But constructing relevance arguments from an entire mass of evidence, by means of a Wigmore inference network, can be an astonishing difficult task. I know this from first-hand experience. Jay Kadane and I constructed an inference network based on 395 items of trial and post-trial evidence in the Sacco and Vanzetti case[97]. Our network was divided into 28 sectors, each concerning an issue that was raised during the trial. I also invite you to examine a similarly complex argument structure in the works of Mark Geller[98] and Terence Anderson[99] in their analysis of inferences concerning the time at which the Sumerian language became extinct. This illustrates how this method of argument structuring knows no disciplinary boundaries.

But not all inference networks have the hierarchical structure of a Wigmore network. An example of a network having a different structure appears in Figure 2B. In another work I have called such networks *process models* in which the objective is to analyse a complex process consisting of variables believed to be involved in this process are linked together in certain ways[100]. What the arrows linking these variables indicate has been the subject of controversy. On various accounts they are said to indicate causal relations, relevance relations, or probabilistic dependencies. I have summarized these various interpretations in another work[101]. Analyses of such process models involve observing how various patterns of evidence influence the probability of any of the variables on the network.
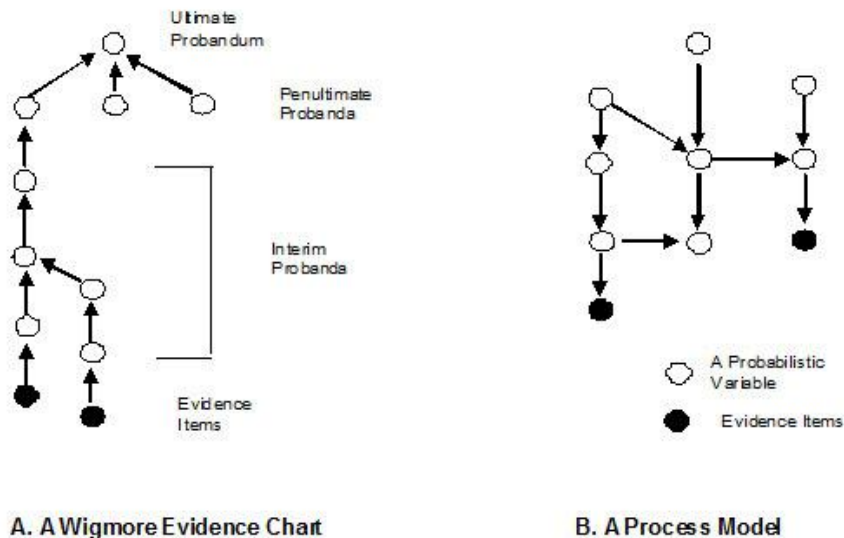

Figure 2

As I noted earlier, mathematics enters our studies of evidence when we examine the structure of arguments we generate from evidence. The mathematical theory of graphs forms an underpinning for the argument structures we generate. The two forms of argument construction just shown in Figure 2 have a different structural appearance but they have a common property. They are both examples of what are called *directed acyclic graphs* [DAGs]. The term *directed* means that there are lines of inference showing relevance, probabilistic influence, or perhaps causal influences among the elements [probanda or variables] in the structure. These are indicated by the arrows. The term *acyclic* means that you cannot follow a chain of linkages or arrows that leads you right back to where you started. This would be a most unfortunate property for an evidence-based argument to have. Imagine an argument that leads you right back to where you started. You would be in an inferential loop and you would never be able to draw any conclusion or generate any meaningful relevance argument. I have found it interesting to note that the arguments Wigmore generated in 1913 had this DAG property well in advance of developments in the mathematics of graph structures.

I now consider two theories concerning relevance. In the first, relevance is not a matter of degree, but in the second it is. I will first consider relevance as it is viewed in the field of law. In America, the Federal Rules of Evidence provide a definition of relevance by means of Federal Rule of Evidence [FRE] 401[102]:

"Relevant evidence" means evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable that it would be without the evidence.

In the Advisory Committee's Notes to FRE 401, several interesting comments are provided regarding this definition. They first note that relevance is not an inherent property of evidence but exists only as a <u>relation</u> between an item of evidence and a matter properly provable in the case. This relation obviously must be in the form of a defensible or logically sound argument linking the evidence with the matter to be proved [a *probandum*, to use Wigmore's term]. I can't help recalling Poincaré's assertion that science is a system of relations[103]. We have evidence of different recurrent forms that we are trying to bring together in their relations to what we are trying to prove or disprove. There is a modern metaphor that has been frequently used to describe what is involved here; it is called "connecting the dots". In Section 4.3 to follow I will have more to say about how one view of a science of evidence is that it is a science of "connecting the dots". Connecting lots of dots is a task that everyone faces in the inference tasks we all perform regardless of our disciplines and or standpoints.

Another important matter addressed in the Advisory Committee's Notes to FRE 401 concerns the following consequence of the definition given in this rule. Notice that FRE 401 does not say how much more probable or less probable relevant evidence should make the existence of any fact that is of consequence to [or is material to] the action. The Advisory Committee noted that any more stringent requirement than "more probable or less probable" would be unworkable and unrealistic. All FRE 401 says

essentially is that relevant evidence must have <u>some</u> probative or inferential weight, force or strength in changing our beliefs about this material fact. So, there is a direct connection between relevance and probative or inferential force. As I will discuss further in Section 4.2.3, we can grade probative or inferential force in various numerical ways, but a similar grading of relevance is not possible. I also noted how Richard Lempert saw the connection between relevance [as given in FRE 401] and probative force[104]. He showed further how an ingredient from Bayes' rule captures this connection.

It is true that relevance, as it is defined in FRE 401, is not a matter of degree. However, in some cases we may require evidence of another matter F to justify the relevance of evidence E. Federal Rule of Evidence 104b covers such situations[105]. This rule FRE 104, Preliminary Questions, Part (b) Relevancy conditioned on fact, asserts:

When the relevancy of evidence depends upon fulfilment of a condition of fact, the court shall admit it upon, or subject to, the introduction of evidence sufficient to support a finding of the fulfilment of the condition.

The basic problem addressed by FRE 104(b) is that evidence is presented seriatim at trial and not all in one lump. Defence of the relevance of evidence E might have to await the presentation of evidence F. But this does not say that evidence E has any particular degree of relevance, even if condition F is satisfied.

I now consider a very interesting alternative view of the concept of relevance proposed by Dan Sperber and Deirdre Wilson[106]. [I note that, at least at the time this book was published, Wilson was on the faculty of Linguistics at UCL]. The view expressed in their work is that the concept of relevance is absolutely basic to an understanding of any form of human communication and to the various cognitive processes involved in what is being communicated. I found this work on relevance especially interesting for a number of reasons. First, I am not the only person to make use of Carnap's assertion of the importance of classificatory, comparative, and quantitative concepts in science[107]. In their work on relevance Sperber and Wilson show us how these same three concepts arise in studies of the concept of relevance[108]. I also note that the authors made reference to studies in semiotics, which is quite understandable in light of their interest in communication and linguistics. However, they disagree with some conclusions reached by semioticians[109].

In Sperber and Wilson's view relevance is a matter of degree[110]. I will try my best to relate this idea to problems we face in defending the relevance of evidence on probanda or matters to be proved or disproved in our inferences. First, the authors define an *assumption*, a critical ingredient in their view of relevance, as follows[111]:

By *assumptions*, we mean the thoughts treated by an individual as representations of the actual world (as opposed to fictions, desires, or representations of representations).

As I read this, it seemed to me that their use of the term *assumption* is closely related to what I mentioned earlier as *generalizations*. I had said that generalizations and ancillary evidence form the "glue" that holds, or fails to

hold, our arguments together. An obvious way of criticizing an argument is to examine whether the underlying generalizations or assumptions, and their evidential backing, seem to make sense. In extreme cases we would have a non sequitur in which one proposition in a chain of reasoning does not follow from a preceding proposition. But the links in chains of reasoning in an argument will vary in strength depending upon a number of factors involving the strength with which we believe an asserted generalization or assumption holds as well as the number of links in the chain of reasoning.

I short, some arguments seem stronger than others, this is just one example of how Carnap's *comparative* concept enters our discussion of relevance. The glue holding together some arguments is stronger then the glue holding together others. In the case of a non sequitur we might assert that the wrong glue is being used. Sperber and Wilson allow that assumptions can vary in their strength[112]. The authors state that the assumptions we make will vary in the strength of the confidence with which we assert them. They even allow such confidence to be expressed in the form of subjective probabilities. This is a most important idea to which I will return when I consider the inferential force of evidence. In the same way, asserted generalizations are always hedged probabilistically in some way, as I will illustrate in a moment. This makes them inductive generalizations.

I now relate assumptions to Sperber and Wilson's view of the manner in which there are degrees of relevance. I will first note that an assumption or generalization, being someone's assertion of how the actual world works, always depends on context. One way of showing that an asserted assumption or generalization holds in the situation in which it is asserted is to back it with appropriate ancillary evidence. What the authors do in showing that relevance is a matter of degree, is to subject the assumptions/generalizations we make to what they call a cost/benefit analysis in terms of the effect and effort related to the assumption/generalization. Here is what they have said about the extent of relevance and its dependence upon the effect and effort of assumptions[113]:

Relevance:

Extent Condition 1: an assumption is relevant in a context to the extent that its contextual effects in this context are large. *Extent Condition* 2: an assumption is relevant in a context to the extent that the effort required to process it in this context is small.

It seems that *Extent Condition* 1 refers to the strength with which we can back up an asserted assumption/generalization in which we have expressed great confidence. In other words, we believe the glue holding together this particular step in our argument is very strong. But *Extent Condition* 2 also requires that we can easily show how the backing evidence we have does support the assumption/generalization we have asserted. In so many cases we would all be quite at home arguing about the extent to which we have appropriately backed an asserted assumption/generalization.

I close my discussion with an example I hope will show the relationship between the two views of relevance I have mentioned. There is nothing about the Sperber and Wilson view that argues against the view of relevance

as expressed in the field of law. Recall that FRE 401 simply says that evidence is relevant only if it has <u>some</u> force in changing one's belief about the probability of an event of consequence in a trial. The example I have in mind comes from an argument I constructed in our analysis of the evidence in the case of Sacco and Vanzetti. At this trial a prosecution witness, a police officer, named Michael Connolly, testified that he had tried, without success, to keep Sacco from putting his hands under his coat. Sacco had acknowledged that he was carrying an automatic pistol at the time of his arrest, but he specifically denied what Connolly said in his testimony. The judge ruled that Officer Connolly's testimony was relevant and it was admitted. The judge later told the jurors that Connolly's evidence was very powerful because it showed Sacco's consciousness of having killed the payroll guard.

But at trial the prosecutor, Frederick Katzman, was never asked to produce an argument showing exactly how Connolly's evidence was relevant in an argument that Sacco knew he had committed the crime with which he was charged. But in our analysis of this case I was obliged to try to construct an argument in defence of Connolly's testimony. My argument or chain of reasoning consisted of eight links which I have described showing the generalizations I was obliged to assert at each link in this argument[114]. These generalizations or assumptions are all instances of my beliefs about the way things work in the world. Here are just three of the generalizations I asserted:

The events reported by police officers testifying under oath usually have occurred.

Persons who intend to intend to use or threaten to use weapons on arresting officers will most often do so because of their intention to escape from custody.

Persons who intend to escape from arresting police officers are usually conscious of having committed a criminal act.

The evidence backing the first generalization was not favourable to it; Connolly may have "cooked up" this story. There was no evidence at all to back up the second generalization. There was evidence bearing on the third generalization, but it acts to diminish its effect [to use Sperber and Wilson's term]. Sacco said he was conscious of having distributed seditious literature which was, at the time, against the law. The long and short of it is that I managed to construct an argument in defence of an item of evidence that I believe has the same degree of strength as the soup that Abraham Lincoln once described that was "made by boiling the shadow of a pigeon that had already been starved to death". The analysis by Sperber and Wilson is valuable in showing how the inferential force of evidence, which can be graded in various numerical ways, depends on the generalizations/assumptions we assert. But they do not suggest that we can grade the relevance of evidence in any numerical way. No Federal Rule of Evidence makes this suggestion either. The overall cogency or defensibility of an argument, and the generalizations/assumptions on which it is grounded, cannot be cast in numerical terms.

4.2.2 On the Credibility of Evidence and Its Sources.

Given an item of information in the form of a tangible object or a testimonial assertion, the question is: can you believe what this item says? You may of course have asked this question before you asked the "so what?' question concerning the relevance of this item. In any case, as I noted in my discussion of Figure 1, the credibility questions you ask of tangible evidence are different from the ones you ask of testimonial evidence provided by human sources. But I also noted that we can have mixtures of tangible and testimonial evidence where we have both kinds of questions to answer, such as instances in which we have a tangible document recording the testimony of a human source. In Section 3.1 I mentioned the epistemological issues we naturally encounter in assessing the credibility or believability of evidence. It is now time for me to consider these issues in more detail. As I proceed, I will continue to focus on matters that will arise in any discussion about there being a science of evidence.

I first need to mention the role of credibility questions in the construction of arguments based on evidence. Credibility questions always form the very foundation of any arguments we generate concerning the relevance of evidence [of any kind]. A very simple example is shown in the chain of reasoning in Figure 3.
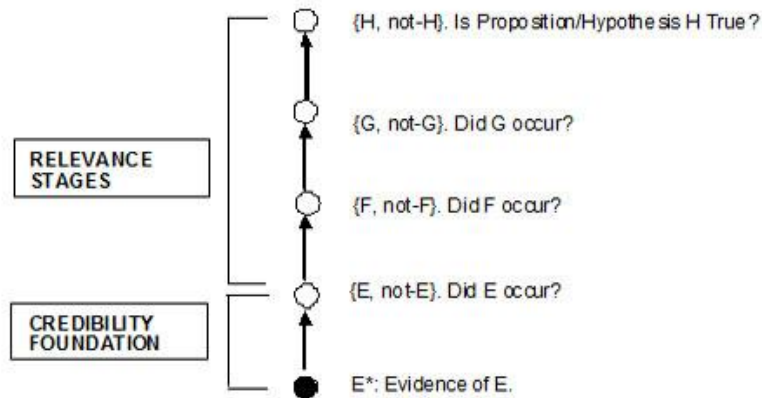

Figure 3

Suppose we have evidence E* that event E occurred. We must always distinguish between evidence for an event and the event itself. Thus evidence E*, that event E occurred, is not the same as the actual occurrence of event E. For the moment forget about whether E* is tangible or testimonial evidence. Suppose we all agree on the suitability of the generalizations/assumptions that license steps in our relevance argument on a probandum of interest: is proposition or hypothesis H true? Here is our

relevance argument: "The occurrence of E licenses an inference of F; F licenses an inference of a G; and, in turn, G licenses an inference of H. Our generalizations licensing this chain of reasoning are all appropriately hedged. We are not asserting that E makes F necessary, F makes G necessary, or that G makes H necessary. What we have in this relevance argument are three sources of doubt.

But the very foundation for this argument comes in the form of the inference we must make about whether or not event E occurred based on evidence E*. Just because we have evidence E* does not entail that E did occur. Based on E* we must first make an inference about the extent to which we can infer that event E did occur. This is the credibility-related foundation stage of the entire argument shown in Figure 3. So, there are four sources of doubt interposed between our evidence E* and the proposition H we are trying to prove from it. These sources of doubt are indicated by the questions shown in Figure 3.

I know of no branch of science, or any other area of serious investigation, in which credibility-related matters are ignored. The chemist or physicist is every bit as concerned about the credibility of what their instruments tell them as the psychologist is in the credibility of various forms of evidence they collect about elements of human behavior. I have already noted the care that Dr. Katritzky has taken concerning the credibility of the evidence she gathers in her studies of theatre iconography. So, a science of evidence must be able to say useful things about the credibility of any kind of evidence encountered in any situation. In my remarks on credibility I will make use of all three of Carnap's concepts in science: classificatory, comparative, and quantitative. In Figure 1 I have already made use of a classificatory concept by distinguishing among the various forms of evidence by means of the credibility-related questions they impose.

Before I dwell further on the attributes of tangible and testimonial forms of evidence, I need to make additional comments about the chain of reasoning shown in Figure 3 that combines relevance stages and a credibility-related foundation stage. A simple truth is that the relevance or credibility stages of any chain of reasoning we might generate can always be decomposed to reveal additional sources of doubt. For example, suppose a critic arrives and examines the relevance stages of our argument as shown in Figure 3. The critic says: "I don't believe you can infer G directly from your F. You need another stage in your argument, namely interim probandum J , that can be inferred from your F, and which will allow you to more appropriately infer your G". In addition, as I will later discuss, the credibility-related foundation stage of this argument can be decomposed to reveal specific concerns about the attributes of either tangible or testimonial evidence. Our concerns about each of these credibility attributes represent additional sources of doubt. Interesting epistemological issues arise at this point. In the case of testimonial evidence I need to explain where the three attributes I have identified come from. I will make use of a theory from epistemology to identify attributes of the credibility of testimonial evidence.

Credibility Attributes: Tangible Evidence

Authenticity. Common experience with tangible evidence reveals concern about its authenticity, reliability, and accuracy. As I noted in explaining tangible evidence in Figure 1, we have first to consider whether this tangible item is what it is represented to be or what we believe it to be. The issue of authenticity often brings to mind the possibility of efforts on the part of others to mislead us in various ways. Documents can be forged, photographs can be altered, and objects can be contrived or misrepresented in various ways as I illustrated in Section 4.1.3 in my example concerning Bullet III in the Sacco and Vanzetti case. A more recent example involves a document that caused no little embarrassment to our [then] Secretary of State, Colin Powell. At a United Nations hearing he presented evidence drawn from a document represented as being an agreement between Saddam Hussein and persons in the country of Niger concerning the purchase by Saddam of certain nuclear materials. This document was later shown to be a rather clumsy forgery.

The trouble is that we often need no other person or group to mislead us; we can easily mislead ourselves. I know of one instance that occurred years ago that led to an incorrect inference that was based on a very high quality intelligence photograph. This photo was simply mislabelled about the time it was taken. We thought it was taken two days ago, when in fact it was taken about two weeks ago. The blood test you experienced last week produced a serious result that surprised both you and your physician. You take the test over again and no such surprising result occured. A short time later the testing service apologized for the lack of authenticity of your first test. Someone had mislabelled the test results of another person as belonging to you. A mistake was made but you would probably not believe the person who mislabelled your blood specimen tried to mislead you in any way.

Apart from the fields of intelligence analysis, law and medicine, historians are obviously concerned about authenticity questions. As Lichtman and French tell us, some forgeries were viewed as authentic for many years until they were carefully examined[115]. The examples they mention include the documents concerning the Donation of Constantine and the Protocols of the Elders of Zion. Inauthentic objects include the Viking rune stones found in Minnesota. The authors emphasize a fact well known in law, namely that it often requires considerable technical skill and highly specialized knowledge to uncover a forgery. Evidence of fabrication, and the degree of similarity between the artifact or document and other items of the period are examined carefully. So is the chain of custody through which the item has passed. In law, great attention is now paid to establishing the chain of custody through which some tangible item, to be presented at trial, has passed. Each stage through which an item passes offers opportunities for the intentional or unintentional alteration of or substitution for some original item. Sadly, such concern about chain of custody was not so evident in 1921 when Sacco and Vanzetti were tried and convicted.

Reliability. The term *reliability* requires special attention. The reason is that this term is so often applied, as a synonym for *credibility*, with reference to human sources and the testimony they provide. I believe this to

be a mistake. What is at issue regarding the testimony of human sources is the extent to which we can *believe* what they tell us. Believability or credibility is a much more complex matter as I will now begin to explain and to which I will return when examining the attributes of testimonial evidence.

Reliable processes are dependable, repeatable, or consistent. In engineering, for example, reliability concerns the probability that some system will continue to function for some specified time in the future. In the behavioural sciences the reliability of some test concerns the likeliness that it gives the same, or nearly the same, result on successive occasions. As we know, there are various statistical indices of the extent to which some result may be expected to occur in repeated samples of the same process.

One reason why reliability becomes a concern regarding tangible evidence is that so many tangible items open to our direct inspection are products of various devices. Examples include photographs and other sensor images such as radar and infra-red, voice recordings, and the traces of physical processes recorded by instruments of various kinds. The issue here concerns the reliability or dependability of the process or device used to make the tangible recording of the process of interest. One matter of interest here concerns the stated reliability of the device in question: How often would you get the same result if you took another reading? Another matter of interest concerns how well the device was maintained. As an example, the prosecutor at your trial on a charge of speeding produces a tangible record that was produced by the arresting police officer's radar gun. This record shows that you were travelling at 75 miles per hour, where the posted speed limit on part of the M-5, where you were travelling, was 60 MPH. You deny that you were exceeding this speed limit. Fortunately for you, your attorney produces evidence that the radar guns used by the police in the area in which you were arrested have a poor record of service. In addition, your attorney offers the testimony of a person who recently serviced the radar gun used in your arrest. This person testifies that this radar gun was damaged and looked like it had been dropped on several occasions.

Accuracy. Another attribute of many items of tangible evidence concern their accuracy. Sensing devices that produce images of some sort vary in the degree of their sensitivity or resolving power. We may not be entirely sure that it was person X in the image captured by the surveillance camera at the bank where the robbery occurred. The tangible document you are examining may contain errors and omissions. Perhaps this document was translated into English from another language. You would of course be concerned about the accuracy of this translation. Tabled entries of measurements of various sorts may contain errors. We are all made aware of various ways in which charts or graphs of statistical data can be misleading A common device used to enhance truly small observed differences is to plot these small differences on an expanded scale.

I come finally to an accuracy issue involving tangible objects. This issue concerns your own detection and recognition accuracy. You examine some tangible object and believe that it reveals an event E that is of interest in the inference task at hand. But a colleague, perhaps having better detection and

recognition capabilities than you possess, argues that this object does not reveal the occurrence of event E. This example illustrates how it is often necessary to rely on the opinion of other persons whose experience and skills allow them to provide advice regarding all three of the attributes of the credibility of the tangible evidence you are examining.

Competence and Credibility Attributes: Testimonial Evidence

I come now to matters I have found to be the most interesting and difficult of any I have encountered in my studies of evidence. As I noted in Section 3.1 concerning the emergence of the concept of evidence, a problem was addressed in the very earliest studies of probability and evidence in the 1600s; this problem was called the *credibility-testimony problem*. I also noted that I found discussion of this problem among probabilists to be more entertaining than useful. The major reason was that I could think of attributes of the credibility or believability of human sources of evidence that these studies did not address. I began my own studies of the credibility of testimonial evidence in the late 1960s. I had already formed the belief that credibility, as a credential of testimonial evidence, has several attributes. So, my question was: how many attributes influence the inferential force of evidence? My studies of this question took me into several fields including law, epistemology, and sensory psychophysics, as well as the common experiences we all have with the testimony given by others. Fortunately, these four areas of thought converge in suggesting the three attributes of the credibility of testimonial evidence that I have already mentioned several times: *veracity*, *objectivity*, and *observational sensitivity* [including the conditions of observation]. But my studies in law convinced me of the importance of another characteristic of human sources of testimony, namely the *competence* of the person providing testimony.

Before I begin my account of where these attributes of competence and credibility come from, I will again be concerned about how the matters I mention can appropriately be termed necessary ingredients in a science of evidence. In the process, I will make use of Carnap's comparative and quantitative concepts in all of science. First, studies of the attributes of the credibility of testimony allow us to make several important comparisons. They allow us to compare the relative credibility of different persons who may provide testimony in a given situation. Second, they allow us to compare the credibility of testimonial and tangible forms of evidence we receive. We might ordinarily believe that tangible evidence, things we can observe for ourselves, are naturally more credible than testimony we receive from other persons. That this belief is not always justified becomes apparent in our studies of credibility. This comparison seems quite important in contrasting credibility issues in various areas of science and in other areas that I will now illustrate.

In most areas of the physical sciences the greatest reliance is placed on tangible evidence in the form of recorded observations that are made as a result of repeated trials involving some observable phenomenon. The researcher explains with care the reasons why these observations were made

and how the results of these observations were obtained. Critics questioning these announced results can make their own observations under the same described conditions to satisfy themselves if they care to do so. But in the social and behavioural sciences and in various areas of the humanities, equally great reliance is placed on testimonial evidence obtained in some, but not all, situations that are replicable. I will return to a comparison of the forms of evidence obtainable in various areas later in Section 5.0 when I consider the benefits to all of us obtained by systematic studies in a science of evidence.

There are ways of grading the credibility of a source of testimonial evidence in quantitative terms. As I will illustrate later in Section 4.2.3, numerical assessments of the credibility of evidence arise naturally in numerical assessments of the inferential force of evidence. These credibility assessments can be combined in various ways with other numerical assessments concerning the strength of the linkages in a relevance argument that relates the event reported in testimony to some probandum or matter to be proved. In short, the strength of all the links in the chain of reasoning shown in Figure 3 can be expressed numerically in probabilistic terms. Appropriate combinations of these assessments can produce assessments of the inferential force of evidence that grounds the chain of reasoning. As I will continue to emphasize, my present purpose is simply to show how such quantitative assessments of the inferential force of evidence are possible. Whether they are employed or not depends on the interests of the person drawing conclusions from the evidence of interest to this person. In short, a science of evidence does in fact include methods for quantitative assessments of the credibility and the force of evidence.

Competence Attributes. I have already briefly mentioned two attributes of the competence of sources of testimonial evidence in Section 4.1 in connection with my description of the classification of evidence in Figure 1. In my studies of the manner in which testimonial evidence is assessed in our legal system I began to learn how the competence and credibility of a source of testimony are entirely different characteristics. I first observed this distinction in a note attached to the definition of competence given in Black's Law Dictionary. This notes says[116]:

Competency differs from credibility. The former is a question which arises before considering the evidence given by the witness; the latter concerns the degree of credit to be given to his story. The former denotes the personal qualification of the witness; the latter his veracity. A witness may be competent and yet give incredible testimony…competency is for the court, credibility is for the jury.

I will add other attributes besides veracity as far as the witness's credibility is concerned. I also noted that the distinction between competence and credibility is not always appreciated in other areas as it is in law; intelligence analysis is one such area I know of. In news reports, as well as in intelligence documents I have read, I have seen numerous statements such as the following: "We can believe what X told us because he had good access". This is an open invitation to inferential miscarriage.

Competence does not entail credibility, nor does credibility entail competence.

In times past persons having various characteristics would automatically be ruled incompetent. Members of certain religious groups, including atheists of course, convicted felons, and even in some times members of certain racial groups as well as the spouses of a defendant were ruled incompetent. There are no such rules today as expressed in the Federal Rule of Evidence FRE 601 which says: "Every person is competent to be a witness except as otherwise provided in these rules"[117].

But FRE 602 requires that ordinary or lay witnesses must have "personal knowledge" of the event(s) about which he/she testifies. This is an important rule to which I will refer again shortly in my account of attributes the credibility of testimony. We will have to sort out what it means to say that a witness has "personal knowledge"; this is where interesting epistemological issues arise. Briefly, what FRE 602 says is that we must have evidence to support the witness's actually observing the event to which he/she testifies, or had access to the information in his/her testimony. Though this rule is frequently cited as relating to credibility I have always thought that it forms perhaps the most important attribute of competence, If the witness did not make a relevant observation or did not otherwise have access to the information in his/her intended testimony, this surely goes against his/her competence.

Another attribute of competence is frequently cited. The witness must have the mental ability to understand what he/she observed so that this person is capable of providing an intelligible and coherent account of what he/she observed. So I have labelled *access* and *understanding* as the two major attributes of the competence of testimonial evidence. But I note that these attributes apply to ordinary or lay witnesses. But expert witnesses will appear in many contexts besides law. The opinion evidence they provide, which I noted in discussing the classification in Figure 1, involves a variety of matters discussed in FREs 702 - 706. There is also a rule concerning opinion evidence provided by ordinary witnesses; it is FRE 701. I will have more to say about opinion evidence in my discussion of credibility that follows.

Credibility Attributes.

I began consulting works on evidence in law in about 1970. The very first topic of interest to me was what evidence scholarship had to say about credibility-testimony problems. Over the next few years I consulted dozens of treatises on evidence law and found a wealth of interesting information of great value in the probabilistic work I was doing. Of particular interest to me was the experience, accumulated over many centuries in our Anglo-American judicial system, of forming questions regarding the impeachment or support of the credibility of witnesses. In Section 3.1 I mentioned how concern about the credibility of external witnesses began to emerge around 1500. Study of these questions became very important to me as I began to think about what are the major attributes of the credibility of persons who provide testimonial evidence. I will return to these questions in a moment.

At the time, I also had interests in the area of signal detection theory [SDT] and its applications in sensory psychophysics. In SDT studies human subjects are given series of trials on which visual or auditory signals at various low levels are presented and always against a background of noise. One objective in such studies was to determine the *absolute thresholds* of our sensory systems, i.e., the lowest level of signal intensity that we can reliably detect. The use of SDT in such studies is interesting because this theory was not generated by psychologists, it was borrowed from the field of electrical engineering[118]. In such studies the subjects report Y = I heard it [or I saw it], or N = I did not hear [see] it; there was only noise. What separated SDT studies of thresholds from previous studies in psychology was that on a certain percentage of trials no signal, but only noise [N], would be presented. This allowed the researcher to determine *hit rates* [h] and *false-positive rates* [f] defined as follows:

h = P(Y|S + N), the probability that the subject says she/he heard [saw] it, given that   the signal was present [S], and f = P(Y|N), the probability that the subject says she/he heard [saw] it, given no signal but only noise [N].

Plots of h and f on what is called *a receiver operating characteristic curve* [ROC] are very revealing because for the first time researchers were able to separate out measures of our observational sensitivity from a variety of other factors, such as our expectancies and our desires, that influence a person's willingness to response Y or N. Work on SDT studies also influenced my work on the credibility of witnesses who provide testimonial evidence as I will explain. I wrote several early papers exploring this connection[119].

But I was finally able to identify and name the major attributes of the credibility of testimony when I began to examine works in the area of epistemology concerning what is meant by the word *knowledge*. Recall my example in which we wish to know whether or not event E occurred and you say we should ask person P, she will know whether E occurred. So, we ask P and she says E*, that E did occur. The question is how do we tell whether or not P knows that E occurred? In epistemology there exists what is termed the *standard analysis* of knowledge, which is traceable to Plato's *Theaetetus*[120]. On this analysis, *knowledge is justified true belief*. In the case of person P, we say on this analysis that P *knows* that event E occurred if: (i) E did occur, (ii) P was justified in believing that E occurred [i.e. that P got good evidence that E occurred] , and (iii) P believed the evidence she obtained.

This standard analysis of knowledge seemed to make perfect sense to me [as a non-epistemologist]. But, as I read further about this analysis, I encountered great debate about its merits. I began to read about paradoxes associated with the standard account of knowledge that were so often linked to the works of Gettier[121] and Radford[122]. So I began to see what the trouble with the standard analysis was all about. In another place I have recorded my own reaction to the wide array of examples and counterexamples of paradoxes associated with the standard analysis[123]. The more I read the more I encountered imaginative but hardly convincing examples of paradoxes.

But I also began to read comments from philosophers concerning their reactions to this imaginative but unconvincing dialogue. Nozick said he became so mired in counterexamples and increasingly complicated conditions for knowledge that he finally stopped reading the literature[124]. Cohen said that discussion of these paradoxes had led many epistemologists into "an imaginary world of freaks, speculation, and science fiction"[125].

I found no account of the requisites for knowledge that is free of controversy, So I decided explore the consequences of the standard analysis to see where it would lead me in my interests in attributes of the credibility attributes of testimonial evidence. But I had one additional concern about where this analysis would lead me. This standard analysis concern whether our source P "knows" whether event E occurred of not. She tells us that event E occurred in her testimony E*. The question then is: Do we also "know" that event E occurred, just because P tells us that it did? Consider the uncertainties we face in answering this question. First, we do not know ourselves whether E occurred; we were not privy to the situation in which event E occurred or not. This is why we are asking P, who we believe to be competent. Further, we do not know how good was the sensory evidence that P obtained. Nor do we know about the actual extent to which P believed her sensory evidence regarding E, if she took this sensory information into account at all. Finally, we do not know if P is reporting in accordance with what she believes about event E. Figure 4 below is an account of the standard analysis regarding P's knowledge of event E [Part A] and how this influences our uncertainties [Part B] ; here is where the three testimonial credibility attributes arise.
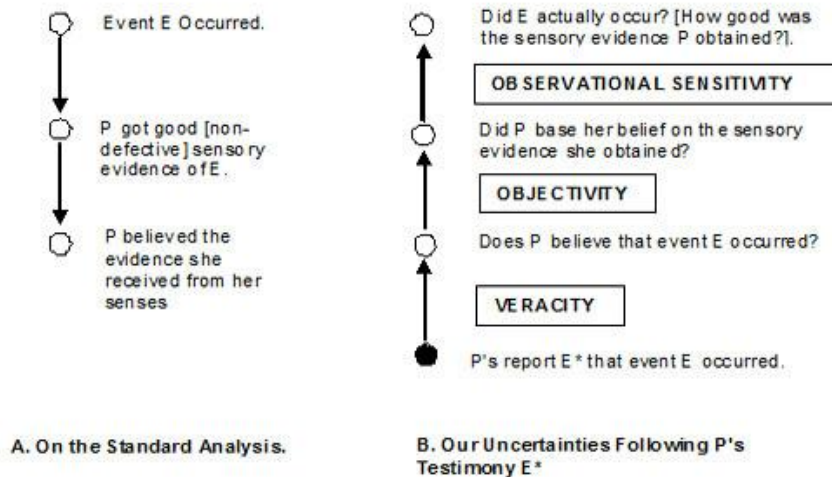

Figure 4

54

Figure 4A shows the sequence of events in accordance with the standard analysis. Event E did happen, P received good evidence for E [or "non-defective evidence" as some epistemologists say], whereupon P believes the evidence she obtained. But we know none of this; all we have is P's testimony E* that event E occurred. You should note that the chain of reasoning shown in Figure 4B is a decomposition of the credibility foundation stage shown in Figure 3 when we have testimonial evidence. As Figure 4B shows, we have three inferences to make from P's testimony E* according to this theory of testimonial credibility stemming from the standard analysis; each stage of this inference defines an attribute of the credibility of a source of testimony. How is this theory to be tested? I offer four tests; the first comes from everyday experiences we all have with persons who report on things they observe.

Experience 1. People do not always believe what they tell us. This is related to the person's veracity. We could not say that this person was being untruthful if this person believed what she/he told us. In our example, the first question we ask is: Does P really believe that E occurred, as she reported to us?

Experience 2. People do not always base their beliefs on the evidence of their senses. On occasion, we are all known to base our beliefs about the occurrence or non-occurrence of events not on what our senses report but simply because we either expect or wish certain events to occur or not occur. This is a matter of a source's objectivity. An objective observer is one who bases a belief on evidence rather than upon expectations/surmise or upon desires. In our example, the next question we ask concerning our source P is: If P does believe what she reported to us, upon what was her belief based? Did she believe what her senses told her about event E, or did she believe E occurred because she either expected or wanted it to occur?

Experience 3. Our senses are not infallible. We all make mistakes in our observations. Sometimes mistakes are due to the conditions under which our observations are made, or to our own physical condition at the time of our observation. So, the final attribute I have termed observational sensitivity [including the conditions of observation]. If the source based a belief on good sensory evidence, this would allow us to infer that the reported event did occur. So the final question is: How good were P's relevant senses at the time of her observation, and what were the conditions under which she made this observation?

But these three questions we ask about the attributes of P's credibility must be answered by ancillary evidence we have bearing upon her veracity, objectivity, and observation sensitivity. What questions should we ask about these three attributes? Answers come from the centuries-old experience in the field of law and constitute my second test of the theory of testimonial credibility I have described. In searching over fifty treatises on evidence in the field of law I found 24 questions that are asked at various times concerning attempts to undermine or support the credibility of a witness. You can find a listing of these 24 questions in other works I have written or have been a part of writing[126]. What I found especially interesting was the all 24 of these credibility-related questions can be sorted out into the three

attributes: veracity, objectivity, and observational sensitivity [including the conditions of observation]. Sixteen of these questions can be sorted into exactly one of the these attributes; four can bear on either of two of these attributes; and three can bear upon all three of these attributes. In short, I found no credibility-related questions regarding testimonial evidence that could not be sorted into one or the other of these three attributes.

My third test involves the "personal knowledge" requirement for ordinary witnesses that I mentioned above concerning our FRE 602. These Federal Rules of Evidence did not exist in Wigmore's day, but there has always been concern about what is meant by "personal knowledge". The basic problem was well expressed by Wigmore[127]:

It is obviously impossible to speak with accuracy of a witness' "knowledge" as that which the principles of testimony requires. If the law received as absolute knowledge what he had to offer, then only one witness would be needed on any one matter; for the fact asserted would be demonstrated.

When a thing is *known* (by a tribunal or other decider) to be, it *is*; and that would be the end of inquiry.

As Wigmore understood, courts and juries can never "know" whether or not a witness "knows" that the event she/he reports did occur. Wigmore interpreted the "personal knowledge" requirement to mean[128]:

The witness made a personal observation,

The witness formed an inference or a belief based on this observation, and

There was adequate observational or sensory data upon which to base this inference.

I have no knowledge about whether Wigmore ever read about the standard analysis of knowledge in epistemology. But his interpretation of personal knowledge seems like it could have been drawn from this analysis that forms the basis for my account of the attributes of testimonial credibility.

My final test of the use of the standard analysis of knowledge comes from SDT. As I noted, this theory allows the separate analysis of observational sensitivity from what I have termed objectivity factors such as a person's desires and expectations in forming beliefs based on observations. But veracity issues rarely, if ever, enter into analyses of SDT studies. The reason is that the subjects in SDT studies have no motive to be untruthful in saying whether they observed a signal or just noise. Instances in which plots of false positives exceed hits are very rarely encountered and researchers usually interpret such cases as a subject's misunderstanding instructions or simple lapses of attention.

I have three final comments to make about the three attributes of credibility of testimony identified above. The first is that second hand testimony based on other testimony involves catenations of the three stages of reasoning described in Figure 4B. The result is that we have the veracity, objectivity and observational sensitivity of each human source in a hearsay chain to consider as I have discussed in detail elsewhere[129]. In the case of opinion evidence given by a person, I have also described inferential

elements other than testimonial credibility that we encounter in evaluating opinion evidence[130].

Finally, if you are wondering whether I regard the chain of reasoning for testimonial evidence given in Figure 4B as being final in any sense, my answer is _no_. I mentioned how any chain of reasoning could be decomposed to reveal additional sources of doubt. This applies to chains of reasoning I construct as well as to those anyone else constructs. Here is an example. Suppose you examine my chain of reasoning in Figure 4B and consider the first link involving the human source's veracity. You say: "I agree that veracity of a source of testimony depends on whether this source is telling us something that she believes to have happened. But what beliefs are you considering, her beliefs while she is now telling you, or the beliefs she had a month ago when she made the observation she says she made? Human beliefs are supple or elastic. What she believes now, and what she believed at the time of her observation might not be the same".

In reply to your entirely reasonable question, I say that I have already considered such instances and can easily see how one's beliefs while giving testimony may be different from this person's beliefs at some earlier time when an observation was made[131]. There are many reasons why these beliefs at different times might be different. Lapses in memory supply one reason, but so does that fact that many events have been interposed between an earlier belief and one presently held. The person may simply have changed her/his belief about what was observed. One form of evidence concerning a changed belief involves what are termed *prior inconsistent statements*. Did this person ever tell someone else a different story than he/she is now telling us? This is one instance of an evidential test that can involve either veracity or objectivity as I have explained.

4.2.3 The Inferential Force, Weight or Strength of Evidence

Considering this final credential of evidence certainly brings Carnap's quantitative concept in science to the fore. Probability and, thus, mathematics now definitely enters our discussion of a science of evidence. You now have an item of information that you are calling evidence because you have established its relevance in the inference at hand, and you have assessed its credibility. The question you now ask is: How strongly does this evidence favour or disfavour propositions or hypotheses you are considering? I mentioned earlier that in all views known to me the force of evidence is graded in probabilistic terms. An interesting consequence is that we presently have different formal or mathematical systems of probabilistic reasoning; in each of these systems there are different views about what the force, weight or strength of evidence means. Each of these formal systems says interesting and important things about the force of evidence, but no one view says it all. I recall Jonathan Cohen once remarking that apples are measured in different ways at the market. They are sold by number, by weight, or by volume and each makes perfect sense in certain situations. As I will mention, the same is true regarding the force, weight or strength of evidence.

In my account of the emergence of the concept of evidence, I mentioned Locke's idea of the *degrees of assent* provided by evidence as we draw conclusions from it[132]. A bit later, David Hume offered a comment that has always interested me. He said[133]:

Thus all probabilistic reasoning is nothing but a species of sensation… When I am convinc'd of any principle, 'tis only an idea which strikes more strongly upon me. When I give the preference to one set of arguments above another, I do nothing but decide from my feeling concerning the superiority of their influence.

In all contexts known to me conclusions reached from evidence are necessarily probabilistic. I can think of five major reasons. Our evidence is always incomplete, commonly inconclusive, often ambiguous, dissonant to some degree, and that comes to us from sources having any gradation of credibility shy of perfection. So, in assessing the inferential force of evidence, these are things we must keep in mind. No single probabilistic view of the force of evidence that I know about accounts for all five of these considerations equally well. I have a general comment to make about the essential ingredients of the force of evidence before I consider four quite different views about how it should be assessed and graded.

Consider the general argument structure shown in Figure 3. In this simple situation depicted in the figure for a single item of evidence, I have shown a chain of reasoning consisting of credibility and relevance links. The force of evidence E* on proposition or hypothesis H depends on the strength of each link in his chain of reasoning. But you can now observe difficulties we face when we have multiple items of evidence each linked to proposition H by often lengthy and possibly interconnected chains of reasoning. Earlier, I used the current metaphor of "connecting the dots" to illustrate problems we face in assessing the overall force of masses of evidence. Figure 3 illustrates

58

how we encounter different forms of "dots". Some dots, such as E*, represent details in our evidence. Other dots represent ideas we have about the meaning of these evidential dots and come in the form of the probanda to which these evidential dots are connected to matters we are trying to prove or disprove. The point is that assessing the force of individual items of evidence is difficult enough, but when we have masses of evidence to contend with the task becomes astonishingly difficult, as Wigmore recognized many years ago.

It is not my intention to summarize the mathematical details of each of the four probability systems I now discuss as they concern the force, weight or strength of evidence. In the first place I have done so elsewhere[134]. In the second place, these details may be uninteresting, even perhaps unintelligible or unimportant, to many readers whom I have such great interest in reaching in my present comments about a science of evidence and why they should have an interest in it. The approach I will take is exactly the approach taken by Anderson, Twining and me in our recent work written essentially for law students. From experience, one way to quickly induce sleep among law students is to write an equation on the board. Fortunately, the essential elements of these alternative views can be expressed in words and in pictures as we have done in our work for law students[135]. The four views of the inferential force, weight or strength of evidence are as follows.

Bayes' Rule and the Force of Evidence

Reasoning from evidence is a dynamic process in which we revise our existing or prior beliefs about hypotheses or propositions on the basis of relevant evidence to form new or posterior beliefs about these hypotheses. There is a consequence called Bayes' rule that tells us how this dynamic process should occur when our beliefs are expressed probabilistically in accordance with three basic rules most of us learn in school. These rules are:

Probabilities are positive numbers or zero [i.e. there are no negative probabilities]

The probability of a "sure" event [one certain to happen] is 1.0,

If two or more events cannot happen together [i.e. they are mutually exclusive], the probability that one or the other of these events occurs is equal to the sum of their separate probabilities.

All probabilities are dependent, or conditional upon, what else we know or find out. Here is an event E whose probability we are interested in determining. But we also learn that event F has occurred; so we have an interest in determining the probability of E, given or conditional on F. Conditional probabilities obey the same three rules given above. There is a consequence of rules for conditional probabilities, called Bayes' rule, that tells us how much, and in what direction, we should revise or *prior beliefs* about some hypothesis based on new evidence we obtain. The result of this determination is what is called a *posterior probability*. This rule originated in the work of Thomas Bayes [1702 - 1761] a dissenting clergyman who lived in Tunbridge Wells. Those interested in finding more about Bayes can consult a recent biography of him[136]. Bayes' rule as been called the first mathematical canon for inductive reasoning.

Now, there are terms in Bayes' rule called *likelihoods* that wlll tell us by how much, and in what direction, we should revise our prior beliefs into posterior beliefs, given some new evidence. We often consider ratios of these likelihoods. In any case, these terms are indications of the inferential force of the evidence we have obtained. Figure 5 is a picture of what these likelihoods express. Suppose defendant Dick is on trial for shooting victim Vick. Here is an item of evidence received during Dick's trial: Dick owned the revolver used in the shooting of Vick.
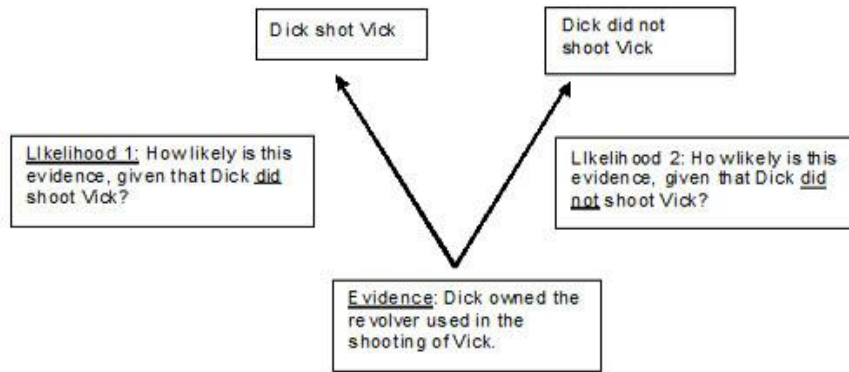
Figure 5.

The force of this evidence on whether Dick shot Vick is given by the ratio of Likelihood 1 to Likelihood 2. If you believe Likelihood 1 is greater than Likelihood 2, you are saying that this evidence favours the proposition that Dick shot Vick by an amount indicated by the size of this ratio. Likelihood ratios express both the inferential force and direction of evidence. If you said that this ratio was 5, you are saying that this evidence is five times more likely if Dick shot Vick than if he didn't shoot Vick. Directionally, this evidence points to Dick shooting Vick. Bayes' rule would say that you are entitled to increase your prior belief that Dick shot Vick by a factor of 5. Thus, in accordance with FRE 401, discussed above concerning relevance, this evidence would indeed be relevant since it allowed you to change your belief in the probability of a material or consequential issue. This is what Richard Lempert observed to be a virtue of the way Bayes' rule grades the inferential force of evidence[137].

For many years now I have studied likelihood ratio formulations for the force of every form and combination of evidence I could think of. I have reviewed many of these studies in another work[138]. Bayes' rule is a marvellous device for capturing a very wide array of evidential subtleties or complexities for study and analysis. Likelihood ratios can be expressed for collections of evidence and not only for individual items as shown in Figure 5. Bayes' rule incorporates a property called *conditional dependence* that is the finest property I know of for capturing evidential subtleties or complexities. I will return to Bayes' rule again when I discuss the discovery of evidence. We normally view probability theories as being involved just in the inductive justification of hypotheses. But application of this rule can prompt us to ask questions we may never have thought of asking. These questions may open up new lines of inquiry and new lines of evidence. But as rich as it is, Bayes' rule does not have all there is to say about the force, weight or strength of evidence.

Evidential Support and Evidential Weight:
Nonadditive Beliefs

The formal system leading to Bayes' rule rests on axioms taken by many to be self-evident. The person who first formed the three axioms I mentioned above was the Russian mathematician A. N. Kolmogorov[139]. In his works Kolmogorov makes clear that his axioms assume situations involving replicable events in which probabilities can be determined by counting. The two basic examples are *aleatory* probabilities in games of chance and *relative frequencies* in statistics. But there are many situations in which we have doubt or uncertainty about events that are the result of processes that cannot be repeated and in which no counting is possible. I refer here to unique, singular, or one-of-a-kind events. These situations are very common in a variety of contexts such as history, law, intelligence analysis, and in everyday experience. I have my own belief about the probability that Nicola Sacco was guilty of killing Berardelli, but I cannot play the world over again 1000 times to observe the number of occasions on which he did it. Various attempts have been made to apply probabilistic concepts in these non-enumerative situations.

Many persons take the view that probabilities can be epistemic, subjective, or judgmental in form and rest on whatever information we happen to have that we believe to be relevant to an assessment of the probability of interest. Applications of Bayes' rule requires at least one epistemic probability; we need an initial prior probability in order to get the dynamic probability revision process started. Regarding some hypothesis or proposition H, we need to assess how likely is H before we begin to gather relevant evidence. Many persons have no hesitation in supplying epistemic judgments of prior probabilities and other ingredients of Bayes' rule, including likelihoods, provided that these probabilities conform to the Kolmogorov axioms. As I mentioned at the close of Section 3.2, this is what led Professor Mario Bunge to refer to colleagues as "charlatans", engaged in "pseudoscience", who are willing to assess the prior probability ingredients of Bayes rule in the form of epistemic or subjective judgements. In short, Bunge and others reject any view of probability as making sense when we have nothing to count. Bunge would really come unstuck if he read the works of the person whose views of the force of evidence I now mention.

Professor Glenn Shafer [Rutgers University] has given very careful thought to epistemic or judgmental probabilities necessary in situations in which we have nothing to count[140]. He begins by denying the self-evident nature of the third of Kolmogorov's axioms; it is called the *additivity axiom*. Recall that this axiom says that if event E and F cannot occur together, then the probability that one or the other occurs is always equal to the sum of their separate probabilities. But there is an added consequence of this axiom. When we have mutually exclusive events that are also exhaustive [one or the other must occur] then the sum of their probabilities is 1.0, in accordance with the second axiom for "sure" events. Thus, if we have two hypotheses H and not-H, their probabilities must sum to 1.0, a priori, or a posteriori, given any evidence. In short, if you believe the probability of H is p, you must also believe the probability of not-H is (1 - p). Thus, if you

increase the probability if H, you must decease the probability of not-H. Shafer says this is an unfortunate property in many situations involving epistemic probability judgments. He offers several reasons why this additivity property causes trouble.

Shafer is well aware of some important ideas that have been around for a long time, such as Jakob Bernoulli's distinction between "mixed" and "pure" evidence that he described in his treatise: *Ars Conjectandi* in 1713. Mixed evidence offers some degree of support to every hypothesis being considered. But pure evidence says nothing about certain hypotheses and offers them no support at all. As an example of pure evidence, suppose that Tom, Dick and Harry are suspects in the theft of a valuable object from the home of the owner. There were no signs that the house had been broken into. Tom is found with a key to this house. This would be pure evidence since it offers support for Tom's having stolen the object; but it says nothing about Dick or Harry.

Shafer says we need a different measure of the support that evidence may provide hypotheses. So, he defines a measure of *evidential support,* S, which he equates to the *inferential weight* of evidence. Like ordinary probabilities, $0 \leq S \leq 1.0$. But S has a different meaning than do the likelihoods discussed above that indicate the force of evidence in Bayes' rule. When $S = 0$, all this means is that evidence provides no support to some hypotheses. But when a likelihood has zero value this means that this hypothesis is impossible. It says the the probability of the evidence we have is zero, given this hypothesis. Bayes' rule then assigns zero probability to this hypothesis. There is an entirely different meaning of the role of zero in ordinary probability and Shafer's S scale. On the ordinary probability scale, zero indicates impossibility or disbelief, On the S scale, $S = 0$ means lack of support or lack of belief. Our belief in hypothesis H can be revised away from zero when we do have evidence that supports it to some degree. But we cannot revise the probability of a hypothesis away from zero that has been determined to be impossible. Disbelief and lack of belief are different judgmental conditions.

There is a very important consequence associated with the manner in which Shafer support S is assigned. We are allowed to withhold support assigned to hypotheses in various ways when we cannot decide what the evidence means. As I will illustrate in an example, this characteristic of Shafer's system leads to conditions in which our beliefs are nonadditive, as they must be using Bayes' rule. Here is how S is assigned. Suppose we have some number n of hypotheses that are disjoint or mutually exclusive, but they are not necessarily exhaustive. We might think of others later on, or revise the ones we are considering at the moment. All the set of n hypotheses represents is how we see or inferential situation at the moment. Shafer call this collection of n hypotheses our frame of discernment, F. We do not assign S to just these n hypotheses by themselves, as we must do in assigning likelihoods in Bayes' rule, but we assign S to subsets of these hypotheses in our frame F. When there are n hypotheses in F, then there are 2n possible subsets of hypotheses in our frame. The set of all 2n hypotheses is called a power set. Here is the simplest case in which we have two

mutually excusive hypotheses that are also exhaustive, H and not-H. The power set of hypotheses in our frame F consists of: {H}, {not-H}, {H, not-H} and Ø, where Ø is the set of none of them [Ø is called the "empty set"]. Also, read the set {H, not-H} as: "either H or not-H". We are allowed to assign S in any way we please across the non-empty subsets of a power set, except that they must sum to 1.0 and with the additional provision that Ø always gets S = 0.

Here is an example of support assignment that involves a very good instance in which being indecisive about what evidence means, and being able to reflect our indecision in our beliefs, is a major virtue of Shafer's evidential reasoning system. This example involves William Twining's favorite law case: *Rex v. Bywaters and Thompson*, that was tried at the Old Bailey, on December 6 -11, 1922[141]. Edith Thompson was charged with either conspiring with Freddy Bywaters to kill her husband Percy Thompson on the particular occasion when he did it [October 3, 1922], or she incited Freddy to kill Percy whenever an occasion presented itself. A classic love triangle appears in this case. Freddy boarded in the Thompson's home; but he was frequently away; he worked aboard ships. Freddy and Edith became lovers and carried on their affair until the time of Percy's death. Edith and Freddy corresponded daily when Freddy was away, either through the mails or by what were then called Marconigrams. Freddy kept all of the correspondence he received from Edith, but Edith kept none of those she received from Freddy.

I have never encountered finer examples of ambiguous evidence than the letters Edith wrote to Freddy. What is clear is that these letters appear to have convinced the twelve male jurors of her guilt. She was hanged on January 9, 1923 at Holloway; Freddy was hanged the same day at Pentonville. Some of these letters mention poisons of various sorts, some mention broken glass, others contain comments suggesting that Edith had tried to kill Percy herself. Other letters seem to give the impression that Edith and Freddy had made plans to do away with Percy. But the Shakespearean scholar Professor Rene Weis [also at UCL] puts a different interpretation on her letters that he provides in a very careful analysis of Edith's case[142]. Twining and Weis agree that Edith was innocent but do so from different standpoints and using different methods[143]. Twining uses this case to give examples of the truly complex situations in which Wigmore's argument structuring methods can be employed.

Using Shafer's method for assigning evidential support, or weight, here is how I view Edith's letters as supporting her being guilty, G , or not guilty, not-G, as she was charged. Let SL represent the support I have assigned to the entire collection of her letters to Freddy. The power set of these hypotheses is: {G}, {not-G}. {G, not-G] and Ø

{G}	{not-G}	{G, not-G}	Ø
SL: 0.3	0.2	0.4	0

Here is what my S assignment means. I think the letter evidence supports her guilt to degree 0.3, and her being not guilty to degree 0.2. But I am undecided to degree 0.4 about what this letter evidence says, and so I assign this amount to the set {G, not-G} because I cannot tell whether this

ambiguous evidence specifically supports G or not-G. This setting of S represents the amount of support I have withheld from either {G} or {not-G}.

The above assignment of support corresponds to my beliefs [Bel] in a way that Shafers system allows . I have Bel{G} = 0.3 and Bel{not-G} = 0.2. My beliefs in this case are non-additive since Bel{H} = 0.3 + Bel{not-G} = 0.2 = 0.5, which is less than 1.0. If I had used a Bayesian approach, I would be required to say that Bel{H} + Bel{not-G} = 1.0 since G and not-G are mutually exclusive and exhaustive. In short, Bayes' rule does not allow me to be indecisive about what I think the evidence means.

Shafer's system, often called a system of *belief functions*, is very useful in capturing elements of our probabilistic beliefs that are difficult, or impossible to capture with ordinary probabilities. Because one of the Kolmogorov axioms is violated, Bayes' rule does not appear is Shafer's belief function system. It is replaced by what is called *Dempster's rule*, which allows us to combine support assessments S for successions of evidence. This rule allows to calculate what is called the *orthogonal sum* of S assignments for different items of evidence. This system has found application in a number of important contexts in which epistemic judgmental assessments are necessary in the reasoning tasks at hand.

Evidential Completeness and the Weight of Evidence in Baconian Probability

Francis Bacon [1561 - 1626] is usually credited with being the first to argue that we can never justify hypotheses about how nature works just by compiling instances that are favourable to them. What he argued was that negative instances are at least as informative as positive instances. In fact, what we should do in testing hypotheses is to perform experiments designed to eliminate possible hypotheses. The one or ones that resist our best efforts to eliminate any of the hypotheses we are considering are the ones in which we should have the most confidence. This view has been called *eliminative induction*. But Bacon was never specific about what eliminative methods could be employed. As I noted in Section 3.1, John Stuart Mill is usually credited with being the first to identify methods designed to eliminate possible causes for the effects we observe in nature. But such methods were known much earlier to the four Oxford scholars I mentioned.

But there is another important element in the eliminative testing of hypotheses. The tests we perform must be variative in the sense that we must establish the array of conditions under which we may expect a hypothesis to continue to remain valid. We cannot do this by performing the same test over and over again. The only thing this repetitive testing would accomplish is to increase our confidence in the reliability of this single test's results. The more varied are the conditions under which some hypothesis holds, the more confidence we can place in it. But this variative testing raises another important question, namely: how complete has been our eliminative testing of our hypotheses? There may be other important tests of our hypotheses that we have not performed whose results might serve to eliminate hypotheses we are still considering.

Neither Bacon, Mill, Popper, nor anyone else was successful in relating problems associated with the eliminative and variative testing of hypotheses to ordinary probabilistic concepts. The first person to study this relation was L. Jonathan Cohen [now emeritus, Queen's College, Oxford]. In a work that had a great influence on probabilistic thinking in law and philosophy, Cohen was the first to generate a theory of probability expressly congenial to the eliminative and variative testing of hypotheses[144]. He refers to this theory as *Baconian probability* to acknowledge its roots in the works of Francis Bacon. On occasion, he also calls it a theory of *inductive probability*. In his works Cohen takes a decidedly ecumenical [or "polycriterial", as he calls it] view of probability in evidence-based reasoning. He allows that conventional views of probability make perfect sense in some but not all situations. He further argues that conventional views of evidence-based reasoning, such as Bayes' rule, overlook *how much* evidence has been considered and how complete is this evidential coverage of matters believed to be relevant in the inference at hand. Eliminative and variative inference requires special considerations. In fact, evidential completeness, in Cohen's view, is the major factor associated with the weight of evidence.

In Figure 6 below is a diagram I have used to illustrate some of Cohen's key ideas in Baconian probability. I have tried my best to generate interest in the importance of Cohen's views among persons in a variety of contexts

who should be aware of his ideas regarding evidential completeness. I have gone to great lengths in some contexts, but not always with any great success[145]. There are two basic questions that arise in Cohen's views about the weight of evidence: (i) How much uncounteracted favourable evidence do we have on some hypothesis that has arisen in answer to relevant questions we have asked?, and (ii) How many relevant questions, that we know about, remain for which we have no evidential answers? In short, the weight of evidence in Cohen's view depends not only on answers to questions we have asked, but also upon how many questions remain unanswered. Cohen's Baconian views about the weight and amount of evidence bring to mind ideas expressed by John M. Keynes in his very influential treatise on probability[146]. Keynes's ideas about the amount and the weight of evidence have often been misunderstood. Cohen has written on various questions that have arisen regarding the views of Keynes on the weight of evidence[147].

Here are some details of the cover story surrounding Figure 6. Some time ago we were asked to assess which of three hypotheses, H1, H2 and H3, is most likely because it will have an important bearing on a decision we must make. Initial evidence pointed very strongly to H1 being true, so we took an action based on H1. What we are now doing is engaging in a post mortem analysis trying to see what went wrong; H3 happened to occur and our decision miscarried. Our decision produced a disastrous result. Someone says: "How could we have gone wrong? We used Bayes' rule to aggregate what our assessments of likelihoods for the evidence we had, and we all agreed that the prior probabilities we were using made perfect sense. Bayes' rule said that the posterior probability of H1 was 0.997 based on the evidence we incorporated in our inference".

If Jonathan Cohen happened to be present during our post mortem, here is what he might have said: "You were out on an inferential limb that was much longer and more slender than you believed it to be, just based on the answers your existing evidence provided. How many relevant questions do you now realise to have been unanswered in your analysis?" We begin to make a list of questions we believe also relevant that we did not attempt to answer; this list grows quite large. It also contains questions we knew about at the time of our analysis. However, we believed the evidence we did take account of was sufficiently strong that we did not hesitate to conclude that H1 was true. Here is a picture of the actual inferential limb we were on.
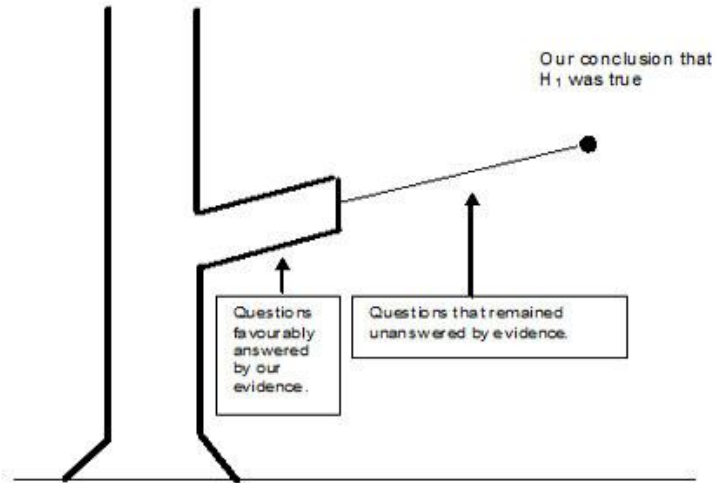
Figure 6

Jonathan Cohen goes on to explain the two parts of this inferential limb on which we found ourselves. He says: "The strong part consists of the evidence you had that was favourably relevant to H1. The weak part consists of relevant questions that remained unanswered. What you did in concluding that H1 was true was to assume essentially that the answers to all of the questions that you did not ask would have been favourable to H1. The problem is that a very high Bayesian posterior probability is not a good indicator of the weight of evidence because it does not grade the completeness or sufficiency of evidence".

In another work I have compared Baconian and Bayesian approaches when we encounter chains of reasoning in arguments we construct[148]. There is nothing incompatible about these two approaches to evidence based reasoning. The reason is that they each respond to different, but equally important, considerations. Bayes' rule provides very useful measures of how strong is the evidence you do have, but Cohen's Baconian probabilities allow us to grade the completeness of our evidence. I ended up concluding that both forms of hedging conclusions would be necessary on many occasions.

Verbal Assessments of the Force of Evidence: Fuzzy Probabilities

In so many situations we talk about the force of evidence, and express the strength of our conclusions, in words rather than in numbers. There are no better examples than those occurring in the field of law. Forensic standards of proof such as, 'beyond reasonable doubt", "clear and convincing evidence", "probable cause", and so on, are verbal assessments that seem to defy efforts to translate them into numerical probabilities. In his analysis of what we now call inference networks, Wigmore understood perfectly well that the arrows linking evidence and probanda, such as those illustrated in Figure 2A, are probabilistic in nature. But he always used words rather than numbers to indicate the force with which one element of an argument is linked to others[149]. He used terms such as "strong force", "weak force", and "provisional force" to indicate the strength of these linkages. The use of words rather than numbers to indicate the force of evidence appears in many other contexts, especially when there is no attempt to employ and combine any of the views of evidential force described above.

There are algorithms for combining numerical probabilities, such as Bayes' rule and Dempster's rule, but how do we combine assessments of the force of evidence that are given in words? Wigmore gave no hint about how we should combine his verbal assignments of evidential force in order to grade the force of an overall mass of evidence. Verbal assessments of probabilities in grading the force of evidence, and in stating the strength of an overall conclusion, are today referred to as *fuzzy probabilities*, in part to acknowledge their imprecision. But thanks to the work of Lotfi Zadeh and his many colleagues worldwide, there is a logic that underlies the expression and combination of verbal or fuzzy probabilities[150]. This system of fuzzy logic and probabilities has found wide acceptance in many situations in which persons must perform a variety of tasks based on fuzzy or imprecise ingredients. But it does have its detractors[151].

I have now completed my comments on the essential properties or credentials of evidence: relevance, credibility, and inferential force, weight, or strength. I have taken some care in discussing these properties in order to illustrate how study of them involves the classificatory, comparative, and quantitative concepts that both Poincaré and Carnap said were involved in science. I next comment on the uses of evidence and will show how these same concepts arise.

4.3 On the Uses of Evidence

We all use evidence every day of our lives in connection with our inferences and decisions, whatever their substance and objectives might be. William Twining has provided a characterization of evidence that seems to cover the use of evidence in any context you can think of. He says[152]:

'Evidence' is a word of relation used in the context of argumentation (A is evidence of B). In that context information has a potential role as relevant evidence if it tends to support or tends to negate, directly or indirectly, a hypothesis or probandum. One draws inferences from evidence in order to prove or disprove a hypothesis or probandum. The framework is argument, the process is proof, the engine is inferential reasoning.

I am going to provide two examples of uses of evidence. The first will illustrate Poincaré's assertion that science relies upon classifications and is the study of relations, some of which can be expressed in quantitative terms. The second involves Carnap's comparative and quantitative concepts and their importance in science and in our everyday lives.

4.3.1 On the Inferential Roles of Evidence

I focus now on Twining's comment given above and his phrase: …"information has a potential role as relevant evidence if it tends to support or tends to negate, directly or indirectly, a hypothesis or probandum". Wigmore has given us a very useful classification of the roles evidence can play in the context of proof[153]. In discussing these evidential roles, Wigmore used terms that are encountered in adversarial processes in our system of laws. But I can make them quite general in their application to inferences in any context. I recall here Stephen Toulmin's comment on adversarial processes and the case we try to make from evidence in any context, and his saying that "logic is generalized jurisprudence"[154].

To make the following comments as general as I can I am going to assume just one person making an inference, namely you. We don't care about the context or substance of this inference. You are trying to reach a conclusion concerning whether hypothesis H is true or not. We might suppose that you have proposed hypothesis H. The only thing we will assume is that you are open-minded and are willing to consider evidence bearing upon both of the hypotheses: H and not-H. The evidence you will encounter may be tangible or testimonial in form and you may receive it as a result of your own efforts as well as the efforts of others from whom you may request information. You will note in this example that several of the forms, mentioned in Figure 1, and combinations of evidence that I said were "substance-blind" will arise in this example. This will satisfy Poincaré's emphasis on the importance of classification in science.

Supporting Hypothesis H.

First suppose you believe that if event E occurred, it would be directly relevant but not conclusive evidence favouring H. You then find evidence E*, that this event E occurred. This evidence may be a testimonial assertion or an item of tangible evidence. In Wigmore's analysis this would be called a *proponent's assertion*, if we regarded you as the "proponent" of hypothesis H. You are concerned of course about the credibility of the source of E*. Suppose you obtain ancillary evidence favourable to the credibility of this source. This would enhance your belief that event E did occur. Suppose, in addition, you query another source who/that corroborates what the first source has said; this second source also provides evidence that event E occurred. You then gather ancillary evidence that happens to favour the credibility of this second source. But you also gather evidence in support of the generalization that licenses the relevance linkage between event E and hypothesis H. This evidence would strengthen this linkage.

But you know about other events, if they occurred, would also support hypothesis H. In particular, you believe that event F would converge with event E in favouring H. You gather evidence F*, that event F occurred. You might gather ancillary evidence favourable to the credibility of the source of F, and you might also gather corroborative evidence from another source that event F occurred. You might, in addition, gather further ancillary evidence to strengthen the generalization that licenses your inference from event F to hypothesis H. So, now you have two lines of evidence which,

together, would increase the support for hypothesis H. You may of course know of think of other events that would converge in favouring H.

Wigmore was no probabilist, that would have been asking too much of him, given all of his other accomplishments regarding evidence and proof. Study of probabilistic matters concerning the inferential force of evidence suggests additional ways in which we you can use evidence to support hypothesis H. First, suppose that evidence E* is testimonial and came from a source named Mary. You begin to think that the fact that Mary told you that event E occurred means more than it would if someone else had told you about E. In fact, what you know about Mary's credibility makes her testimony more valuable than if you knew for sure that event E occurred. In short, what we know about a source of evidence can often be at least as valuable as what the source tells us. I have captured this subtlety associated with testimonial evidence mathematically in a recent report[155].

Another subtlety, this time involving events E and F. can be captured that would often greatly increase the support that evidence of these events could provide for hypothesis H. Earlier in Section 4.1.2 I mentioned how items of convergent evidence, favouring the same hypothesis, can often be synergistic in their inferential force. You might believe that evidence of events E and F, taken together, would favour hypothesis H much more strongly than they would if you considered them separately or independently. This synergism can be captured in probabilistic terms[156].

These are all roles that evidence can play in support of some hypothesis. But we agreed that you are open-minded and will carefully consider counterevidence that would negate H or would favour not-H.

Negating Hypothesis H

A colleague appears who has never shown any great enthusiasm for a belief that hypothesis H is true. So far you have said that events E and F would favour H. But your colleague challenges the credibility of the sources of evidence about these events by first producing ancillary evidence that disfavours their credibility. In addition, your colleague might produce contradictory evidence from sources who/that will say that either or both of the events E and F did not occur. Wigmore referred to this evidence as constituting *opponent's denial* that events E and F occurred.

But your colleague might instead, or in addition, have ancillary evidence that weakens the generalizations you have asserted that link events E and F to hypothesis H. Such evidence would tend to *explain away* what you have said was the significance of events E and F. This ancillary evidence would allow your colleague to say: "So what if events E and F did occur, they have little if any bearing on hypothesis H". Wigmore termed this situation *opponent's explanation*.

Your colleague has yet another strategy for negating H. She might say the following: "So far you have only considered events [E and F] that you say would favour H. Are you only going to consider events you believe favour H? I have gathered evidence J* and K* about events J and K that I believe disfavours H. Wigmore termed this situation *opponent's rival evidence*. In this situation you would have to cope with what I termed

divergent evidence. There is no contradiction here since J and K involve different events and might have occurred together with events E and F. Events J and K simply point in a different F.

inferential direction than do events E and F.

But your colleague has a final approach to undermining an inference that H is true; it involves what you might have said about the synergism involving events E and F. Your colleague says: "You have said that evidence about events E and F taken together have much more force than the would have if we considered the separately. In other words you are saying that event F has more force in light of the occurrence of E than it would have if we did not consider E. But I argue that the occurrence of E would make F redundant to some degree and so I argue that they mean less when taken together than they would do when considered separately.

In providing this illustration of the various roles evidence plays, it was not actually necessary for me to suppose that you experienced a colleague who would use these evidential strategies for undermining a belief in H. If you were indeed open minded in your inferential approach, you would have played your own adversary by considering how evidence favourable to H could be attacked or countered in these various ways by other evidence. Your conclusion may well have been that not-H is true in spite of your initial belief that H was true. What this example illustrates is the necessity for us to be unbiased or objective in the gathering and evaluation of evidence in the inferences we make from evidence. We might say that we are well served when we play the role of our own "loyal opposition" in the inference tasks we face. This is true even if we often face the often "not so loyal opposition" from our critics.

4.3.2 Stories from Evidence and Numbers

I begin by acknowledging the many studies currently being undertaken in which masses of evidence and complex processes are being analyzed in probabilistic ways in a variety of important contexts including law, medicine and intelligence analysis. Consider Figure 2 again that shows two very simple illustrations of the different forms of inference networks that have been analysed probabilistically. I will use both the Wigmore analysis of a mass of evidence and a certain process model to illustrate how it is possible, and usually necessary, to construct alternative stories that might be told about the inferential force of a mass of evidence. My examples will involve the use of Bayes' rule, but similar analyses can be performed using Shafer's belief functions or Cohen's Baconian probabilities.

Both of the examples I will discuss involve what is termed "task decomposition" or "divide and conquer". In such decompositions, an obviously complex inference task is broken down into what are believed to be its basic elements. Wigmore's analytic and synthetic methods of analyzing a mass of evidence is a very good example. We first list all of the evidence and sources of doubt we believe appear in arguments from evidence to what we are trying to prove or disprove from it [a key list], and then construct a chart [or inference network] showing how we believe all of

these pieces fit together. As I have mentioned, we can describe this process as one of trying to "connect the dots".

Suppose we have a mass of evidence in our analysis and an inference network based on this evidence that has survived a critical analysis designed to uncover any disconnects or non sequiturs in the arguments we have constructed. The next step is to assign probabilities that will indicate the strength with which the probanda, sources of doubt, or probabilistic variables are linked together. These probabilities come in the form of the likelihoods I described in Figure 5. All the arrows in the two diagrams in Figure 2 indicate these probabilistic linkages expressed in terms of likelihoods. Let us suppose that we all agree that the inference network we have constructed captures the complex arguments or elements of the process we are studying. But where do these linkage probabilities come from? In some very rare instances we might have a statistical basis for estimating these probabilities from relative frequencies. But in most instances many or most of these probabilities will rest on epistemic judgments we make. In the two examples I will provide, <u>all</u> of the probabilities rest on subjective judgments. Here is where the necessity for telling alternative stories arises.

Though we agree about the structure of our inference network we may find ourselves in substantial disagreement about the likelihoods linking elements of our argument together. Suppose we are interested in determining the overall force or weight of the evidence we are considering. How do our differences in these likelihood ingredients affect the force of the evidence we are considering? We might say that our different beliefs about these probabilistic ingredients allow us to tell different "stories" about the force of the evidence we are considering. The "actors" in our stories consist of the items of evidence we have. The "plots" of our stories are provided by the likelihoods we have assessed. When your likelihoods are different from mine we are essentially telling possibly different stories based on the same evidence, or involving the same actors. We are of course interested in the extent to which our stories end in telling us about the force of the evidence we have. How your story ends may be quite different from the ending of my story, but not necessarily. The metaphor of telling stories from evidence is certainly appropriate. It describes a process that is repeated every day involving the different stories told from the same evidence by opposing attorneys in trials at law. They use the same actors to tell different stories.

How do we tell how our different stories about the force of evidence will end? This is where mathematics comes to our assistance. It involves the process of what is termed *sensitivity analysis*. We have equations stemming from Bayes' rule that tell us how to combine your likelihoods and my likelihoods in calculations of the force of all the evidence we are considering. In short, these equations supply endings to your story and my story about the aggregate force of the evidence. Here comes Carnap's comparison and numerical concepts again in a science of evidence. Our stories are told numerically but they can easily be translated into words. Then we can compare our stories to observe the extent to which our differing likelihood ingredients have affected determinations of the force of the evidence.

But there is an important characteristic of the equations we use to combine these likelihood ingredients; they are all <u>non-linear</u>. What this means is that these equations can produce many "surprises" that would never result from linear equations in which "the whole is always equal to the sum of its parts". What will happen is that, on some occasions, the fact that our assessed likelihoods are quite different makes little difference in the ending of our two stories; we are telling two stories that have the same or nearly the same ending. But on other occasions even exquisitely small differences in our likelihood ingredients will produce drastic differences in our stories about the force of our evidence.

Jay Kadane and I used the process of sensitivity analysis just described in our probabilistic analysis of parts of the evidence in the case of Sacco and Vanzetti. We used this process to tell different stories on behalf of the prosecution and of the defence in this case[157]. As I mentioned above, we used the same evidence in telling these stories on behalf of the parties involved in this trail. Here are two examples of the kinds of stories we told and how we told them.

The first involved what Wigmore termed *concomitant evidence,* that involved what Sacco was doing at the time of the crime. Two of the witnesses I have already mentioned, Lewis Pelser and Lewis Wade. They were the prosecution's "star" witnesses. Recall that Pelser said he saw Sacco at the scene of the crime when it occurred, and Wade said he saw someone who looked like Sacco at the scene of the crime when it occurred. But there were five defence witnesses, four of whom were just across then street from where the payroll guard Berardeiil was shot. They all said the neither Sacco nor Vanzetti was at this scene when the shooting occurred. A fifth defence witness testified that Sacco was not at the scene about 15 minutes before the crime occurred.

The Wigmorean argument structure for the evidence just described is simple enough so that were able to write out the exact equations necessary for combining the likelihoods in this argument structure. Using these equations we told ten different stories, five on behalf of the prosecution and five on behalf the defence[158]. In the process of telling these ten stories we varied the value of likelihoods such as those that concern the credibility of the witnesses, and other matters such as those concerning the probability of someone looking like Sacco being at the scene, given that Sacco was not there. The ten stories we told all have different endings as far as the weight of this evidence favouring either the hypothesis that Sacco did shoot the payroll guard or the hypothesis that Sacco did not shoot the payroll guard. As I noted above, these stories often have surprising endings. For example, by varying Pelser's credibility by apparently just a small amount we can effectively destroy the contribution of his testimony to the weight of the combination of evidence being considered. Some of the stories we told illustrated how Bayes' rule captures the possible redundance of Pelser's and Wade's testimonies.

We also considered much more complex aggregates of the evidence in this case[159]. One combination of evidence concerned Bullet III I mentioned earlier and the Colt automatic alleged to be the one Sacco had in his

possession when he was arrested. The argument structure here is sufficiently complex that it prevents us from writing out equations for the force of evidence. The inference network we constructed for this ballistics evidence requires 42 assessments of likelihoods. But thanks to the truly outstanding works of several probabiliists, we now have computer-based systems that "know" what the necessary equations should be for any appropriate argument structure [a DAG] we can construct, whether it is a Wigmore evidence chart or a process model such as the ones I illustrated in Figure 2[160]. The stories we told on behalf of the prosecution and defence were designed to capture important conditional dependencies among probanda or propositions appearing on our inference network. For example, evidence that Bullet III was fired through Sacco's Colt points toward Sacco's guilt, but not very strongly. Someone else may have fired this weapon during the crime. But this evidence means more when we take into account other evidence that Sacco fired a weapon at the scene of the crime when it was committed. The credibility of all of this evidence is so important.

We were pleased to note that the stories we told from numbers in the Sacco and Vanzetti case were included in a work by John Allen Paulos, a mathematician whose books enjoy a very wide following. In a recent book on what he terms the hidden mathematical logic behind stories[161], he describes the usefulness of the stories we told from numbers in the Sacco and Vanzetti case and argues that the Wigmorean analysis underlying these stories would have been useful in other complex cases such as the trial of O. J. Simpson.

I make one final point about telling stories from numbers based on inference networks constructed from evidence. I have said nothing so far about the role of experiments in a science of evidence. The process of sensitivity analysis is a form of experimentation in which we vary the probabilistic ingredients of equations based on a given inference network. We do so in order to see how the equations will behave [i.e what different stories they will tell] in response to these changes in their ingredients. But do all of the stories told based on an inference network make sense? Sensitivity analysis is also a process for testing the inference network itself. As I have mentioned, inference networks we construct are products of our imaginative and critical reasoning. How do we test to see if the network we have constructed makes sense in allowing us to draw conclusions of interest to us? Sensitivity analysis allows one kind of test: Does our network allow us to tell stories that make sense when we vary the ingredients of these stories in more or less systematic ways? This form of experimentation is one I have used for years in testing the inferential consequences of arguments I have constructed from a wide variety of forms and combinations of evidence[162]. A science of evidence does allow experimental tests of our views about how we believe evidence to be related to matters we are trying to prove or disprove.

4.4. Discovery in the Science of Evidence

The third OED definition of science, that I mentioned in my opening remarks in Section 4.0, is restrictive in the sense that it makes necessary for a science to have "reliable methods for the discovery of new truths in its own domain". I have two responses here, the first concerns how we extend our knowledge <u>about</u> evidence and its properties and uses; this involves how we learn more about evidence itself. The second concerns how a science of evidence can provide methods for enhancing discovery-related activities in any discipline or context in which this science may be applied. So in this sense we have both basic and applied interests to consider as far as concerns discovery in the science of evidence.

4.4.1 Discoveries about Evidence

In Section 3.0 I considered how the concepts of *evidence* and of *science* have emerged and changed over the ages. At no point did I argue that these two concepts are now fixed for all time. Thus, a science of evidence is in the process of emergence in which we hope to learn more about evidence as we continue to study it. For example, I have offered a categorization of recurrent forms and combinations of evidence that I advertised as being "substance-blind" only in the sense that it tells us what kinds of evidence we have. But I never advertised that my classification could not be improved or that other useful classifications could never be made. You might, for example, think of a type of evidence that I have left out of the rows in Figure 1. If so, please tell me about it.

What we wish to avoid are problems the philosopher Imre Lakatos identified in connection with what he called *research programs*[163]. He called research programs *progressive* to the extent that they continue to have heuristic power in generating new facts about the phenomena of interest. If they fail to do this, and just continue to account for already observed facts, Lakatos called them *regressive* research programs. Given the richness of the contexts in which evidence is used and studied, we run little risk of a science of evidence becoming regressive any time soon, provided that we share our experiences and thoughts about evidence with each other. Phil Dawid's emphasis in his proposal to the Leverhulme Foundation emphasized an <u>integrated</u> science of evidence. I have always interpreted an integrated science of evidence to involve the necessity of identifying and exploiting our varied experiences and thoughts about evidence. I will give just two examples of how such sharing has benefited me in my work on evidence. I return to the integrated nature of a science of evidence in Section 5.0.

I have always marvelled that there are so many important evidential and inferential subtleties or complexities that often lie just below the surface of even the "simplest" of evidence-based reasoning tasks [if there are any such things as "simple" inferential tasks]. I spent many years of my work on evidence trying to identify and understand these subtleties so that they could be exploited in the conclusions we must draw from evidence. These subtleties involve any or all of the evidence credentials I have mentioned: relevance, credibility and inferential force or weight. Many years ago I recognised that nearly every law case I read revealed some new subtlety I

should examine. Many of those I have studied were first revealed to me in law cases, some that took place centuries ago.

But I have often observed subtleties revealed in the works of others interested in the study of various kinds of evidence. As an example, I now refer to the work of C. A. J. Coady, a philosopher interested in the study of testimony[164]. As expected, Coady carefully examines the epistemological foundations for testimonial evidence. But one chapter of his work was especially interesting to me, his Chapter 10 entitled; *Astonishing Reports*. This chapter concerns testimonial evidence of events either contrary to the laws of nature, or testimony about unusual events that would conform to these laws. The study of testimony about miracles has occupied the attention of probabilists and others for centuries[165]. And it is still a still a topic of research today[166]. I hope to give further study to these topics in future. In my probabilistic studies there are very interesting terms appearling that concern the rareness or improbability of the events reported in testimony. Since the 1600s it has been expected that the rareness of the event reported, in addition to the credibility of the source reporting the event, are factors that determine the force of testimonial evidence about this event. But the relationship between these ingredients was never made clear in any studies I could find. I am just vain enough to conclude that my analysis of this rareness-credibility problem makes the nature of this relationship clear. It turns out that I was "murdered" for doing so[167].

4.4.2 Evidence Science and the Discovery of New Evidence

I understand that the discovery of new evidence depends on the nature of the investigations that take place in the substantive area of concern. The evidence we generate or discover in any context depends upon the questions we ask. People in different contexts will naturally ask different questions. Thus, a historian, a sociologist, an accident investigator, and a physician will ask different questions and will generate the different kinds of evidence of interest to them. The path to expertise in any area depends in no small part on knowing what kinds of questions to ask and how answers to them might be obtained. It is also true that persons having different standpoints will ask different questions in the same context. Suppose a police officer, a news reporter, and an attorney all arrive at the scene involving an injured or a dead person. The police officer will ask whether a crime has been committed. The reporter wil ask whether there is a story here. And the attorney will ask whether there is a civil complaint or a criminal charge that can be levied in a legal action of some sort.

But it is also true in so many situations involving discovery or investigative efforts that we have hypotheses in search of evidence at the same time we have potential evidence in search of hypotheses. It turns out that we can generate new evidence in both of these situations. I have already noted that from hypotheses we generate new evidence and new lines of reasoning. But new evidence can be generated by several other ways as I now illustrate.

Generating Evidence from Argument Construction

I now consider the situation when we have potential evidence in search of hypotheses. I say "potential" here because we must be able to link information to some hypothesis before we can call it *evidence*. Charles S. Peirce referred to the imaginative reasoning involved in the generation of hypotheses from our observations as *abductive reasoning*. From some item of information, or combination of items, we generate a hypothesis to explain the observations we have made. I have often thought that Peirce's writings on abductive reasoning lead one to conclude that the generation or discovery of a new hypothesis always takes place in one glorious episode of such reasoning. In another work I have identified sixteen species of abductive reasoning that I believe capture more of the complexities of the process of discovery or investigation as it is played out over time[168]. These sixteen species result from four levels of the actual creativity of a hypothesis as identified by Umberto Eco, and Paul Thagard's four classes of matters that are to be abduced. On this view, discovery takes place as we mix together these various species of abductive reasoning with other steps involving inductive and deductive reasoning.

What I wish to show is how Wigmore's methods of argument construction, however cumbersome they might seem, also provide an elegant means for generating new evidence. I consider two cases. First, suppose you have just abductively generated a hypothesis you believe accounts for an observation you have made. In order to convince someone that this hypothesis does in fact account for this observation, you would need to construct an argument showing why this new hypothesis does so. In another case, suppose you have already generated some hypothesis from other observations. But you have a new item of information you are attempting to show is relevant to this hypothesis. In either case, the argument you construct involves further stages of imaginative or inductive reasoning. Recall that the interim probanda that your argument contains, however many of them there are, each represents a source of doubt about some proposition. Each source of doubt represents a new line of evidence you might be able to gather.

As an example, have a look at the chain of reasoning shown in Figure 3. Your argument from evidence E* concerning event E involves sources of doubt about events F and G. You identified these stages of your argument abductively. Thus, in constructing this argument you have identified two new items of evidence that you may be able to gather. You should notice that evidence regarding events F and G would be more direct on hypothesis H than is your evidence about event E. In short, arguments from evidence can allow you to generate additional evidence.

Evidence Marshaling and Discovery

The science of evidence also now includes study of ways in which we organize or marshal our thoughts and our evidence during the process of discovery or investigation. The argument for such study is quite simple. The manner in which we organize or marshal our *existing* thoughts and evidence strongly influences how successful we will be in generating *new* thoughts

and *new* evidence. Discovery or investigation in any area is a complex activity that unfolds over time. We learn different things at different times and we may begin an episode of discovery with different amounts of information. On some accounts it has been said that most of what discovery involves is using sophisticated methods for search[169]. The trouble is that there are so many investigations that begin with our having nothing to search. The problem is unique and there is no background of information relevant to this unique situation. We begin to generate thoughts and evidence by asking questions. It seems that, in any case, the process of inquiry is at least as important as our methods for search. Studies we have performed on evidence marshaling are designed, in part to stimulate the process of inquiry as we generate hypothese and evidential tests of them.

Professor Peter Tillers [Cardozo School of Law] and I began studies of how we might marshal our thoughts and our evidence during the process of discovery or investigation. In several works we have given an account of a linked network of different marshaling operations[170]. It became quite obvious in the early stages of our research that there will be no single method for evidence marshaling that will satisfy all the intellectual demands placed upon an investigator during the process of discovery. Every episode of discovery is unique and proceeds at a different pace. We learn different things at different times in response to the questions we ask. Our network of marshaling operations consists of five tiers containing 15 different operations[171]. Where one starts using this network depends, in part, on how quickly the pace of discovery proceeds. One of the marshaling operations has relevance only in the field of law; all the others are relevant to discovery in any context. This system is now being incorporated in the field of intelligence analysis. It forms a major step in our efforts to be more adept at "connecting the dots".

Mathematics and the Discovery of Evidence

I know that there will be many readers who will never contemplate using any of the formal or mathematical methods I mentioned in Section 4.2.3 for grading and combining the inferential force or weight of evidence. I also anticipate that the methods I have described for inference network construction may not be appealing to everyone. One reason is that such methods rest upon very careful argument construction that is as laborious as it is useful. In might be thought that the only use for probabilistic methods, applied in such study, is to allow the sensitivity analysis I described in which we tell various stories about the inferential force or weight of the evidence we have. But such mathematical analyses can do much more than this.

Well-posed equations for the force or weight of evidence can be especially important during the process of discovery as it proceeds in some inferential problem of interest. What happens is that the equations themselves can suggest questions you can ask that you might never have thought about asking if you had not done this analysis. These questions may suggest new lines of reasoning or interpretations of your evidence and the argument you are constructing. I have found this out myself in many studies

I have performed concerning the force or weight of various forms and combinations of evidence. As an example, my identification of the conditions under which testimonial evidence can have more force than knowing for sure that the event testified did indeed occur came as a result of the mathematical analysis I performed[172]. In short, mathematics can have heuristic value in studies of evidence as it has in other areas of science.

4.5 A Stronger Definition of a Science of Evidence

I have now gone to some lengths in my attempt to show that a science of evidence goes well beyond the three weaker definitions of the word *science* as given by the OED. All these three definitions say is that science involves study, in a recognized department of learning, that leads to an organized body of knowledge. In my account of studies of the properties, uses and discovery of evidence, I have shown how the three concepts that have been associated with science are regular features of such studies: classification, comparison and quantification. In the process, I have attempted to show that studies of evidence involve elements of the two stronger definitions of science given by the OED. Evidence can be systematically classified with respect to the kinds of evidence we encounter, that reliable methods for the discovery of new truths about evidence are possible, and that there are both intellectual and practical ends that are served by systematic studies of evidence.

I have also attempted to go beyond what the OED says scientific activity involves. Theories or hypotheses offered in explanation of phenomena of interest are necessary features of science together with methods for testing these theories or hypotheses. In my account of the three major credentials of evidence [relevance, credibility and inferential force or weight] I have offered alternative theories concerning each of these three credentials. But the testing of these theories or hypotheses will not rest upon conventional empirical tests. For example, we cannot conduct conventional experimental tests of the four views of the inferential weight or force of evidence to see which one is "best". The reason is that each of these four views "resonates" to different attributes of the task of assessing the force or weight of evidence. I have long believed that evidence-based reasoning is far too rich an intellectual activity for us to be able to capture all of this richness in any single axiom-based system any of us is likely to construct. I noted that on some occasions it will be advisable to weigh evidence using more than just one of these views. However, on some occasions we can tell stories based on some of these views and ask: do the stories being told make sense? The testing here seems to be logical rather than empirical in nature.

I have also mentioned the various ways in which mathematics can enter a science of evidence when we consider the task of assessing the force or weight of individual items or masses of items in drawing conclusions. The mathematics for doing so is at hand, whether it will be used in any context is another question. Persons in many disciplines do quite well in expressing the intensity of their beliefs about the force of evidence of interest to them without resorting to specific numerical expressions of such intensity. But, as I have noted, mathematical expressions for the force of evidence have

heuristic value in suggesting additional questions we might ask about and of our evidence.

I return briefly to where I started in Section 2.0 with Israel Zangwill and his view that the science of evidence is the "science of science". There I mentioned that others have offered views about what a science of science should involve. I now quote from the biologist Ernst Mayr whose views of a science of science will lead me to matters I will mention in Section 5.0. Mayr said[173]:

Increasingly often one reads references to a 'science of science'. What is meant by this designation? It relates to an evolving discipline that would combine the sociology of science, the history of science, the philosophy of science, and the psychology of science with whatever generalizations one can make about the activities of scientists and about the development and methodology of science. It would also include generalizations on the intellectual growth and style of work of the great leaders of science and, for that matter, also of the great army of other scientists who make contributions to the gradual progress of our knowledge and understanding.

I take no position on the issue of whether a science of evidence is also the science of science, as Zangwill claimed. But it seems that Mayr's comments apply equally well to a science of evidence in emphasizing how many persons and disciplines are naturally involved in it. I now turn to the multidisciplinary and integrated nature of a science of evidence and attempt to answer a question I posed at the outset: Who should care about a science of evidence?

5.0 AN INTEGRATED SCIENCE OF EVIDENCE

I begin with a quotation I have used on other occasions; it comes from the French historian Marc Bloch[174]. Bloch said:

Each science, taken by itself, represents but a fragment of the universal march toward knowledge. …. in order to understand and appreciate one's own methods of investigation, however specialized, it is indispensable to see their connections with all simultaneous tendencies in other fields.

So it is with our march toward a greater knowledge and understanding of evidence. We all make specialized uses of evidence and study it in ways that appeal to us. But we would be foolish indeed to ignore what others have said about evidence, their experiences with it, and their methods of studying it. Bloch's thoughts are as good as any I can think of to set the stage for my remarks on an integrated science of evidence in which we actively share our thoughts and experiences with evidence, its properties, its uses and its discovery.

In my remarks so far on a science of evidence, I have used thoughts and examples drawn from a wide variety of disciplines. Many of these examples come from areas of research that I have never previously examined and cited as being important in my own previous work. I acknowledged at the outset that I completely agree with William Twining's view that a science of evidence must be multidisciplinary in nature. I begin with Twining's thoughts on this matter and then I will provide what I regard as a model we might follow in our efforts to develop a truly integrated science of evidence.

5.1. The Science of Evidence: A Multidisciplinary Venture

William Twining emphasizes that the study of evidence can involve anyone with an interest in the subject. In a recent work he specifically identifies the study of evidence as being a multidisciplinary subject[175]. Notice that I have not said that Twining refers to this subject as a "science of evidence". In our conversations, Twining has not objected to the term *science* in connection with the study of evidence, provided that this word is used in a weak sense. I hope that my arguments in this present paper will persuade William that we are entitled to view a science of evidence in a stronger sense of the word *science*.

I have said on many occasions that the field of law has provided us with the richest legacy of experience and scholarship on evidence of any field known to me. Twining's own contributions form a most important part of this legacy. However, he has been among the first to acknowledge that a rich legacy of experience and scholarship on evidence comes from other disciplines. In the minds of many persons the word *evidence* immediately brings the field of law to mind, as if the only persons interested in evidence are those who appear in our courts. This point has also been made by Peter Murphy, another evidence scholar in law. He says[176]

The word 'evidence' is associated more often with lawyers and judicial trials than with any other cross-section of society or form of activity. So it is ironic that the basic questions of what evidence is and what its philosophical and scientific properties are, as opposed to questions of what the law of

evidence may be, have received relatively little attention in legal scholarship, and are rarely considered by judges and practitioners.

Murphy's and Twining's point here is that persons in the field of law have as much to learn about evidence from other areas as they can contribute themselves. Twining mentions Jeremy Betham's often-cited statement: "The field of evidence is no other than the field of knowledge"[177]. This encompasses all of us. On another occasion I said that the house of evidence has many mansions and that we should visit as many of these mansions as we can[178].

Twining mentions that the topic of evidence is now a high profile subject and he gives a variety of examples including our work on the Leverhulme project. He further argues that the core of the subject of evidence is inferential reasoning. He states that all disciplines having empirical elements share common problems concerning evidence and inference. He cites, as an example, common evidential and inferential problems encountered in history and law as they are revealed in a recent work[179]. But in a mild criticism of Twining's views about the commonality of evidential and inferential problems across disciplines, Professor Eileen Scallen argues that we would perhaps have more to learn by focusing on differences among disciplines as far as evidential questions they raise[180]. She also argues against Twining's claim that inferential reasoning forms the core of the subject of evidence. She prefers to put evidence in the realm of rhetoric and its importance in advocacy.

So we should all share our thoughts and our experiences with evidence, its properties and its uses as they appear in our own contexts. What prevents us from doing so? There are certainly incentives issues to consider. Why should a sociologist spend time and effort discussing her/his evidential and inferential problems with a historian or a chemist? Some cross-disciplinary interactions may be more interesting and profitable than others. For example, the field of history and sociology are linked in the study of social history[181]. Another reason, I believe, concerns the kinds of evidence encountered in certain disciplines.

I have often noticed that very few persons from the physical sciences have taken an interest in discussions of evidence. William Twining has made this same observation. I think the basic reason is that, for the most part, the only evidence considered in the physical sciences is the tangible evidence I described in Figure 1. Such evidence is usually observed under replicable conditions in which statistical information in the form of relative frequencies of observations can be tabulated and analyzed.

Some years ago, I would also have included engineers in this group of persons basically uninterested in other forms of evidence in the research they perform. But I have now had twenty years of experience teaching about evidence and inference to a large number of students from a very wide array of fields of engineering. They have readily taken an interest in all of the forms and combinations of evidence that I mentioned in Section 4.1. The basic reason is that they are not only interested in the design of systems and the physical environments in which they will operate, but they are also interested in the people who will use these systems or be affected in some

way by their use. In fact, the first edition of my work on evidence and probabilistic reasoning was published as part of a series in systems engineering[182].

By what means can we make a science of evidence truly multidisciplinary, or integrative that will include persons who do not always interact and share their experiences with evidence and inference? I think we have a very good model to follow, as I now discuss.

5.2 The Science of Complexity: A Model

Nearly three years ago, when I gave my first talk at UCL, I attempted to find some common ground in the work being proposed for the Leverhulme research by the eleven persons at UCL whose proposals I had seen. The obvious common ground I noted in all of these proposals was the complexity of the processes or situations of interest to these eleven persons, who I am now so pleased to call colleagues and friends. I spent most of my talk discussing how each of these proposals reflects at least one attribute of processes we can say are *complex*. Where did I find these attributes? The answer is: several years ago, I began to read avidly the increasing number of studies performed by scholars in Santa Fe, New Mexico at the *Institute for Studies in the Science of Complexity*. The Santa Fe Institute has a very interesting history and has been quite remarkable in its ability to bring together scholars, from a truly wide assortment of disciplines, who in the past may never have thought about collaborating. Persons actively collaborating at the Santa Fe Institute include those having interests in nearly every area of the physical, behavioral and social sciences imaginable. All appreciate the complexity of the processes they are studying. Several works giving an account of the history of the Santa Fe Institute are available[183].

The best way to sample the flavour of an integrated science of complexity, as it has been studied at the Santa Fe Institute, is to examine the multiple volumes now available in the following series: *Proceedings Volumes*, *Lectures Volumes*, *Lecture Notes Volumes*, and *Reference Volumes*. To my knowledge, all are published by Addison-Wesley, New York, NY. I currently have three of the volumes in the *Proceedings Series* and one volume in the *Lectures Series*. I made extensive use of these works in a paper I wrote on probabilistic reasoning and the science of complexity[184]. There are certainly many attributes of complexity in the evidence-based reasoning tasks we all encounter. But my current interest in complexity as it has been studied at the Santa Fe Institute goes deeper than its obvious relation to a science of evidence.

I have wondered recently what will happen to the efforts of all of us currently involved in various evidence-related research when the Leverhulme and ESRC money runs out. Will we all go our separate ways and never collaborate or interact further in our studies? I hope not. I now entertain a proposal that I ask Phil Dawid to take seriously. Why not have a *Center for the Science of Evidence* established at UCL? I can't think of a better place for such a center, and I can't think of a person better able to manage such a center than Phil Dawid. As William Twining notes, the topic

of evidence is a currently vibrant one, both in the UK and here in the USA. Given our now expanding legacy of research on evidence and inference in a variety of contexts, we should be able to attract additional funds for such a center, that I believe would be an absolutely unique venture. We will have the experience of the Santa Fe Institute and its many accomplishments to draw upon.

5.3 A Science of Evidence: Who Should Care?

If you are convinced at all by my arguments regarding the existence of a science of evidence, who should care about its existence? And, who would benefit from it? I mentioned at the outset that the remaining burden I have is to be able to answer these questions. I began this final section of my thoughts about a science of evidence with some thoughts from the historian Marc Bloch. He gives us some very good reasons why we should care about an emerging science of evidence and its applications in many contexts. In order to better understand our own evidential and inferential problems, Bloch says it is indispensable that we consider the manner in which others, in different contexts, have thought about these problems. Here is in example I have drawn from the *Santa Fe Institute Proceedings Volume XVI*. It is entitled*: Understanding Complexity in the Prehistoric Southwest.* A lingering problem in such study is: what caused the *sudden* disappearance of Native American cultures, such as the Anasazi, from large and well organized communities they had occupied for centuries? And, what happened to them? Who would have thought that possible answers to such questions would come from the interaction of archaeologists, physicists, economists, and the many others involved in the science of complexity? One of the two co-authors of this volume is a Nobel Laureate in physics named Murray Gell-Mann, one of the founders of the Santa Fe Institute.

Here is a short answer to my question: Who would benefit from there being a science of evidence? We would all benefit from an established science of evidence, provided that we all share our thoughts and experiences with evidence-based reasoning. Earlier I mentioned that one of the most important evidential subtleties is the synergism that can often occur when we consider two or more items of evidence jointly rather than separately or independently. I believe that a science of evidence would promote such synergism on a larger scale. When the ideas of two or more persons having different thoughts and experiences with evidence are combined, the joint ideas resulting from their interaction may well be more defensible and persuasive than they would be in the absence of such interaction. In short, I have no idea how many persons or organizations would benefit from continued work in a science of evidence. But my strong expectation is that the number of persons and organizations would draw upon research on a science of evidence would be very large and would represent interests that would surprise all of us.

6.0 IN CONCLUSION

I have taken on the task of trying to defend the idea of a science of evidence seriously for the major reason that we have obtained a substantial amount of money to study this science and how it might be integrated in beneficial ways. It would be more than embarrassing for us to conclude that there is really no such thing as a "science" of evidence, especially embarrassing to those of us who have signed on to a project having the title*: Towards an Integrated Science of Evidence*. If we do not think a science of evidence exists, why should we worry about whether it is integrated or not?

But I have had another reason for taking the task of defending the idea of a science of evidence seriously. In my remarks on my standpoint in Section 1.0, I did not mention that I have had another objective, which I have waited until now to acknowledge. I have been honoured and pleased more than I can say at having been made an Honorary Professor of Evidence Science at UCL. If there is no such thing as a science of evidence, I would have to give up this title. But I think I will hold on to this title a bit longer until someone demonstrates that my arguments in favour of there being a science of evidence are not defensible or persuasive.

NOTES

1 Schum, D. *Evidential Foundations of Probabilistic Reasoning*. John Wiley & Sons, NY 1994; Northwestern University Press, Evanston, IL, 2001 [paperback ed], 12 - 21.

2 Hayes, D., Undisciplined Science. *American Scientist*. Vol. 92,No.4, July-August, 2004, 306-310.

3 Anderson, T., Schum, D., Twining, W. *Analysis of Evidence*, 2nd ed. Cambridge University Press, 2005.

4 Twining, W. Evidence as a Multi-Disciplinary Subject. *Law, Probability and Risk*. Vol. 2, No. 2, June, 2003, 91 - 105

5 See note 1 above.

6 See note 1 above at pages 6-8.

7 See http://tinyurl.com/d6f7t

8 Zangwill, I. *The Big Bow Mystery [1891]*. Carroll & Graf Publishers, New York, 1986 edition, 143 - 144.

9 Wigmore, J. H. *The Science of Judicial Proof: As Given by Logic, Psychology, and General Experience, and Illustrated in Judicial Trials*. 3rd ed. Little, Brown and Company, Boston, MA,.

1937

10 id. At page 3.

11 Twining, W. *Theories of Evidence: Bentham & Wigmore*. Stanford University Press, Stanford CA., 1985, 116.

12 id

13 See note 9 at page 8

14 See note 9 at page 119.

15 Mayr, E. *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*. Belknap Press of Harvard University. Cambridge, MA, 1982, 829 - 858.

16 Allen, J. *Inference from Signs: Ancient Debates about the Nature of Evidence*. Clarendon Press, Oxford, 2001.

17 Colapietro, V. *Glossary of Semiotics*. Paragon House, New York, 1993, 179.

18 See note 16 at page 1.

19 Schama, S. Dead Certainties: (Unwarranted Speculations). Vintage Books, New York, 1992, 320

20 Hacking, I. The Emergence of Probabil*ty:* A Philosophical Study of Early Ideas about Probability, Induction, and Statisical Inference. Cambridge University Press, 1978, 31 - 48.

21 Kneale, W., Kneale, M. The Development of Logic. Clarendon Press, Oxford, 1984, 241

22 e,g, see Holdsworth, W. *A History of English Law*, vol.1, Methuen, London, 1903; Plucknett, T. *A Concise History of the Common Law of England*. 5th ed. Little, Brown & Co,. Boston, MA, 1956; Pollock, F., and Maitland, F. *The History of English Law before the Time of Edward I*. Cambridge University Press, 1968; Cockburn, J., Green, T. *Twelve Good Men and True: The Criminal Jury Trial in England, 1200 - 1800.* Princeton University Press, Princeton, NJ, 1988

23 Wells, C. The origin of the petty jury. *Law Quarterly Review*. Vol. 107, 1911, 347 - 361; Wells, C. Early Opposition to the Petty Jury in Criminal Cases. *Law Quarterly Review*. Vol. 117, 1914, 97 - 110

24 id. See Wells, 1911 at page 360.

25 See note 22 Plucknett 1956 at page 436.

26 Hale, M. *The History of the Common Law of England* [1739]. C. M. Gray ed. University of Chicago Press, Chicago, IL 1971, 164.

27 Toulmin, S. *The Uses of Argument*. Cambridge University Press, 1964 ed. 7.

28 Twining, W. The Rationalist Tradition of Evidence Scholarship. In: W. Twining. *Rethinking Evidence: Exploratory Essays*. Basil Blackwell, Oxford, 1990, 32 - 91.

29 Id. at page 73.

30 Twining, W. *Blackstone's Tower: The English Law School*. Stevens & Sons/Sweet & Maxwell, London, 1884, 182.

31 David, F. *Gods, Games, and Gambling*. Griffin, London, 1962.

32 See note 20 at pages 1- 2.

33 Todhunter, I. *A History of Mathematical Probability: From the Time of Pascal to that of Laplace*. 1865. Chelsea Publishing, NY, 1965 reprint.

34 Keynes, J. *A Treatise on Probability*. 1921. Macmillan, London, 1957 reprint, 180 - 185.

35 Daston, L. *Classical Probability in the Enlightenment*. Princeton University Press, Princeton, NJ, 1988, 370 - 386.

36 Zabell, S. The Probabilistic Analysis of Testimony. *Journal of Statistical Planning and Inference*. Vol. 20, 1988, 327 - 354.

37 Schum, D. Knowledge, Probability, and Credibility. *Journal of Behavioral Decision Making*. Vol. 2, 1989, 39 - 62.

38 Locke, J. *An Essay Concerning Human Understanding* [1689]. Ed P. Nidditch, Clarendon Press, Oxford, 1991, 655 - 664

39 See note 9 at pages 13 - 15.

40 Toulmin, S. *The Uses of Argument*. Cambridge University Press, Cambridge, 1958.

41 Wiigmore, J. The Problem of Proof. 8 *Illinois Law Review*, No 2, 1913, 77 - 103.

42 Id.

43 Toulmin, S., Rieke, R., Janik, A. *An Introduction to Reasoning*, 2nd ed, Macmillan, NY, 1984, 73 - 77.

44 See note 9 at page 8.

45 Achinstein, P. [ed] *The Concept of Evidence*, Oxford University Press 1983.

46 Achinstein, P. *The Book of Evidence*. Oxford University Press, 2001.

47 Eco, U., *A Theory of Semiotics*. Indiana University Press, Bloomington, IN, 1979, 7

48 Eco, U., Sebeok, T. eds. *The Sign of Three: Dupin, Holmes, Peirce*. Indiana University Press, Bloomington, IN, 1983.

49 *The New Shorter Oxford English Dictionary*. Vol. 2, Clarendon Press, Oxford, 1993, 2717

50 Ochoa, G., Corey, M. *The Timeline Book of Science*. Stonesong Press, Ballantine Books, New

York, NY, 1995.

51 Hempel, C. On the Nature of Mathematical Truth. In: J. R. Newman, ed. *The World of Mathematics*. Vol. III, Simon and Schuster, New York, NY, 1956

52 Oldroyd, D. *The Arch of Knowledge: An Introductory Study of the History of the Philosophy and Methodology of Science*. Methuen, New York, NY, 1986

53 Schum, D. Teaching about Discovery and Invention in Engineering. *Technological Forecasting and Social Change*. Vol. 64, 2000, 209 - 223

54 Baring-Gould, W.S. The Annotated Sherlock Holmes, Vol. I, Clarkson N. Potter, New York, NY, 1967, 231

55 Popper, K. *The Logic of Scientific Discovery*. Harper Torchbooks, New York, NY, 1968, 31

56 Hintikka, J., Hintikka,M. Sherlock Holmes Confronts Modern Logic: Toward a Theory of Information Seeking Through Questioning. In: Eco, U., Sebeok T., [eds], The Sign of Three: Dupin, Holmes, Peirce. Indiana University Press, Bloomington, IN, 1988, 154 - 169.

57 Achinstein, P. [ed] *Science Rules: A Historical Introduction to Scientific Methods*. The Johns Hopkins University Press, Baltimore, MD, 2004

58 Gjertsen, D. *Science and Philosophy: Past and Present*, Penguin Books, London, 1989, 113

59 http://www.scientificmethod.com/smvsm.html. This address is just an index to the links provided for each of the eleven steps that are given.

60 Bauer, H. *Scientific Literacy and the Myth of the Scientific Method*. University of Illinois Press, Urbana, IL., 1994.

61 Ibid, at page 32

62 Haack, S. *Defending Science - Within Reason: Between Scientism and Cynicism.* Prometheus Books, Amherst, NY., 2003.

63 Ibid at page 23

64 Twining, W., Hampsher-Monk, I. *[eds] Evidence and Inference in History and Law: Interdisciplinary Dialogues.* Northwestern University Press, Evanston IL, 2003

65 Katritzky, M. The Mountebank: A Case Study in Early Modern Theater Iconography. In: Twining, W., Hampsher-Monk, I. *[eds] Evidence and Inference in History and Law: Interdisciplinary Dialogues.* Northwestern University Press, Evanston IL, 2003 231 - 286.

66 Bunge, W. Causality in Modern Science. Dover Publications, New York, NY., 3rd ed 1979.

67 Bunge, M. In Praise of Intolerance to Charlatanism in Academia, In: Gross, P. Levitt, N., Lewis, M. [eds] *The Flight from Science and Reason.* The New York Academy of Sciences, New York, NY, 1996, 96 - 115.

68 Ibid at page 103

69 Ibid at page 103

70 Lempert, R. The New Evidence Scholarship: Analyzing the Process of Proof. 66 Boston University Law Review, Nos. 3 - 4, 1986, 439 - 477

71 See note 67 at page 103 - 104

72 Cohen, L, J. *The Probable and the Provable.* Clarendon Press, Oxford, 1977.

73 Shafer, G. *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, NJ., 1976

74 Poincaré, J. H. *The Value of Science* [1905]. In: Gould, S. J. [ed] *Henri Poincaré, The Value of Science: Essential Writings of Henri Poincaré.* The Modern Library, NY, 2001, 347

75 Carnap, R. *An Introduction to The Philosophy of Science.* [ed] Gardner, M. Basic Books, NY, 1966, 51

76 note 74 at pages 364-365

77 e.g. see note 1 at pages 114 - 120.

78 Schum, D. Hearsay from a Layperson. 14 *Cardozo Law Review*, Vol 1, 1992, 1 - 77

79 Bloch, M. *The Historian's Craft.* Vintage Books, NY., 1953, 51

80 see note 38.

81 Ekelof, P. O. My Thoughts on Evidentiary Value. In*: Evidentiary Value: Philosophical, Judicial, and Psychological Aspects of a Theory.* [Eds] Gardenfors, P., Hansson, B., Sahlin, N. Gleerups, Lund, Sweden, 1983, 11.

82 See note 3 at 76 -77

83 See note 1 at pages 366 - 449

84 Kadane, J., Schum, D. *A Probabilistic Analysis of the Sacco and Vanzetti Evidence.* J. Wiley & Sons, New York, NY, 1996.

85 See note 1 at pages 114 - 130; note 3 at pages 71 - 77.

86 See note 7.

87 See note 65.

88 See note 84.

89 E.g. Edelson, M. *Hypothesis and Evidence in Psychoanalysis.* University of Chicago Press, Chicago, IL. 1984.

90 See note 75.

91 Bothamley, J. *Dictionary of Theories: One Stop to more than 5,000 Theories.* Visible Ink, Detroit, MI, 2002.

92 As quoted in: Copi, I. M. *Introduction to Logic*, 6th ed., Macmillan, New York, NY. 1982, 47.

93 See note 9 at pages 9 -22.

94 See note 40 at pages 97 - 107.

95 E.g. see note 3 at pages 262 - 288.

96 Wigmore, J. H. The Problem of Proof. 8 *Illinois Law Review*, No. 2, 1913, 77 - 103. Also see note 9 at pages 858 - 919.

97 See note 84.

98 Geller, M. The Last Wedge. In: Twining, W. Hampsher-Monk. I. *Evidence and Inference in History and Law: Interdisciplinary Dialogues*. Northwestern University Pres, Evanston, IL., 2003, 122 - 139.

99 Anderson, T. Wigmore Meets the "Last Wedge". In: Twining, W. Hampsher-Monk. I. *Evidence and Inference in History and Law: Interdisciplinary Dialogues*. Northwestern University Pres, Evanston, IL., 2003, 140 - 215.

100 Schum, D. Alternative View of Argument Construction from a Mass of Evidence. 22 Cardozo Law Review, Nos. 5 -6, July 2001, 1461 - 1592.

101 See note 1 at pages 140 - 156.

102 *Federal Rules of Evidence: 2005 - 2006 Edition*. Thompson-West. 2005, 28 - 30.

103 See note 74 and accompanying discussion.

104 See note 70.

105 See note 103 at page 12.

106 Sperber, D., Wilson, D. *Relevance: Communication and Cognition*. Harvard University Press, Cambridge, MA., 1986

107 See note 75.

108 Note 107 at pages 79 - 83.

109 Ibid at pages 6 - 21.

110 See note 107 at pages 123 - 132.

111 Ibid at page 2.

112 Ibid at pages 75 - 83.

113 Ibid at page 125.

114 See note 1 at pages 86 - 92.

115 Lichtman, A. French, V. *Historians and the Living Past; The Theory and Practice of Historical Study*. Harlan Davidson, Inc. Arlington Heights, IL.,1978, 200 - 201.

116 *Black's Law Dictionary*, 4th ed, West Publishing Co., St. Paul, MN., 1951, 355.

117 See note 103 at pages 73 - 75.

118 e.g. Swets, J. *Signal Detection and Recognition by Human Observers*. John Wiley & Sons, New York, NY, 1964.

119 Schum, D. Du Charme, W. Comments on the Relationship Between the Impact and the Reliability of Evidence. *Organizational Behavior and Human Performance*. Vol 6., No.2, 1971, 111 - 131; Schum, D. Sorting Out the Effects of Witness Sensitivity and Response Criterion Placement on the Inferential Value of Testimonial Evidence, . *Organizational Behavior and Human Performance*. Vol 27. 1981, 153 - 196.

120 Cohen, L. *The Dialogue of Reason*. Clarendon Press, Oxford,1986, 19.

121 Gettier, E. Is Justified Belief Knowledge? *Analysis*, Vol.23, 1963, 121 -123

122 Radford, C. Knowledge - By Examples. *Analysis*, Vol. 27, 1966, 1 - 11.

123 Schum, D. Knowledge, Probability and Credibility. *Journal of Behavioral Decision Making*. Vol.2, 1989, 39 - 62.

124 Nozick, R. Philosophical Explanations. Belnap Press of Harvard University, Cambridge, MA., 1981, 169.

125 See note 121 at page 23.

126 E.g. see note 1 at page 107; note 3 at page 68; note 78 at page 34.

127 Wigmore, J. H. 2 *Evidence in Trails at Common Law*. Little, Brown & Co. Boston, MA., § 650, at page 755, 3rd ed. 1940.

128 Ibid at page 762 - 771.

129 See note 78.

130 See note 1 at pages 447 - 449.

131 See note 78 at pages 28 -32.

132 See note 38.

133 Hume, D. *A Treatise of Human Nature* (1739), [ed[Selby-Bigge, Oxford University Press, 1975, 103

134 See note 1 at pages 200 - 269.

135 See note 3 at pages 246 - 261.

136 Dale, A., *Most Honorable Rembrance: The Life and Work of Thomas Bayes*. Springer-Verlag, New York, NY, 2003.

137 See note 70.

138 See note 1, Chapters 6 - 8.

139 Kolmogorov, A. N. Foundations of a Theory of Probability (1933), 2nd English edition, Chelsea Publishing, New York, NY., 1956.

140E.g., Shafer, G. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ., 1976.

141 E. g. See note 3 at pages 159 - 223.

142 Weis, R. *Criminal Justice: The True Story of Edith Thompson*. Hamish Hamilton, London,1988; Penguin, London, 2001.

143 See note 64 Chapter 2, pages 63 - 121.

144 Cohen, L. J. *The Probable and the Provable*. Clarendon Press, Oxford, 1977. Also see: Cohen, L. J., *An Introduction to the Philosophy of Induction and Probability*. Clarendon Press, Oxford, 1989

145 Schum, D. *Evidence and Inference for the Intelligence Analyst*, [2 Volumes], University Press of America, Lanham, MD. 1987. Multiple references to Cohen's work in both volumes.

146 Keynes, J. M. *A Treatise on Probability* (1921). Reprint of 1st ed. Macmillan, London, 1957.

147 Cohen, L. J. Twelve Questions about Keynes's Concept of Weight. *Brit. J. Phil. Sci*., Vol. 37, 1985, 263 - 278.

148 Schum, D. Jonathan Cohen and Thomas Bayes on the Analysis of Chains of Reasoning. *Probability and Rationality: Studies on L. Jonathan Cohen's Philosophy of Science*. Posnan Studies in the Philosophy of Sciences and the Humanities. Vol.21. Eds. Eellls,E., Maruszewski, T. Rodopi Press, Amsterdam, 1991, 99 - 145

149 See note 9 at pages 864 - 866.

150 E.g., Zadeh, L. Fuzzy Sets. *Information and Control*. Vol. 8, 1968, 338 - 353; Zadeh, L. The Role of Fuzzy Logic in the Management of Uncertainty in Expert Systems. *Fuzzy Sets and Systems*, Vol. 11, 199 - 227.

151 E. g. Haack, S. *Deviant Logic, Fuzzy Logic: Beyond the Formalism*.University of Chicago Press, Chicago, IL., 1996.

152 See note 4 at page 97.

153 See note 9 at pages 18 - 48.

154 See note 27 at page 7.

155 Schum.D. Capturing an Interesting Subtlety Involving a Source of Testimonial Evidence. *UCL Studies of Evidence Science Report*, May 4, 2004.

156 See note 1 at pages 401 - 409.

157 See note 84 at pages 175 - 183; 198 - 239.

158 Ibid pages 198 - 215.

159 Ibid pages 215 - 239.

160 E.g. see Pearl, J. *Probabilisic Reasoning in Expert Systems: Networks of Plausible Inference*.Morgan Kauffman, San Mateo, CA., 1988; Lauritzen, S., Spiegelhalter, D. Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems. Journal of the Royal Statistical Society: Series B, Vol 50., No.2, 1988, 157 - 224.

161 Paulos, J. A. *Once upon a Number: The Hidden Mathematical Logic of Stories*. Basic Books, New York, NNY, 1998, 71 - 73

162 See note 1, Chapters 7 and 8.

163 Lakatos, I. *The Methodology of Scientific Research Programs*. Ed. Worrall, J., Currie,G. Cambridge University Press, 1980; Lakatos, I. *Mathematics, Science and Epistemology*. Ed. Worrall, J., Currie,G. Cambridge University Press, 1980.

164 Coady, C. A. J. *Testimony: A Philosophical Study*. Clarendon Press, Oxford, 1992.

165 See note 35.

166 E. g. Earman, J. *Hume's Abject Failure: The Argument Against Miracles*.Oxford University Press, , Oxford, 2000.

167 Edwards, W., Schum, D., Winkler, R. Murder and (of) the Likelihood Principle: A Trialogue. *Journal of Behavioral Decision Making*. Vol. 3, 1990, 75 - 89.

168 Schum, D. Species of Abductive Reasoning in Fact Investigation in Law. *Cardozo Law Review*, Vol. 22, Nos. 5 - 6, July 2001,1645 - 1681.

169 E.g. Langley, P., Simon, H., Bradshaw, G., Zytkow, J. *Scientific Discovery: Computational Explorations of the Creative Process*. MIT Press, Cambridge, MA., 1987.

170 E. g. Tillers, P., Schum, D. A Theory of Preliminary Fact Investgation. 24 *University of California-Davis Law Review*, No.4., 1991, 931 - 1012; Schum, D. Marshaling Thoughts and Evidence During Fact Investigation. 40 *South Texas Law Review*, No. 2, 1999, 401 - 454.

171 Ibid, see Schum, 1999 at page 437.

172 See note 156.

173 See note 15 at page 829.

174 See note 79 at page 18 - 19.

175 See note 4.

176 Murphy, P. *Evidence, Proof, and Facts: A Book of Sources*. Oxford University Press, Oxford, 2003,

177 See note 4 at page 92.

178 See note 1 at page xiv.

179 See note 64.

180 Scallen, E. 'Mere' Rhetoric about Common Ground and Different Perspectives: A Comment on Twining's 'Evidence as a Multi-disciplinary Subject'. *Law, Probability and Risk*. Vol. 2, 2003, 109 - 116.

181 E.g. Davidson, J., Lytle, M. *After the Fact: The Art of Historical Detection*. McGraw-Hill, New York, NY.1986,

182 See note 1, 1994 edition.

183 E.g. Waldrop, M. *Complexity: The Emerging Science at the Edge of Order and Chaos*.Touchstone, New York, NY., 1993; Lewin, R. *Complexity: Life at the Edge of Chaos*. Collier Books, New York, NY. 1992.

184 Schum. D. Probabilistic Reasoning and the Science of Complexity. In: Shanteau, J., Mellers, B., Schum, D. [eds] *Decision Science and Technology: Reflections on the Contributions of Ward Edwards*. Kluwer Academic Publishers, Boston MA., 1999, 183 - 209.