

**ASSESSING THE COMPETENCE AND CREDIBILITY OF
HUMAN SOURCES OF INTELLIGENCE EVIDENCE:
CONTRIBUTIONS FROM LAW AND PROBABILITY***

**David A. Schum, Professor
Volgenau School of Information Technology and Engineering
School of Law
George Mason University
Honorary Professor of Evidence Science
University College London**

**Jon R. Morris, Affiliate Professor
Volgenau School of Information Technology and Engineering
George Mason University, and
United States Central Intelligence Agency
Directorate of Science and Technology**

6 July, 2007 [Revised]

*** This work was supported in part by a research grant to University College London provided by the Leverhulme Foundation and the Economics and Social Research Council. The authors are most grateful for this support.**

ABSTRACT

These are perilous times in which our security is under continual threat by persons and organizations around the world and at home. Information supplied to us by human sources concerning the capabilities and intentions of these persons and organizations is crucial to our ability to recognize and prevent these threatening actions. We have all seen the many news accounts of our need for more and better HUMINT (Human Intelligence). But a major issue is: When we obtain an item of HUMINT, to what extent can we believe it? Our ability to make these assessments involves determining the competence and credibility of the human source providing this item. These assessments are very difficult and the persons making them need all the help they can get. Fortunately, much assistance comes from the fields of law and probability.

This paper provides an account of how we are now exploiting the very rich legacy of experience and scholarship accumulated in the field of law over the past 500 years or so regarding questions to ask about the competence and credibility of witnesses. For quite some time we have been attempting to bring valuable insights from the field of law concerning various evidential and inferential matters to the attention of persons in the Intelligence Community. On this occasion we will describe a computer-assisted system called MACE (Method for Assessing the Credibility of Evidence) we are developing to exploit what we have learned from law regarding assessments of the competence and credibility of sources of HUMINT. This system is designed to assist in addressing the question: To what extent can we believe this particular item of HUMINT that has just been supplied to us by a human source? MACE employs two different, but entirely complementary, probability systems to help us answer this question. Baconian probability methods help us answer the question: How much evidence do we have about this human source and how completely does it answer questions about the source's competence and credibility? Bayesian probability methods allow us to determine how strong is the evidence we do have about this particular human source, and it also provides an assessment of the posterior odds favoring the extent to which we can believe or disbelieve what this source is telling us.

**ASSESSING THE COMPETENCE AND CREDIBILITY OF HUMAN SOURCES
OF INTELLIGENCE EVIDENCE: CONTRIBUTIONS FROM LAW AND PROBABILITY**

David A. Schum

Jon R. Morris

1.0 PREFACE

Every day of our lives we rely upon information we receive from others based on events they claim to have observed. In some situations such reliance forms the basis for very important conclusions we must reach. One example concerns the testimony of witnesses who appear at a trial in which we are asked to serve as jurors. What these witnesses have to tell us will often involve serious consequences to the defendants, plaintiffs and other stakeholders in the matters in dispute. Another example involves the persons charged with evaluating and analyzing the testimony of persons providing intelligence information that we believe concerns our nations' security. We all know about the terrorist actions in New York and Washington DC in September, 2001 and in London in July, 2005. As we write this we have just heard about the terrorist attempts in London and Glasgow in July, 2007. In such instances the stakes involved may be enormous and involve the lives, freedom and fortunes of untold numbers of people. In our everyday lives we also rely on the testimony of others in order to draw conclusions about many events that are important to us. An obvious example concerns the accounts given in the news media by reporters or other persons who claim to have observed events and situations we care about.

The trouble is that reliance on the testimony of persons in all of these situations may not be justified. In some cases a person will report observing events when we later discover that no such observation was ever made; this would be evidence that the person is trying to mislead us. But even in situations in which we have confidence that a person did make a relevant

observation, we may have serious reasons for not believing what this person has told us. From common experience we know that people do not always tell us what they believe actually happened; in such cases their veracity or truthfulness is at issue. Even if a person tells us something this person does believe, we have questions about how this person formed this belief. Was it based on the sensory evidence this person obtained during an observation, or did this person believe the event occurred simply because this person expected or desired it to occur? This involves the person's objectivity. Finally, even if this person did form this belief objectively on the basis of sensory evidence, we must ask how good this sensory evidence was in the conditions under which the person made the relevant observation. In this case we must inquire about the accuracy or sensitivity of this person's senses.

Who knows when concern about the credibility of testimony first arose in human affairs? What we do know is that there are references to credibility matters in some very old documents. One example refers to the necessity for corroborating witness testimony: "...for any inquiry...at the mouth of two witnesses or at the mouth of three witnesses shall the matter be decided" [*Deuteronomy*; 19:15]. Our present concern about the history of credibility issues in testimonial evidence begins with the development of the jury system and the adversarial nature of trials in England that took place over several centuries. As this system developed witnesses began to appear for both parties in a dispute and concern began to emerge about how the credibility of these witnesses could be impeached or supported. But it was not until the 1600s that the discipline of probability enters our history. At the very time at which the first mathematical thinking about probability emerged, there was great interest expressed by persons in several disciplines including law, history, business and even theology about what was then called the "credibility-testimony" problem. Persons in many disciplines were eager to make probability calculations they believed appropriate to various situations in which we have testimonial evidence about various events of interest.

What follows is an account of work we have done in developing a computer-assisted system for assessing the competence and credibility of human sources of intelligence information. This system rests on some old and some new ideas. First, it would be very foolish to ignore the very rich legacy of experience and scholarship in the field of law that has accumulated over at least the past five hundred years concerning assessment of the competence and credibility of trial witnesses. This legacy provides us with experience-tested questions to ask about both of these characteristics of sources of testimonial evidence. In the process of classifying these questions we obtain additional verification that we have identified the essential attributes of the competence and credibility of persons providing testimonial evidence. Although it is true that there is a past legacy of interest in credibility-testimony issues in the field of probability, we have found it necessary to adopt newer approaches to assessing the credibility of sources of testimonial evidence and its bearing on the inferential or probative force of testimonial evidence. The approach taken by early (and some later) probabilists to credibility-testimony issues is not helpful for several reasons that we will mention. Here again in our work we have combined the old with the new. We employ terms necessary in using Bayes' rule, which has been around for quite some time. But we have also found it necessary to employ ideas from much newer ideas in what has been termed Baconian probability.

Readers interested in the consequences of failing to carefully examine human sources of intelligence information [called HUMINT] can do no better than to read the account provided in open-source literature by Drumheller and Monaghan regarding our experience with a person code-named "Curveball" [2006, 231 - 282]. Curveball, an Iraqi national, fled from Iraq to Germany where he supplied many reports regarding Saddam Hussein's alleged development of weapons of mass destruction [WMD]. On February 5, 2003 our then Secretary of State, Colin Powell, gave a speech concerning Saddam's having available WMDs to the United Nations Security Council. While this speech was being prepared, reliance was placed on Curveball's testimony. But many intelligence analysts were convinced that Curveball had fabricated his accounts of Saddam's having stockpiles of WMDs. In spite of this, a short time later in 2003 we invaded Iraq and it was

discovered, much to the embarrassment of Colin Powell and many others, that Curveball was indeed a fabricator. This notorious incident illustrates the hazards of relying on HUMINT without careful assessment of the competence and credibility of its sources. Many questions were not asked about Curveball that should have been asked. In fact, Curveball should have been subjected to a detailed competence and credibility analysis of the sort that witnesses so often experience in our trials at law.

2.0 A BIT OF HISTORY

The history of thoughts about testimonial evidence is a fascinating subject all on its own and we could easily spend many pages describing the interesting characters we meet along the way in this history. Our interest in testimonial evidence concerns the competence and credibility of its sources and how these characteristics affect the probative or inferential force or weight of such evidence. Though we provide a brief account of historical developments in law and probability that bear upon our present work, we pause to mention the thoughts of a person in neither field, but one who thought about testimonial evidence: Samuel Johnson. In his account of the life of Samuel Johnson, L. L. D, Boswell records a conversation involving Johnson and several other persons on May 30, 1784. The conversation concerned arguments and testimony in the House of Commons. Boswell records the following comment by Johnson [Boswell, 1791, Croker edition, 1890, 757]:

There is a beautiful image in Bacon upon this subject: testimony is like an arrow shot from a long bow; the force of it depends upon the strength of the hand that draws it. Argument is like an arrow shot from a cross bow, which has equal force though shot by a child.

Two comments are necessary regarding Johnson's image. First, it seems that his memory failed him on this occasion since the image he quotes was due to Robert Boyle and not

Francis Bacon. Second, Johnson was almost certainly referring to the credibility of the source of testimony when he speaks about the force of testimony being dependent on the strength of the hand that draws the long bow. The trouble with this image is that the credibility of evidence forms the very foundation of all arguments based on any form of evidence. So, if this evidence is testimonial in nature, we must have the arm strength of a person drawing a long bow in order for a child to shoot the cross bow of a defensible argument.

2.1 I History of Testimony in Anglo-American Law

When the Normans came in 1066 they brought with them an inquisitorial system for dispensing justice. Under this system a person accused of a crime had the burden of proving his innocence. The proof process involved a series of three steps: judgment - trial - sentence. At the judgment stage a defendant had a choice of three ways of proving his/her innocence, none of which involved evidence of any kind. The accused could chose trial by combat against an accuser; submit to an ordeal such as carrying a hot iron over a certain distance or being bound and then submersed in water; or obtain a certain number of persons willing to swear on oath to support the accused innocence. The trial then proceeded in accordance with the accused choice of these three options. No evidence was presented and proof was left judicium dei, to the judgment of God. It was believed that God would not side with a guilty defendant in a trial by combat; God would not allow the burnt hand of a guilty defendant to show signs of healing in a short period of time [usually three days]; and God would strike dead any person who gave a false oath about the innocence of any guilty defendant. The verdict and sentence followed in accordance with God's apparent wishes during a trial. However, at the Fourth Lateran Council in 1215, Pope Innocent III ruled that clergy members could no longer bless these ordeals. But another reason was given for the decline in these evidence-free methods of proof; the crown did not always get the verdicts they expected or desired (Holdsworth, 1903, Vol.1, 143).

But another event happened in 1215 when, on Monday, June 15, King John signed the document we call Magna Carta. Clause 39 of this document says that the King agrees that he will not proceed with force against any man, or send others to do so, except by the lawful judgment of his peers (judicium parium). Some persons have interpreted this clause to mean trial by jury, but this seems not to have been its original intent. Some historians have said that Magna Carta belonged to the future and became important not for what it actually said but for what people thought it said (Poole, 1955, 476 - 477). There were at the time two forms of juries, which actually dated from Anglo-Saxon times (Stenton, 1962, 502 - 503). The first, called presentment juries, performed an accusatorial function; they presented charges against a person in criminal cases. Juries in civil matters were called assize juries. These presentment juries were the forerunner of our modern grand juries. They were also the forerunner of our petty juries, the ordinary trial juries we have today.

But the first trial juries hardly resembled our trial juries of today. To fill the gap left by the demise of proof by combat, ordeals, and oaths, presentment juries were asked to rule in criminal cases against a defendant. There were no set procedures for these early trial juries and they had a characteristic that made them greatly undesirable to defendants. The first trial juries consisted of accusers, witnesses, and others having a vested interest in the case. It comes as no surprise to learn that most defendants preferred the older proof methods. The history of the development of our jury system certainly makes interesting reading and there are many fine works on the subject (e.g. Holdsworth, 1903; Plucknett, 1958; Pollock and Maitland. 1968; Cockburn and Green, 1988).

What is especially interesting for our present purposes is when jurors ceased being witnesses and were chosen because they had no knowledge of the accused or the offense being tried and would agree to be objective and unbiased in their evaluation of what external witnesses testified. There is some disagreement among legal historians about when jurors ceased being witnesses themselves. Holdsworth places this time as being in the seventeenth century (1903,

Vol. 1, 157), but it might have occurred earlier. For example, there were several developments that enhanced this transformation from jurors being witnesses and having vested interests in a case. In 1352 a statute of Edward III allowed an accused to challenge the suitability of any juror who had joined in his/her indictment (Wells, 1911, 36). And in 1563 during the reign of Elizabeth I appeared a statute that compelled the attendance of witnesses at trial and made perjury a criminal offense (Plucknett, 1956, 436).

As the rights of an accused to representation by an advocate became more extensive, trials became more adversarial in nature. In this emerging adversarial climate it became apparent that one side of a matter in dispute would seek to challenge the evidence and arguments provided by the other. This emerging adversarial process led Sir Matthew Hale to claim that the questioning by parties in contention, by the advocates, judges, and juries is a better process of "beating and boulding the truth" than any other system lacking this adversarial quality (Hale, 1739, Grey Edition, 1971, 164). We can reasonably infer that this emerging adversarial process generated interest in developing questioning strategies designed to impeach or support the competence and credibility of external witnesses who provide testimonial evidence at trial. In Section 4.0 below we list five major questions asked of witnesses regarding their competence and 23 major questions asked about their credibility. Another very matter we will mention is how it became apparent in trials at law that witness competence and credibility are two quite distinct characteristics involving different attributes and different questions. These questions come from our examination of a very large assortment of current evidence treatises in which the impeachment and support of witness competence and credibility are discussed. We have no way of knowing the precise times during the emergence of this adversarial process at which the importance of asking a particular question became important. As we will mention, some of these questions were undoubtedly asked of witnesses in much earlier times and in other cultures.

These brief historical comments simply show how reliance on external witnesses only gradually emerged during the history of our legal system and only when the process became truly

adversarial did concern about witness competence and credibility become such important matters. The field of law has provided persons in other disciplines with such a rich legacy of experience and scholarship on the properties and uses of evidence; the assessment of the credibility of testimonial evidence just being one example. William Twining has told us that in acknowledging these contributions we would do well to emphasize the importance of experience in this legacy. He says that most of this legacy has accumulated in the day to day "crucible of adversarial argument" in trials at law (Twining, 1994, 181 - 182). This is precisely why we have said that the grounds we will mention for impeaching and supporting the competence and credibility of intelligence sources are experience-given in law.

2.2 Probability Theory and Witness Testimonial Credibility

Roots of interest among probabilists in the credibility of testimony do not go back near as far as the roots we have discussed that were planted in law. The major reason is that interest in mathematical calculations of probabilities only go back to the sixteen hundreds and are usually associated with the name Blaise Pascal (1623 - 1662). In fact, the historian of probability Todhunter supplies an exact date of July 29, 1654 at which interest in probability calculation was first generated in a letter written by Pascal to another mathematician named Pierre de Fermat (Todhunter, 1865, 8). This letter concerned how to divide the stakes among players in a game of chance. It is often believed that early interest in probability calculations was confined only to games of chance. But the work of current historians suggests that persons with interests in many other areas also became immediately interested in probability calculations. For example, Hacking tells us that it would be a mistake to infer that probability could only be measured with reference to concepts devised for games of chance (Hacking, 1978, 10). He goes on to mention a variety of applications for probability that were sought at the time in law, theology, mortality studies and annuities. Theologians were interested in probabilities of miraculous events having religious significance and how these probabilities might seem to decline over time. In 1699 John Craig made calculations concerning the time at which the story of Christ would expire; according to his

calculations this would happen in 3150 AD (Stigler, 1986, 883). Daston also provides additional examples of emerging interests in probability among historians, politicians and persons in other areas; she refers to applications in these areas as representing the moralization of mathematics (Daston, 1988, 296 - 369).

Our present interest concerns probabilistic analyses of what were called credibility-testimony problems. Zabell has provided us with a chronological listing of 23 persons who had interest in the probabilistic analysis of testimony (Zabell, 1988, 350). His list begins with George Hooper (1640 - 1727), Bishop of Bath and Wells, and ends with the American philosopher Charles Sanders Peirce (1839 - 1914). There are two credibility testimony problems of interest to us because they concern how probabilities associated with sources of testimony were defined at various times among probabilists. These two problems and their solutions offered by early probabilists are mentioned by Todhunter but he does not say who first attempted their solution (Todhunter, 1865, 55). Zabell says it was Hooper and calls the results *Hooper's rules*; although he allows that they might have originated with John Craig (Zabell, 1988, 336). This first problem involved what was called simultaneous or concurrent testimony in Hooper's time. In this situation we have some number N of witnesses, all of whom testify that event E has occurred. Today we would say the testimony of these N witnesses is corroborative in nature since they all report the same event. The question posed was: how does our belief that E did occur increase as we hear a greater number of witnesses concurrently telling us that E did occur? According to this rule, if we have N witnesses, each of whom have the same credibility p , after the N 'th witness the probabilistic degree of our belief that E occurred should be $P(E) = 1 - (1 - p)^N$. So, after the first witness $P(E) = 1 - (1 - p) = p$; after the second witness $P(E) = 1 - (1 - p)(1 - p)$, and so on.

Hooper's second rule concerns what was then called successive testimony. In this case we have a succession or chain of N sources, each of whom has credibility p of faithfully reporting what was learned from the preceding source. This sounds very much like what we would refer to today as a hearsay chain, or a chain of secondhand evidence. Suppose we hear from the N th

source in the chain that event E occurred. How strongly should we believe that event E occurred as presumably reported by the first source to the second source in the chain? The answer Hooper gave is that we should have $P(E) = p^N$. As you see, according to Hooper's rules, $P(E)$ approaches 1.0 asymptotically for simultaneous testimony and $P(E)$ approaches zero asymptotically for successive testimony.

There are many reasons for criticizing early probabilistic studies such as these. The first concerns how the term credibility was defined in many of these studies. In some cases credibility was defined as the probability that a witness was "speaking the truth". On some occasions it was said that a witness is credible if the event this witness reported actually occurred. This definition confounds the several attributes of credibility. For example, a witness might be truthful about what he observed but inaccurate in his observation. In some cases witness accuracy and veracity are mentioned but the definition of veracity was also troublesome since it was said that a witness was being truthful only if the reported event actually occurred. This also confounds credibility attributes of witness credibility. What it comes to is that the early probabilists did not spend much time decomposing witness credibility in an effort to reveal its several attributes. And in none of their studies do we find any concern about witness competence and its relation to whether a witness actually made the observation claimed in his/her testimony. For reasons of simplicity in doing probability calculations it was usually assumed that collections of witnesses all have the same credibility, however it was defined. This hardly captures the many situations in which we have multiple witnesses each of whom may have any gradation of credibility shy of perfection. An additional trouble is that probabilists believed that witness credibility could be graded using a single probability. In our current studies we observe that it takes at least two probabilities to grade witness credibility.

But we should also not be overcritical of these early probability studies; they did not overlook all factors naturally involved in credibility-testimony problems. One very important matter they did not overlook was the rareness or improbability of the event reported in testimony. The

force of testimonial evidence depends of course on how important inferentially is the event reported and the credibility of the source who reports it. But it also depends on how rare or improbable this event is. This fact was noted by both Hume (1777, 113) and LaPlace (1795, 114). As we will discuss later on, certain ingredients of Bayes' rule show us exactly how the rareness of the event reported interacts with event importance and source credibility in determining the inferential force or weight of testimonial evidence.

As both Daston (1988, 370 - 386) and Zabell (1988, 327 - 354) tell us, interest in credibility-testimony problems among probabilists all but expired; Zabell puts the expiration date in the late 19th century. Part of the reason they note is that probabilists had accepted a frequentistic interpretation of probability, believing that probabilities could only be determined or estimated enumeratively, or by counting operations. Such instances require replicable or repeatable events whose past occurrence can be tabulated in the form of relative frequencies, as in statistics. But, fortunately or unfortunately, no one keeps any statistics on attributes of our credibility as witnesses. In addition, as in law and intelligence analysis, testimonial evidence so often concerns singular, unique, or one-of-a-kind events that can never be repeated. Consequently, we must consider views of probability involving testimonial evidence that permit us to deal with the non-enumerative nature of testimonial credibility assessment.

3.0 ON THE ATTRIBUTES OF WITNESS COMPETENCE AND CREDIBILITY

One major contribution of the experience with testimonial evidence in law to our current efforts is the necessity for distinguishing between the competence and the credibility of witnesses; they are two quite distinct characteristics. Unfortunately, these two characteristics are frequently confused in some situations and have led to serious inferential errors. In fact, an inference that a person is credible just because he is competent is a glaring non sequitur.

3.1 Witness Competence Attributes

The manner in which testimonial evidence is assessed in our legal system illustrates how the competence and the credibility of a source of testimony are entirely different characteristics. This distinction appears in a note attached to the definition of competence given in Black's Law Dictionary. This note says (1951, 355):

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.

We will add other attributes besides veracity as far as the witness's credibility is concerned. We 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 we know of. In news reports, as well as in intelligence documents we have read, there are 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 American Federal Rule of Evidence FRE 601 which says: "Every person is competent to be a witness except as otherwise provided in these rules" (2006, 73 - 74).

But FRE 602 requires that ordinary or lay witnesses must have "personal knowledge" of the event(s) about which he/she testifies (2006, 75 - 760). This is an important rule to which we will refer to again shortly in our 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 the witness testifies, or had access to the information given in this testimony. This rule 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 the intended testimony, this surely goes against witness 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 was observed. So we have labeled access and understanding as the two major attributes of the competence of testimonial evidence. But there are, in addition, other attributes of witness competence that can be identified in intelligence analysis that we will mention in Section 4.0.

3.2 Witness Credibility Attributes: The Standard Analysis

Suppose a witness Mary has made a relevant observation of an event she has reported to us and has given us an intelligible account of it. Now the question is: can we believe that the event Mary has just reported to us actually occurred? The first crucial matter in assessing Mary's credibility is to distinguish between her testimony that a certain event occurred and the actual occurrence of it. Suppose Mary has reported E^* that event E occurred. Clearly, E^* and E are not the same events. Just because the Mary reports E^* does not entail that event E did occur. The possibility always remains open that event E did not occur, which we label as E^C , read "E-complement, or not-E". So, from evidence E^* we must infer whether event E occurred or did not.

Suppose we later learn with certainty that event E did not occur; but Mary has told us that it did occur. We might naturally believe that Mary was either untruthful or she was mistaken during her observation. This suggests that veracity and observational accuracy are the only attributes of witness credibility. But there is another credibility attribute to consider when we consider the very important relationship between knowledge and evidence, testimonial evidence in our present case. To illustrate, suppose we ourselves were not privy to the occurrence or nonoccurrence of event E, an important event in an inference task we are performing. One of us says: "Mary is a competent source we should consider; she will know whether event E occurred or did not". She may in fact have volunteered to tell us about her observation. But if we believe she is competent, she should also know whether E occurred or did not.

In 1989 one of us wrote a paper concerning knowledge, probability and credibility (Schum, 1989, 39 - 62). This work explored what it means to say that a person knows that an event has occurred. This happens to be an issue arousing great controversy among epistemologists. In our present example, Mary tells us E*, that event E did occur. The question is how do we tell whether or not Mary knows that E occurred? In epistemology there exists what is termed the standard analysis of knowledge, which is traceable all the way back to Plato's Theaetetus (Cohen, L. J., 1986, 19). On this analysis, knowledge is justified true belief. In the case of Mary, we say on this analysis that Mary knows that event E occurred if: (i) E did occur, (ii) Mary was justified in believing that E occurred (i.e. Mary got good evidence that E occurred), and (iii) Mary believed the evidence she obtained.

In the paper cited above (Schum, 1989, 39 -62) appears an account of the kinds of controversy surrounding the standard analysis of knowledge and what other philosophers have said about this controversy. It happens that there is no account of the requisites for knowledge that is free of controversy. Even so, the standard analysis seemed a good place to start in exploring its consequences as far as the credibility attributes of testimonial evidence are

concerned. But there is one additional issue about where this analysis would lead us. This standard analysis concerns whether our source Mary "knows" whether event E occurred or 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 Mary 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 Mary, who we believe to be a competent source of information about event E. Further, we do not know how good was the sensory evidence that Mary obtained. Nor do we know about the actual extent to which Mary believed her sensory evidence regarding E, if she took this sensory information into account at all. Finally, we do not know if Mary is reporting in accordance with what she believes about event E. Figure 1 is an account of the standard analysis regarding Mary's knowledge of event E (Part A) and how this influences our uncertainties [Part B] ; this is where the three major testimonial credibility attributes that we have taken into account arise.

INSERT FIGURE 1 ABOUT HERE

On this analysis we have a credibility-related chain of reasoning suggested by the standard analysis. The three links in this chain identify the three major credibility attributes we have taken into account in our analysis. The standard analysis has allowed us to identify a very important attribute that is rarely mentioned in other accounts of witness testimonial credibility; witness objectivity. The following comments illustrate the importance of this attribute and also illustrate the first of several tests we have made showing the degree to which this analysis captures what really matters in the assessment of witness credibility.

3.3 Verifying Witness Credibility Attributes

We have tried our best to identify attributes of the credibility of human sources of evidence that are defensible and as complete as we can determine. The first test we have made

of this analysis of witness credibility based on the standard analysis of knowledge is to consider the extent to which it corresponds with the common experiences we all have had in hearing people report about events they have observed.

3.3.1 From Common Experience

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 Mary really believe that E occurred, as she reported to us? Relating veracity to a witness's beliefs is quite natural among persons who have thought carefully about what veracity means. For example, in her book on lying, Sisela Bok says (1989, 13):

When we undertake to deceive others intentionally, we communicate messages meant to mislead them, and make them believe what we ourselves do not believe.

The American evidence scholar John H. Wigmore tells us (1937, 605):

A lie is the intentional introduction into another's mind of a belief that is not in harmony with what the actor himself supposes to be the truth.

Thomas Paine tells us (1794, 8):

Infidelity does not consist in believing or disbelieving; it consists in professing to believe what he does not believe.

Of the three credibility attributes veracity has certainly been the most frequently misinterpreted. It does not refer to what actually happened but to a person's belief about what happened. Notice in Figure 1 that veracity forms the foundation stage of our credibility-related chain of reasoning.

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 surmise or upon desires. In our example, the next question we ask concerning our source Mary is: If Mary 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? However, our beliefs are elastic and are known to change over time. This is where a witness's memory becomes such an important consideration. Suppose Mary made her observation concerning event E six weeks ago. Does she believe the same thing about this event now, while providing her testimony, as she did at the time she made her observation? The role of memory in witness's testimony is a much-investigated issue as Gillian Cohen explains (1999, 3 - 18).

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 sensory and general physical condition at the time of our observation. So, the final attribute we 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 Mary's relevant senses at the time of her observation, and what were the conditions under which she made this observation?

3.3.2 Wigmore on Personal Knowledge

Our second test of the standard analysis interpretation of attributes of witness credibility comes from the American evidence scholar whose works we so frequently draw upon. His comments on the "personal knowledge" requirement in the American rule of evidence FRE 602 are so valuable. 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 (1940, §650, 755):

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:

- 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.

We 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 the standard analysis that forms the basis for our present account of the attributes of testimonial credibility. The reader can observe how close Wigmore's chain of reasoning compares with the chain of reasoning shown in Figure 1-B, when we also include Wigmore's views about veracity.

3.3.3 Trial Experience in Law

We come now to the final and most important test of our analysis of the attributes of witness credibility. This test involves observing how well the three attributes veracity, objectivity, and observational sensitivity capture the kinds of witness credibility questions revealed in the five

hundred year history of our Anglo-American adversarial system of trials at law. Figure 2 shows the twenty - three credibility-related questions we have identified.

INSERT FIGURE 2 ABOUT HERE

First observe that of these twenty-three questions or tests, sixteen of them can be categorized as being specific to one of the three attributes; seven under veracity, three under objectivity, and six under observational sensitivity. For example, questions regarding sensory defects and observational conditions obviously concern the observational sensitivity of the witness under the conditions of observation. Questions regarding what a witness expected to observe and memory-related factors concern the objectivity attribute. Questions regarding a witness's character and reputation, and questions regarding influence or corruption among witnesses, bear upon veracity. As we have noted, not all of the questions shown involve possible answers that would be admitted at trial. But they are commonly listed because answers to them may assist in fact investigation in preparation for trial.

Next observe that there are four questions that are less specific to these three attributes and can bear upon two of them. For example, any observational instructions or objectives given to a witness in advance of an observation may influence either witness objectivity or sensitivity. Questions regarding witnesses who contradict themselves may reveal only failures of memory, and thus objectivity, or they may bear upon a witness's veracity; the witness has difficulty in keeping his story straight.

But we come to three classes of credibility questions that can bear upon any of the three credibility attributes. Two of these questions are age-old and were certainly asked long before the emergence of our Anglo-American adversarial trial system. The best examples refer to the existence of other evidence that either contradicts or conflicts with what a witness has just told us. A good example of contradicting evidence involves situations which witness like Mary testifies E*

that event E occurred, when another witness Harold testifies E^{C*} that event E did not occur. Leaving aside Harold's credibility for a minute, the existence of this contradictory evidence does not specify which of Mary's credibility attributes may be in question. Perhaps she was untruthful but she could also have been mistaken or not objective in forming her belief. But now we come to Harold's credibility; just because he tells us that E did not occur, does not mean that it did not occur. Perhaps he was the one lying, mistaken or not objective. The same issues apply to other witnesses who provide conflicting or divergent evidence about other events. And, some forms of prior inconsistencies can be related to any of our witness's credibility attributes.

We have taken our classification of questions shown in Figure 2 to be the strongest evidence of the suitability of our identification of the three attributes of witness credibility suggested by the standard analysis of knowledge. Here we have an example of a controversial conjecture that, though imperfect in many ways, nevertheless provides us with a very useful heuristic for examining a very complex process. We did not find it necessary to list any other attributes in addition to veracity, objectivity, and observational sensitivity under the conditions of observation in order to capture the credibility-related questions asked in law. With the comments we have made so far in mind, we are now ready to describe MACE, the system we have developed for assessing the competence and credibility of the informants who provide us with HUMINT evidence.

4.0 A SYSTEM CALLED MACE

The MACE system has been under development for quite some time; it was first suggested in a prototype version constructed in 1990. Events since the attacks on the World Trade Center and the Pentagon in America on September 11, 2001, and the more recent attacks in London on July 7, 2005, have aroused great interest in the necessity for having more and better sources of HUMINT. It is vital of course that we have experience-based and useable methods for assessing the competence and credibility of these HUMINT sources. In so many

cases intelligence professionals have told us that they are not sure which questions they should be asking about HUMINT sources. As we have discussed, we have a very old legacy of important questions suggested by the experience in our Anglo-American adversarial trial system to ask about our sources of HUMINT regarding their competence and credibility. MACE allows users two different but entirely compatible ways to draw probabilistic conclusions about the extent to which we can believe what sources of HUMINT report to us based upon evidence we have about their competence and credibility.

4.1 Marshaling Competence and Credibility Evidence about a HUMINT Source

Suppose we have a HUMINT source who has just provided us with a report of an interesting event this person claims to have observed. To what extent should we believe that this event actually occurred? In answering this question we must first marshal the existing evidence we have regarding this source's competence and credibility. MACE provides a coherent and convenient way to marshal evidence we have about this source's competence and credibility that have been identified in the many centuries of experience in law and summarized in Figure 2. Some of the questions or tests of a source's competence and credibility have been altered slightly to make them relevant to the intelligence context. We have also added a few questions or tests that have been suggested to us by experienced intelligence professionals. This marshaling operation is crucial since it forms the basis for both of the probability methods we employ in grading the extent to which we might believe what this source has reported to us. Of great importance is that it also allows us to keep track of what we know about a source in a systematic way showing what specific questions about a source's credibility and competence have been answered by the evidence we have. Finally, MACE alerts the user to make separate judgments regarding a source's competence and credibility.

4.1.1 Marshaling Evidence on Competence Questions

Here are the five questions or tests of a HUMINT source's competence MACE asks the user to try to answer based on the evidence we have that is relevant to answering these questions.

1. Did this source actually make the observation being claimed or have access to the information reported?
2. Does this source have an understanding of what was observed or have any knowledge or expertise regarding this observation?
3. Is this source generally a capable observer?
4. Has this source been consistent in his/her motivation to provide us with information?
5. Has this source been responsive to inquiries we have made of him/her?

Now, the user of MACE marshals existing evidence in answer to these five competence-related questions. We may have more than one answer to any of these questions. The next step requires a judgment on the part of the user regarding each of these five questions; there are four possible responses allowed:

- The evidence on this question **favours** this source's competence,
- The evidence on this question **disfavours** this source's competence,
- **I cannot decide** whether the evidence on this question favours or disfavours this source's competence,
- There is **no available evidence** bearing on his question.

4.1.2 Marshaling Evidence on Credibility Questions

Here are the questions MACE asks the user to try to answer about this source's veracity, objectivity, and observational sensitivity under the conditions of observation, based on the available evidence we have about his source.

For Veracity:

1. Has this source told us anything that is inconsistent with what this source has just reported to us?
2. Is this source subject to any outside influences?
3. Could this source have been exploited in any way in this report to us?
4. Is there any other evidence from other sources that contradicts or conflicts with what this source has just reported?
5. Is there any evidence from other sources that corroborates or confirms what this source has just told us?
6. What evidence do we have about this source's character and honesty?
7. What does this source's reporting track record show about the source's honesty in reporting to us?
8. Is there evidence that this source tailored this report in a way that this source believes will capture our attention?
9. Are there collateral details in this report that reflect the possibility of this source's dishonesty?
10. Evidence regarding the demeanor and bearing of this source during the interview?

Here the user of MACE marshals existing evidence in answer to these ten veracity-related questions. We may have more than one answer to any of these questions. The next step requires a judgment on the part of the user regarding each of these ten questions; there are four possible responses allowed:

- The evidence on this question **favours** this source's veracity
- The evidence on this question **disfavours** this source's veracity,
- **I cannot decide** whether the evidence on this question favors or disfavours this source's veracity,
- There is **no available evidence** bearing on his question.

For Objectivity:

1. Is there evidence about what this source expected to observe during the reported observation?
2. Is there evidence about what this source wished to observe during the reported observation?
3. Was this source concerned about the consequences of what this source believed during the observation?
4. Is there any evidence concerning possible defects in this source's memory?
Also, how long ago did this source's observation take place?
5. Is there any other evidence from other sources that contradicts or conflicts with what this source has just reported?

The user of MACE marshals existing evidence in answer to these five objectivity-related questions. Again, we may have more than one answer to any of these questions. For each objectivity questions the user makes one of the following judgments:

- The evidence on this question **favours** this source's objectivity
- The evidence on this question **disfavours** this source's objectivity,
- **I cannot decide** whether the evidence on this question favors or disfavours this source's objectivity,
- There is **no available evidence** bearing on his question.

For Observational Sensitivity under the Conditions of Observation:

1. The source's sensory capacity at the time of the observation?
2. The conditions under which the observation took place?
3. The source's track record of accuracy in previous reports?
4. Is there any other evidence from other sources that contradicts or conflicts with what this source has just reported?
5. Are there collateral details in this report that reflect the possibility of this source's inaccuracy?

As the user did for veracity and objectivity, the user marshals existing evidence in answer to these five observational sensitivity questions. Again, we may have more than one answer to any of these questions. For each question the user makes one of the following judgments:

- The evidence on this question **favors** this source's observational sensitivity
- The evidence on this question **disfavors** this source's observational sensitivity,
- **I cannot decide** whether the evidence on this question favors or disfavors this source's observational sensitivity,
- There is **no available evidence** bearing on his question.

As this evidence marshaling and judgmental process proceeds, MACE keeps track of the judgmental responses the user provides to each competence and credibility questions. At the time a decision is required about the extent to which we can believe what this source has reported to us, MACE provides a summary of these responses that is illustrated in Figure 3. The numbers shown in the boxes in this figure are just examples showing the kinds of summary information MACE provides the user regarding the source's present report. This summary sets the stage for the first of two probability methods MACE provides for assessing the extent to which

we can believe what the source has reported, based on the evidence we have about this source's competence and credibility.

INSERT FIGURE 3 ABOUT HERE

4.2 Baconian Probability Assessments

The first issue MACE allows the user to address is: how much evidence do we have about this source's competence and credibility and how completely does it answer what we believe to be relevant questions in assessments of this source's competence and credibility? These two very important questions stem from L. Jonathan Cohen's work on a system of probability he links to the work of Sir Francis Bacon (e.g. Cohen, 1977; 1989). Bacon had advised that we not seek to verify hypotheses simply by accumulating evidence favorable to their occurrence. What we should do instead is to subject our hypotheses to evidential tests designed to eliminate them; the hypothesis that best resists our eliminative tests is the one in which we should place the greatest confidence. But Bacon also advised that the evidential tests to which we subject hypotheses should all involve different relevant matters. The variative nature of our evidential tests will show the array of conditions under which any of our hypotheses will continue to hold up or resist being eliminated. In short, there is no benefit to performing the same test over and over again. This may of course increase our confidence in the reliability of this single test, but it says nothing about the array of different conditions under which we might expect hypothesis to hold up. The major contribution we have received from the field of law is an account of the array of different tests to which we should subject sources of HUMINT as far as their competence and credibility is concerned. So, variety is the spice of life in competence and credibility assessments.

Two matters are of the greatest importance in the use of the Baconian part of MACE. The first is: how much favorable evidence do we have concerning answers to the questions or tests involving this source's competence and credibility? This calls for judgments on the part of the

users of MACE as we noted above. The second, and very important matter, involves our consideration of how many relevant questions remain unanswered by the evidence we have so far? This is one of the major reasons why we have adopted this Baconian view of probability. Jonathan Cohen's view is that the weight of evidence depends not only on how much favorable evidence we have but also on the number of questions that remain unanswered by available evidence. Cohen was the first to generate a system of probability congenial to Bacon's views about the eliminative and variative nature of the inductive testing of hypotheses. His view of the weight or force of evidence involves completeness issues not addressed by any other view of probability known to us, including the Bayesian view that we will discuss below. Bayes' rule does provide very useful ways of grading the strength of evidence we do have, but it remains silent on the issue: how do the questions that remain unanswered affect our judgment about whether to believe what this HUMINT source has just reported to us?

Figure 3 will allow us to give an account of how the user of MACE makes judgments using the Baconian part of MACE. First, we believed that many users of MACE would share the same aversion to assessing subjective probabilities on a numerical scale that so many other intelligence professionals exhibit. No such subjective probability assessments are required using the Baconian part of MACE. MACE does require subjective judgments regarding the following questions:

- Does this item of evidence favor or disfavor a competence or credibility attribute to which this evidence is relevant, or can the user not decide whether this evidence is favorable or unfavorable? In so many situations, evidence we have about a HUMINT source may be ambiguous and we cannot say what it is telling us.
- How relatively important are the questions being asked about this source's competence and credibility in this particular situation involving this particular report?
- On balance, does the available evidence favor or disfavor this source's competence in providing this report, keeping in mind the number of relevant questions that remain unanswered?; and does the available evidence favor or disfavor each of the three

credibility attributes: veracity, objectivity, and observational sensitivity, keeping in mind the number of relevant questions that remain unanswered?

- On balance, how strongly does the accumulated evidence favor or disfavor our believing the report this source has just given us, keeping in mind the total number of questions that remain unanswered by the evidence we have?

INSERT FIGURE 3 ABOUT HERE

For examples of these judgments, now consider Figure 3 and the competence assessment shown in the first row where we have five questions concerning this source's competence. Here we show a situation in which we have evidence items favorable to three of the questions concerning this source's competence and disfavoring evidence on one question regarding this source's competence. But we also have one competence question unanswered. Here the user might judge the competence evidence about this HUMINT source to be, on balance, favorable provided that the unanswered question did not involve something of major importance, such as whether the source did have access to the information being reported or whether the source did make a relevant observation. The Baconian judgment here is the only judgment users of MACE make regarding a source's competence. There are no Bayesian assessments of the probabilistic strength of this evidence. The major reason is that we do not have a model that says how we should combine any conventional probabilities that may be associated with the five questions we ask about a source's competence. But we do have a model that shows us how to combine the evidence regarding this source's veracity, objectivity, and observational sensitivity.

For the credibility assessments for this source, consider the next three rows of Figure 3. First, Mace asks us to try to answer ten questions regarding this source's veracity. In the example shown, the user has judged the evidence to be favorable to just one of these questions, disfavoring to four of these questions, undecided about the evidence on two of these questions,

and three of the questions remain unanswered. This account is hardly an endorsement of this source's veracity and so the user would be justified in concluding that this source has reported an event this source does not believe happened. Regarding objectivity, for which there are five questions, the user has judged that none of these questions has been answered favorably, one was answered unfavorably, the user was undecided about two of these questions, and there were two questions that remain unanswered. This is not good evidence that the source based a belief on sensory evidence and may have done so for other reasons. Finally, regarding observational sensitivity, for which there are five questions, the user judged that two questions were answered favorably, one was answered unfavorably, was undecided about one question, and one question remained unanswered. The user's conclusion here might well rest upon our obtaining evidence bearing on this unanswered question if it might have been crucial. Suppose the unanswered question concerns the source's sensory and general physical condition at then time of the observation. If the source was intoxicated at this time, or if the source suffered from a sensory disability, this would not favor this attribute.

Finally, consider the credibility summary shown in the last row of table 3. Here MACE provides the user a summary account of how the entire collection of twenty credibility questions were answered or unanswered by the available evidence. In the example, only three questions were answered favorably, six were answered unfavorably, five questions produced the undecided response, and six questions remain unanswered. If there is no further evidence bearing on these six unanswered questions, the MACE user might rest content on a conclusion that we should not believe what this source has reported to us. This example simply illustrates how a source judged to be competent might not be credible.

The user of MACE might quit at this point and report a decision based just on this Baconian analysis about whether we should believe the report provided by the person whose competence and credibility have been examined. The Baconian analysis provides no number or range of numbers indicating how strongly the user believes this decision to be correct. But what

MACE does provide is a careful account of the evidence-based reasons why this conclusion was reached. Summaries can easily be provided of the favoring and disfavoring evidence bearing on the source's competence and credibility, as well as an account of relevant questions that have not been answered by available evidence. We return to the source "Curveball" mentioned in Section 1.0. If this person had been subjected to the Baconian analysis just described, the events of the past four years might have taken an entirely different course. All but a very few of the competence and credibility questions identified in MACE were ever answered about Curveball. The few that were answered were quite unfavorable to his competence and credibility.

There is one matter we need to address before we consider the Bayesian methods MACE provides. We have said nothing yet about how a user might judge the relative importance of the credibility attributes: veracity, objectivity, and observational sensitivity. In our account of the Bayesian view of credibility assessment we will make use of the credibility chain of reasoning illustrated in Figure 1-B. There are three links in this chain corresponding to the attributes: veracity, objectivity, and observational sensitivity. If there is a weak link in this chain of reasoning, does it matter where it is located? As we have shown elsewhere, the answer depends on whether we adopt a Bayesian or a Baconian point of view (Schum, 1994, 301-306). On a Bayesian view, it does not matter where in the chain the weak link is located, since the chain cannot be any stronger than its weakest link. So, in a Bayesian view, these three links seem to have equal importance. But on the Baconian account, it does matter where in the chain a weak link is located. As we now explain, this view suggests that the veracity-related foundation link is the most important.

As Cohen tells us (1977, 73): "Proof is something that may depend on what is probably inferable from known facts, but not on what is certainly inferable from probable ones". In other words, a strong foundation for a weaker argument is preferable to a weak foundation for a strong argument. First, consider the top link in the chain involving the source's observational sensitivity. The only interface we, the users of MACE, have with the reported event(s) comes only indirectly

through the medium of the source's senses. Therefore, we might believe it obvious that the sensitivity or accuracy of the source's senses is of primary importance. Suppose we have a competent source who has, on evidence, nearly perfect sensory capability that he/she exercised under nearly perfect observational conditions. By itself this evidence seems to provide a strong argument that the events reported by this source did occur.

But now consider the foundation stage for this argument involving the source's veracity. Suppose we have evidence-based reasons for believing that this source's veracity is highly questionable; the source does not believe what he just told us. In this case we have a weak foundation for a strong argument. Cohen would say that we would be better off having evidence from a source with strong veracity credentials but weaker observational sensitivity credentials. In this case we have a strong foundation for a weaker argument. On a Bayesian interpretation, we would say that the strength of the chain is weakened whether the weak link involves veracity or observational sensitivity, or objectivity for that matter. In another work, we have shown how Jonathan Cohen's Baconian view and the Bayesian view are not natural adversaries but are quite complementary (Schum, 1991, 99 - 143). Each view tells us something important about assessing the credibility of persons who tell us about what they have observed.

4.3 Subjective Bayesian Probability Assessments

As we have just illustrated a user of MACE could draw a conclusion about whether we should believe an item of HUMINT from a certain source based just on the Baconian methods we have described. But it can be argued that this method does not tell us how strongly the evidence we do have favors or disfavors each of the credibility attributes; veracity, objectivity, and observational sensitivity. Further, the Baconian method does not supply us with any numerical gradation concerning the extent to which we might believe an item of HUMINT provided by a human source. As we now describe, a Bayesian analysis provides both of these assessments. But this method does require the user of MACE to provide subjective probability assessments of

propositions concerning the HUMINT source's veracity, objectivity, and observational sensitivity. We note here that we will never have any statistics bearing upon these three attributes for any source, and so these probabilities must be in the form of subjective judgments. We first look at what probabilities are required; we then look at how the MACE user provides them; and look finally at how MACE combines them in calculating a final assessment about the extent to which we can believe what the source has reported to us, based on evidence we do have about this source's credibility.

We first consider again the chain of reasoning suggested by the standard analysis we mentioned in Figure 1-B. We repeat this chain of reasoning in Figure 4 showing an abstract situation in which we have testimonial evidence E^* , from a source assumed to be competent, that event E occurred. Our major inference here concerns the extent to which we can believe that event E occurred, based on the evidence E^* . Using the odds-likelihood version of Bayes' rule, we have:

$$\frac{P(E | E^*)}{P(E^c | E^*)} = \frac{P(E)}{P(E^c)} \frac{P(E^* | E)}{P(E^* | E^c)} \quad (\text{Eq. 1})$$

The term on the left is called the posterior odds of event E to event E^c [not- E], given evidence E^* . The term $P(E)/P(E^c)$ is called the prior odds of event E to event not- E . The term $P(E^*|E)/P(E^*|E^c)$ is called the likelihood ratio for evidence E^* . In Bayes' rule this likelihood ratio indicates the inferential force or weight of evidence E^* on E and E^c . This likelihood ratio, which we label L_{E^*} , shows the extent to which we have revised our prior odds on event E to event E^c into the posterior odds of E to E^c , now that we have evidence E^* . If $P(E^*|E) > P(E^*|E^c)$, the evidence favors E over E^c and so the posterior odds of E to E^c will be greater than the prior odds of E to E^c . If $P(E^*|E) < P(E^*|E^c)$, the evidence favors E^c over E and so the posterior odds of E to E^c will be less than the prior odds of E to E^c . If $P(E^*|E) = P(E^*|E^c)$, there will be no change in the prior odds of E to E^c .

INSERT FIGURE 4 ABOUT HERE

As Figure 4 illustrates, the linkage between E^* and events $\{E, E^C\}$ is not direct, there are two links interposed. One link involves whether the source believes the event being reported and the other concerns whether the source based this belief on sensory evidence about event E . This figure also shows what Bayes' rule says are the necessary probabilistic ingredients at the three stages of reasoning from evidence E^* to the events $\{E, E^C\}$. These ingredients are also in the form of likelihoods as defined as follows.

For the veracity link, we need:

- $PE^*|E_B$): How likely is this report of E , given that the source believes that event E occurred?
- $PE^*|E_B^C$): How likely is this report of E , given that the source does not believe that event E occurred? Notice that this includes two possibilities: the source believes that E did not occur, or the source has no belief at all about event E . Perhaps this source was told what to tell us.

For the objectivity link, we need:

- $P(E_B|E_S)$: How likely is this source to have believed event E , given that this source based this belief on sensory evidence of event E ?
- $P(E_B|E_S^C)$: How likely is this source to have believed event E , given that this source did not base a belief on sensory evidence of event E . This would indicate that the source based a belief that event E occurred because he/she expected or desired this event to have occurred, or the source may have been told to believe that event E occurred.

For the observational sensitivity link, we need:

- $P(E_S|E)$: How likely is the source to have based a belief on sensory evidence of event E , given that event E occurred?
- $P(E_S|E^C)$: How likely is the source to have based a belief on sensory evidence of event E , given that event E did not occur. These two likelihoods tell us how good the source's sensory evidence might have been at the time of observation.

Decomposing the link between evidence E^* and events $\{E, E^C\}$ shows that we need three pairs of likelihood judgments, one pair for each credibility attribute. In doing so, we have just decomposed the two probabilities in $L_{E^*} = P(E^*|E)/P(E^*|E^C)$. As shown elsewhere (Schum, 1994, 2001, 333 - 337) these two probabilities become:

$$P(E^*|E) = P(E_S|E)[P(E_B|E_S) - P(E_B|E_S^C)][P(E^*|E_B) - P(E^*|E_B^C)] + P(E_B|E_S^C)[P(E^*|E_B) - P(E^*|E_B^C)] + P(E^*|E_B^C). \quad (\text{Eq. 2})$$

$$P(E^*|E^C) = P(E_S|E^C)[P(E_B|E_S) - P(E_B|E_S^C)][P(E^*|E_B) - P(E^*|E_B^C)] + P(E_B|E_S^C)[P(E^*|E_B) - P(E^*|E_B^C)] + P(E^*|E_B^C). \quad (\text{Eq. 3})$$

Again, the ratio $L_{E^*} = P(E^*|E)/P(E^*|E^C)$ indicates the force of testimony on events $\{E, E^C\}$.

We next faced the task of trying to make these probability judgments as easy as we could for persons having an aversion to providing numerical probability assessments. We did in fact provide a way in MACE for users to generate the necessary three pairs of likelihoods without using any numbers at all; but by drawing boxes in a two-dimensional probability space. The likelihood spaces are shown in Figure 5 for the attributes: veracity, objectivity, and observational sensitivity.

INSERT FIGURE 5 ABOUT HERE

Users of MACE are provided with elaborate instructions about how to provide their judgments on these likelihood spaces. First, MACE provides summaries of the available evidence regarding a source's credibility attributes: veracity, objectivity, and observational sensitivity. MACE instructs them to consider all of the evidence we have about an attribute before making their judgments. We also considered the precision with which MACE users could make these judgments. Perhaps no one would believe exact numerical judgments for any credibility attribute. For example, if the user said that $P(E^*|E_B) = 0.67$ and $P(E^*|E_B^C) = 0.21$, few persons would take these assessments seriously. So, the user is encouraged to make judgments about intervals of likelihoods. Readers may recognize these probability judgment intervals as second-order

probabilities, as they are often called. They reflect uncertainty about what a probability might be. This is where the concept of a response "box" arises. The manner in which MACE users make these pairs of likelihood judgments is illustrated in Figure 6 for the veracity attribute. The same discussion will apply to the other two attributes.

INSERT FIGURE 6 ABOUT HERE

The location of a user's response box first indicates his/her belief about whether the available evidence favors or disfavors the source's veracity. If the box is above the positive diagonal this means the user judges the evidence to favor the source's veracity, since the box will show values of $P(E^*|E_B) > P(E^*|E_B^C)$. The closer this box is to the upper left corner, the stronger the user believes it to be in favoring the source's veracity. A box below the diagonal indicates the user's belief that the evidence disfavors the source's veracity; in this case the box will show values of $P(E^*|E_B^C) > P(E^*|E_B)$. The closer this box is to the lower right corner, the stronger the user believes the evidence disfavor's the source's veracity. The size of a response box indicates the user's confidence in the probability assessments being made. Larger boxes reflect greater judgmental uncertainty in assessing a pair of likelihoods. Reflecting back on the Baconian assessment, the user might recall how many "I can't decide" assessments were made and how much evidence was available on an attribute. This would surely bear on judgmental confidence. Of course the user constructs just one response box for each of the credibility attributes.

MACE provides a variety of forms of user assistance in making these probability assessments. This first thing it does is to alert the user to any inconsistencies that occur between the Baconian and Bayesian assessments. For example, suppose the user decided in a Baconian assessment that, on balance, the evidence disfavored the source's veracity. If the user then, in the Bayesian analysis, constructs a box above the diagonal, a message will appear alerting the user to this inconsistency and suggest that he/she reconsider both assessments. MACE also

provides a Bayesian "Wizard" that can provide a verbal statement about what a constructed response box is telling us. Following is one basis for this Wizard.

Some readers will notice that each of the three response spaces can provide the same kinds of information as do receiver operating characteristic [ROC] curves common in the theory of signal detection. These ROC curves capture many subtleties of a user's responses. First, they show the strength of evidence favoring or disfavoring a source's credibility attribute. But they can also indicate any source biases revealed in the evidence for this source. The reader should note in Figure 2 that we have listed a form of bias associated with each credibility attribute. For example, suppose in Figure 6 that the user constructs a box near the center and top of the response space, but slightly above the positive diagonal. In this box both values of $P(E^*|E)$ and $P(E^*|E^C)$ are very large, even though $P(E^*|E) > P(E^*|E^C)$. What this indicates is that the user believes this source has a bias in favor of making this report whether he believed it or not. Finally, MACE shows the user that boxes other than perfect squares are perfectly appropriate since the user may be more confident in the assessment of one of the probabilities in a pair than he/she is about the other.

Now suppose that the user has provided one response box for each of the three credibility attributes for the source under consideration. MACE takes over at this point and first identifies the four numerical corner points of each of the user's response boxes. There are 64 possible combinations of these corner points, taken one each from each box. Then, using Equations 2 and 3 above, MACE calculates values of $P(E^*|E)$ and $P(E^*|E^C)$ and takes their ratio. Thus we have 64 values of $L_{E^*} = P(E^*|E)/P(E^*|E^C)$ for the user's assessments and MACE determines an interval containing all of them. This interval thus contains all of the assessments of the force of testimonial evidence E^* on events $\{E, E^C\}$. Now, the current Bayesian version of MACE also asks the user to assess a range of prior probabilities of the event(s) reported by the source, before he/she begins marshaling the competence and credibility evidence for this source. This judgment is quite important since the prior odds on E relative to E^C , $P(E)/P(E^C)$, says

how likely the reported event seems before we had evidence of its occurrence. This term in Baye's rule allows the user to assess how probable or improbable the reported event seems. By such means the user can indicate the extent to which the reported event is rare or improbable. We mentioned in Section 2.2 that the early probabilists recognized the importance of event rareness in credibility-testimony problems.

Taking this prior probability interval in to account, MACE calculates an interval of posterior odds of the form $P(E|E^*)/PE^C|E^*)$ for every combination of priors and calculated L_{E^*} values and displays the ranges of values as shown in Figure 7. The ranges shown in this figure are just examples. These posterior odds ranges form the basic product of the Bayesian analysis in MACE. They show ranges of uncertainty about the extent to which we may believe or disbelieve a HUMINT source's report, given what the user of MACE has assessed to be the strength of the credibility evidence we do have about this source.

INSERT FIGURE 7 ABOUT HERE

What is certainly true for HUMINT source credibility assessments is that there may often be great disagreements among various persons who do the assessments or among other persons who use these assessments. What MACE readily provides is the capability for doing instant sensitivity analysis showing what would be the consequences of changing any of the three assessments of veracity, objectivity, or observational sensitivity, as well as changes in a prior odds assessment. In addition, different analysts making credibility judgments about the same source and the same report can easily compare the consequences of their different assessments using MACE. Thus, MACE has a direct interactive capability. In the example shown in Figure 7, someone may argue that the prior odds on event E are set too low; the example shows prior odds in the range of about 10:1 to nearly 70:1 favoring not-E over E. This says that E is quite rare or improbable to begin with. Or, another analyst may believe the evidence to be more charitable regarding the credibility of the source of the report being examined. Such alterations introduced

would cause changes in the location and breadth of the posterior odds interval MACE calculates using Bayes' rule.

5.0 IN CONCLUSION

Competence and credibility assessment for human sources of evidence is an astonishingly difficult task. We have certainly not answered all the questions that might be asked about this task in our present version of MACE. In fact our work so far shows that we have just begun to answer some of the questions that arise concerning HUMINT sources. Our work on MACE is continuing at present and we hope to make advancements in the following areas. First, a certain source, who we may call a primary source, provides us with an item of testimonial evidence whose believability is at issue. So we query other human or other sources to gain some evidence concerning this primary source's competence and credibility. But we understand that the competence and credibility of these secondary sources of evidence about our primary source is also at issue. In fact we encounter chains of human sources who support or impeach each others' competence or credibility. Such a situations has been likened to being in a "wilderness of mirrors" (Martin, 1980). At present MACE allows us to deal with chains of up three sources, but we may need to consider longer chains.

One major difficulty, that we all recognize, is that human sources of intelligence evidence often have more than one employer. The possibilities for deception are endless and may involve highly organized attempts to mislead us in various ways. Our hope is to make improvements in MACE that will allow us to more readily identify deceptive efforts. Finally, human sources of evidence provide other things besides testimonial reports on events they have allegedly observed. In many cases they provide tangible evidence in the form of documents, objects, images, and many other kinds of tangibles. But assessing the credibility of tangible evidence requires us to consider different credibility questions such as the authenticity of a tangible item including the chain of custody through which this item passed. We are also concerned about the

accuracy and reliability of the sensing or other device that produced the tangible. But we are also concerned about the competence and credibility of the human source who delivered this tangible item to us. Further versions of MACE will allow us to cope with these very difficult situations in which we can be deceived by others or be deceived by ourselves in the way we process tangible evidence.

In closing, we express our gratitude to the many practitioners and scholars in the field of law who have, over the centuries, recorded their experiences and their thoughts about assessing the competence and credibility of witnesses who provide testimonial evidence. We also express our thanks to the many probabilists who have made contributions to the analysis of the credibility-testimony problem. We hope our continuing work on the MACE system will be a worthy addition to the legacy we have inherited from these two fields. Failing this, we hope that at least we will have kept discussion alive on some issues that are as important as they are difficult.

REFERENCES

- Black's Law Dictionary 1951, 4th Ed. West Publishing Co. St. Paul, MN
- Bok, S. 1989 Lying: Moral Choice in Public and Private Life. Vintage Books, New York, NY
- Boswell, J. 1791 The Life of Samuel Johnson, LL. D. 1890 Croker ed. John Murray, Albemarle Street, London
- Cockburn, J., Green, T. 1988 Twelve Good Men and True: The Criminal Trial Jury in England, 1200 - 1800. Princeton University Press, Princeton, NJ
- Cohen, G. 1999 Human memory in the real world. In: Analysing Witness Testimony, eds Heaton-Armstrong, A., Shepard, E., Wolchover, D. Blackstone Press, London
- Cohen, L. J. 1977 The Probable and the Provable. Clarendon Press, Oxford
- Cohen, L. J. 1986 The Dialogue of Reason. Clarendon Press, Oxford
- Cohen, L. J. 1989 An Introduction to the Philosophy of Induction and Probability. Clarendon Press, Oxford

- Daston, L. 1988 Classical Probability in the Enlightenment. Princeton University Press, Princeton. NJ.
- Drumheller, T., Monaghan, E. 2006 On the Brink: An Insider's Account of How The White House Compromised American Intelligence. Carroll & Graf, New York
- Federal Rules of Evidence, 2005-2006 Edition. Thompson West, St. Paul, MN
- Hacking, I. 1975 The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction, and Statistical Inference. Cambridge University Press, Cambridge
- Hale, M. 1739 A History of the Common Law of England. 1971 C. M. Gray ed. University of Chicago Press, Chicago, IL
- Holdsworth, W. S. 1903 A History of English Law, vol. 1. Methuen, London
- Hume, D. 1777 Enquiries Concerning Human Understanding. 1989 ed P.H. Nidditch, 3d ed. Clarendon Press, Oxford
- Laplace, P. S. 1795 A Philosophical Essay on Probabilities. 1951 Dover reprint, New York
- Martin, D. C. 1980 A Wilderness of Mirrors. Ballantine, New York.
- Paine, T. 1794 The Age of Reason. 1984 ed Prometheus Books, Buffalo, NY
- Plucknett, T. 1956 A Concise History of the Common Law, 5th ed. Little Brown, Boston, MA
- Pollock, F., Maitland, F. 1968. The History of English Law before the Time of Edward I. 2nd ed Cambridge University Press, Cambridge
- Poole, A. L. 1955 From Domesday Book to Magna Carta: 1087 - 1216. Clarendon Press, Oxford
- Schum, D. 1989 Knowledge, probability, and credibility. Journal of Behavioral Decision Making. Vol. 2, 39 - 62
- Schum, D. 1991 Jonathan Cohen and Thomas Bayes on the analysis of chains of reasoning. In: Probability and Rationality; Studies on L. J. Cohen's Philosophy of Science. Eds. Eells, E., Maruszewski, T, Rodopi, Amsterdam, The Netherlands, 99 - 145
- Schum, D. 1994 The Evidential Foundations of Probabilistic Reasoning. John Wiley & Sons, NY; 2001 Northwestern University Press, Evanston, IL (paperback ed).
- Stenton, F. M. 1962 Anglo-Saxon England. Clarendon Press, Oxford
- Stigler, S. 1986 John Craig and the probability of history from the death of Christ to the birth of

- Laplace. Journal of the American Statistical Association, Vol. 81, No. 396, 879 - 887
- Todhunter, I. 1865 A History of the Mathematical Theory of Probability from the Time of Pascal To that of Laplace. 1965 Chelsea Publishing Co., New York
- Twining, W. L. 1994. Blackstone's Tower: The English Law School. Sweet & Maxwell, London
- Wells, C. 1914 The origin of the petty jury. Law Quarterly Review 107: 347 - 361
- Wigmore, J. H. 1937 The Science of Judicial Proof: as Given by Logic, Psychology, and General Experience and Illustrated in Judicial Trials. 3rd ed Little, Brown & Co. Boston, MA.
- Wigmore, J. H. 1940 Evidence in Trials at Law, Vol. 2, 3rd ed. § 650, at 755, Little, Brown & Co, Boston, MA
- Zabell, S. 1988 The probabilistic analysis of testimony. Journal of Statistical Planning and Inference 20, 327-354

FIGURES

ALL FIGURES DONE IN MS WORD

POINTS OF INSERTION NOTED IN THE TEXT

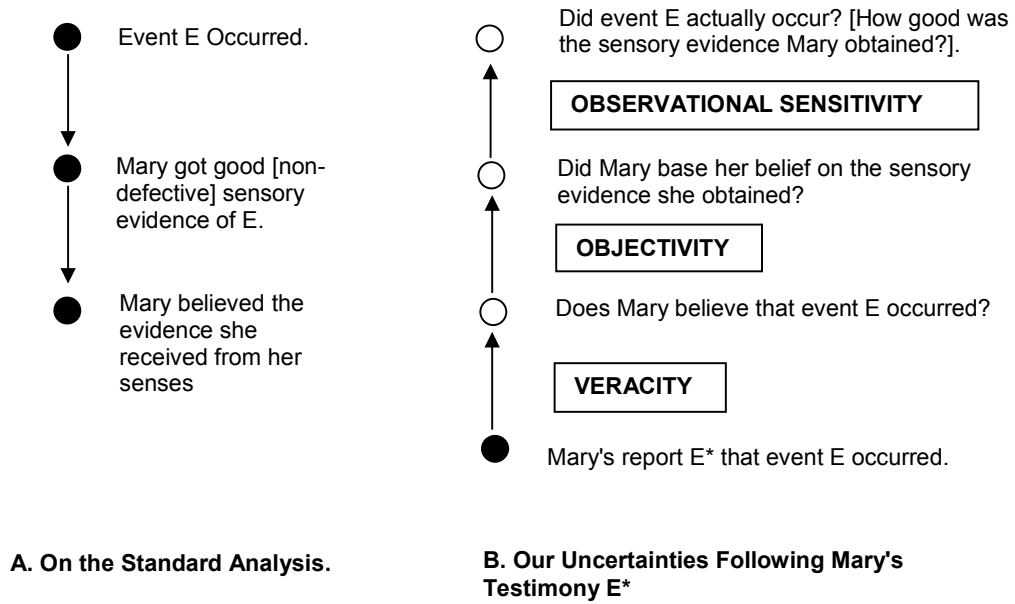


Figure 1. The Standard Analysis and Credibility Attributes.

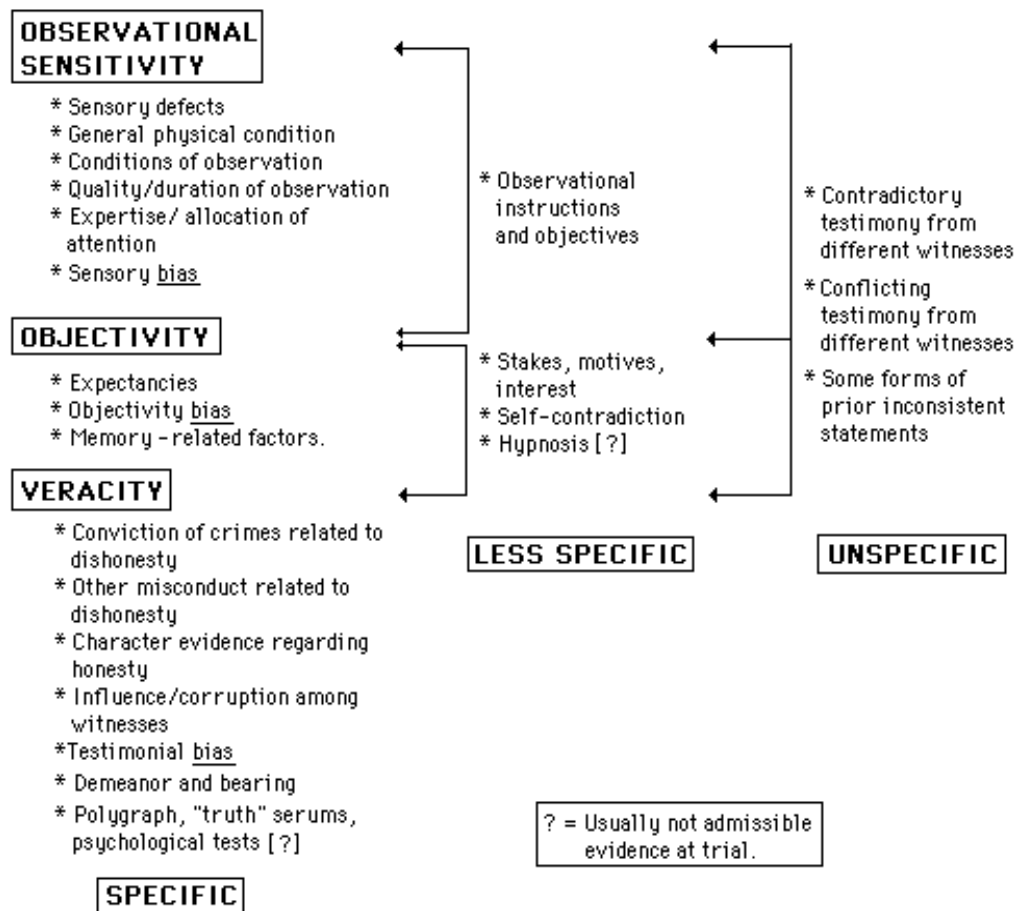


Figure 2. A Categorization of Credibility-Related Questions

JUDGMENTAL SUMMARY				
	<u>ANSWERED QUESTIONS</u>			<u>UNANSWERED QUESTIONS</u>
	<u>FAVORING</u>	<u>DISFAVORING</u>	<u>CAN'T DECIDE</u>	
COMPETENCE [5 Qs]	3	1		1
<hr/> <hr/>				
VERACITY [10 Qs]	1	4	2	3
OBJECTIVITY [5 Qs]		1	2	2
SENSITIVITY [5 Qs]	2	1	1	1
<hr/>				
CREDIBILITY SUMMARY	3	6	5	6

Figure 3. Competence and Credibility Attributes Judgments

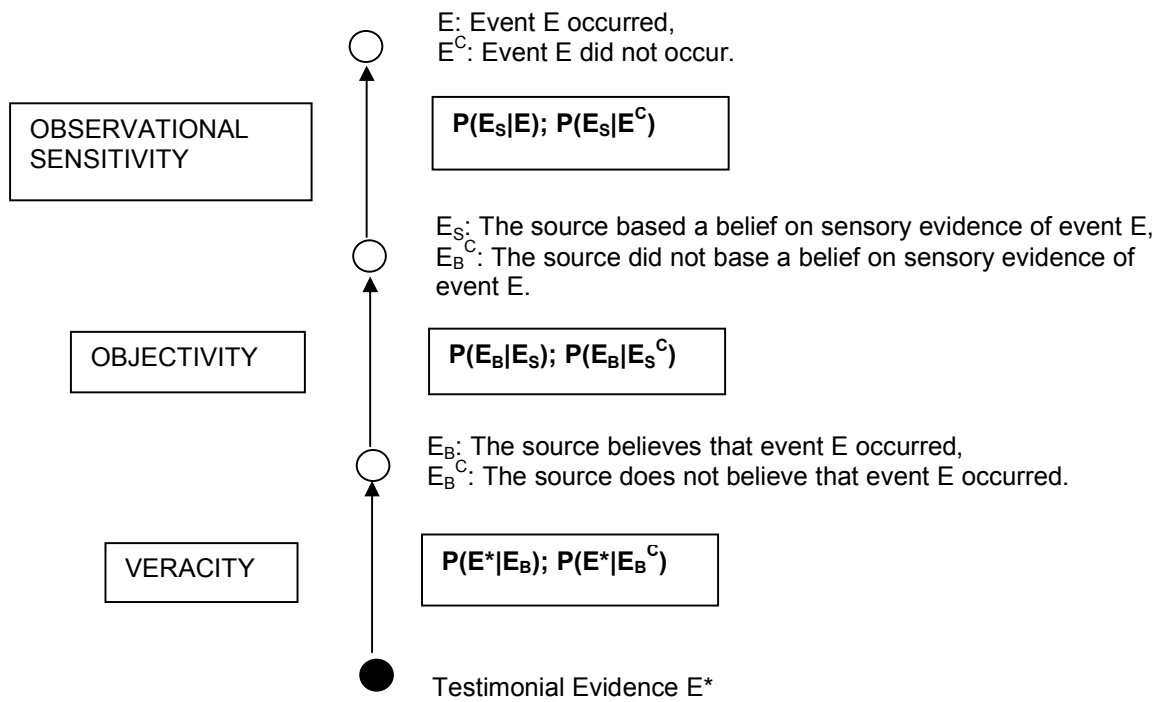


Figure 4. A Credibility Chain of Reasoning with Likelihoods at Each Link

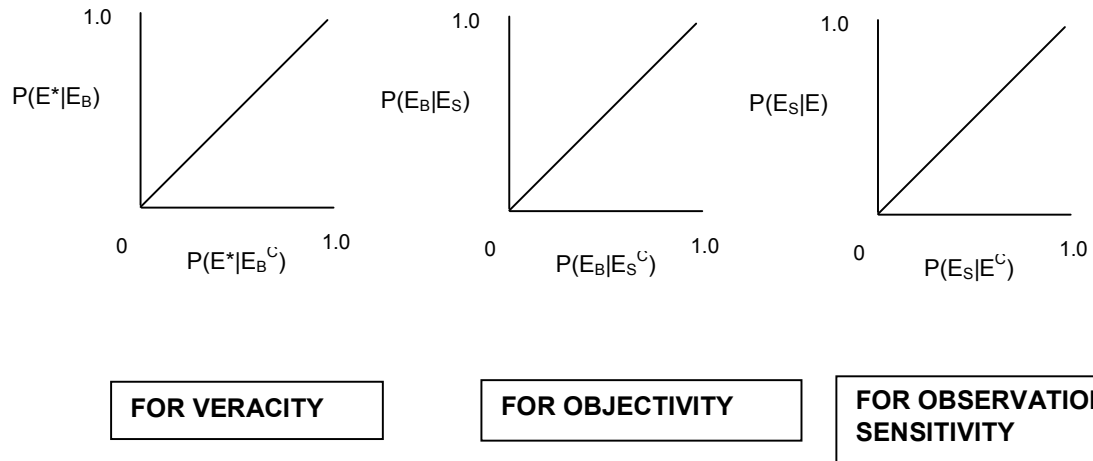


Figure 5. Likelihood Judgment Spaces for Each Credibility Attribute

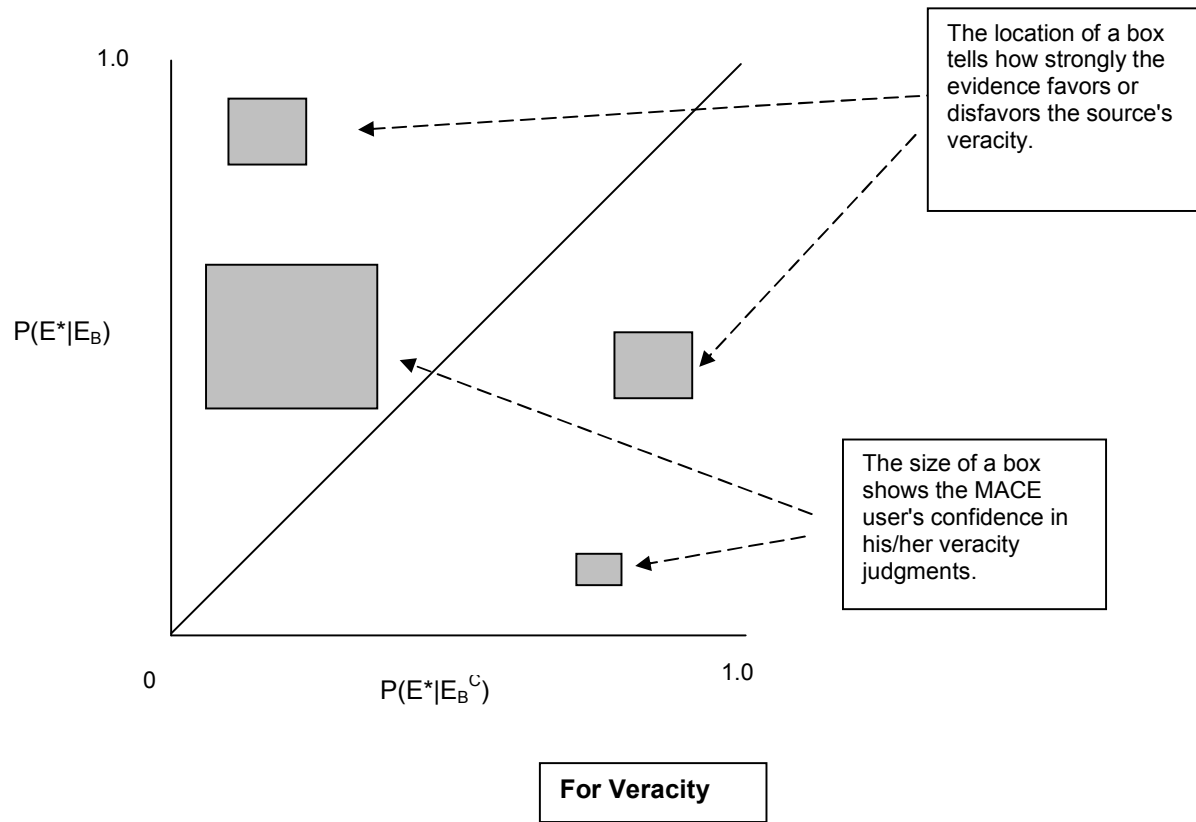


Figure 6. An Example of a MACE User's Likelihood Judgments

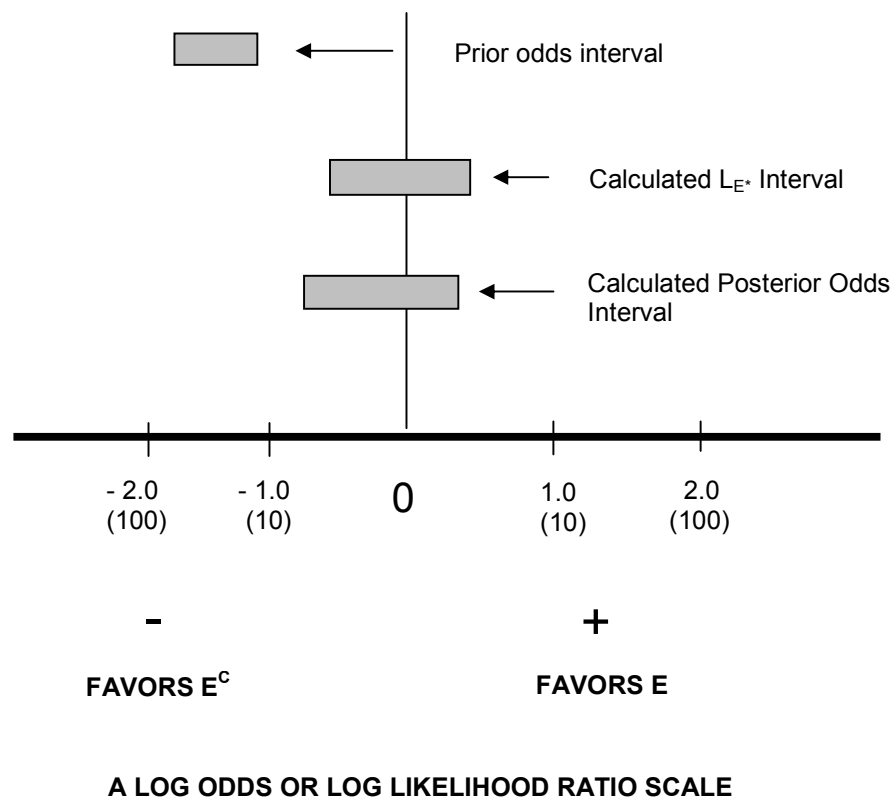


Figure 7. Bayesian Credibility Analysis Results Provided by MACE

