Substance-Blind Classification of Evidence for Intelligence Analysis

David Schum, Gheorghe Tecuci, Mihai Boicu and Dorin Marcu

Abstract—Intelligence analysis requires the development of arguments that link evidence to hypotheses by establishing and fusing the relevance, believability and inferential force or weight of a wide variety of items of evidence of different types. This paper presents several substance-blind classifications of evidence which are based on these inferential characteristics and facilitate the clarification of many uncertainties lurking in intelligence analysis. It also shows how the Disciple-LTA cognitive assistant uses these classifications to develop Wigmorean probabilistic inference networks for assessing the likelihood of hypotheses.

Index Terms—evidence classification, relevance, believability, inferential force, Wigmorean networks, cognitive assistant, ontology, evidence-based hypothesis analysis, high-level fusion

I. Why Is a Substance-blind Classification of Evidence Needed?

'Evidence' is word of relation used in the context of argumentation: e.g. "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, some hypothesis about a contested matter. One draws inferences from evidence in order to prove or disprove a hypothesis. The framework is argument, the process is proof, and the engine is inferential reasoning from information [1]. Thus *evidence* differs from the words *data* or *items of information*, since data or items of information only become evidence when their relevance is established regarding some hypothesis at issue. The term *evidence* must also be distinguished from the term *fact*. We may all agree that it is a fact that we have evidence about event E. But whether it is a fact that event E did occur is another matter since we have questions about the credibility of the source of this evidence. This makes it necessary to distinguish between evidence for an event and the event itself. Evidence can be any species of proof consisting of tangible items such as objects, documents, images, and records of any kind, or testimony from human sources or witnesses.

Evidence may have any possible substance or content. Therefore, attempts to categorize it in terms of its substance or content would be an endless and fruitless task. Why should anyone wish to be able to categorize evidence? First, *it is often necessary to compare the force or weight of different lines of argument based on different evidence in a particular analysis.* For example, here is a line of argument based on HUMINT evidence; how does this argument compare with a different line of argument based on IMINT or one based on MASINT? Second, *there are different uncertainty issues that arise when we have different kinds of evidence.* Third, *how does the strength of our conclusions in one analysis compare with those reached in another analysis, given the fact that these two analyses are based on entirely different mixtures of evidence?* Fourth, *how will we ever resolve differences among analysts themselves, or among analysts and their "customers", regarding interpretations of evidence forming the basis for conclusions reached in an analysis?* Finally, *how do we ever say anything general about evidence given that its substance or content varies in a near infinite fashion? What is badly needed in so many situations is an evidence categorization scheme for allowing us to say what kinds of evidence we have without resorting to discussions about its substance or content.*

In this paper we present a foundation for such an evidence categorization scheme that will tell us what kinds and combinations of evidence we have in any intelligence analysis regardless of the substance or content of the evidence and the objectives of the analysis. First, we present a general approach to evidence-based hypothesis analysis which consists in developing a Wigmorean probabilistic inference network that shows how evidence is linked to a hypothesis through a potentially very complex argument that establishes and fuses the *relevance*, the *believability* and the *inferential force or weight* of a wide variety of items of evidence of different types [2, 3]. Then we present three substance-blind classifications of evidence, one based on believability, one on relevance and one on inferential force, which support the development of Wigmorean networks for hypotheses analysis. This approach to hypothesis analysis and the substance-blind classifications of evidence are implemented in Disciple-LTA, an analyst's cognitive assistant that can learn complex analytic expertise directly from expert analysts, can support analysts in hypothesis analysis, collaboration and sharing of intelligence, and can teach its analytic expertise to new analysts [4, 5, 6].

II. Wigmorean Networks

Disciple-LTA assists an analyst in assessing the likelihood of various hypotheses, such as "Al Qaeda has nuclear wea-

This work was supported in part by several U.S. Government organizations, including the Air Force Office of Scientific Research (AFOSR, FA9550-07-1-0268), the National Science Foundation (NSF, 0750461), and the Department of Defense (DOD). The US Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official polices or endorsements, either expressed or implied, of AFOSR, NSF, DOD or the U.S. Government.

Dr. David Schum is Professor in the Volgenau School of Information Technology and Engineering (VS IT&E) and in the School of Law at George Mason University (GMU), Fairfax VA 22030, USA. (phone: 703-993-1694; fax: 703-993-9275; e-mail: dschum@gmu.edu).

Dr. Gheorghe Tecuci is Professor of Computer Science in the VS IT&E and Director of the GMU Learning Agents Center (e-mail: tecuci@gmu.edu).

Dr. Mihai Boicu is Assistant Professor of Applied IT and Associate Director of the GMU Learning Agents Center (e-mail: mboicu@gmu.edu).

Dr. Dorin Marcu is Research Assistant Professor in the GMU Learning Agents Center (dmarcu@gmu.edu).

pons" or "The United States will be the world leader in non-conventional energy sources within the next 10 year" or "Iran is pursuing nuclear power for peaceful purposes" [4]. This is accomplished by developing an argument in the form of an Wigmorean inference networks, through the use of a general problem-reduction/solution-synthesis reasoning approach which is illustrated in Fig. 1 and discussed in the following.

A complex hypothesis is first reduced to simpler and simpler hypotheses and the simplest hypotheses are assessed through evidence analysis. For example, in Fig. 1, the hypothesis H_1 (or problem [P1]) is reduced to three simpler hypotheses, H_{11}, H_{12}, and H_{13} (problems [P2], [P3] and [P4]). Each of these hypotheses is assessed by considering both *favoring evidence* and *disfavoring evidence* (i.e., problems [P5] and [P6]). Let us assume that there are two items of favoring evidence for H_{11}: E_1 and E_2. For each of them (e.g., E_1) Disciple-LTA assesses the extent to which it favors the hypothesis H_{11} (i.e., [P7]). This requires assessing both the *relevance* of E_1 to H_{11} (problem [P9]) and the *believability* of E_1 (problem [P10]). Let us assume that Disciple-LTA has obtained the following solutions for these two last problems:

If we believe E_1 then H_{11} is almost certain.

It is likely that E_1 is true.

In this example, "almost certain" and "likely" are symbolic probabilities for likelihood, based on the DNI's standard estimative language. By compositing the solutions [S9] and [S10] (e.g., through a "min" function) Disciple-LTA assesses the *inferential force or weight* of E_1 on H_{11}:

Based on E_1 it is likely that H_{11} is true. [S7]

Similarly Disciple-LTA assesses the inferential force or weight of E_2 on H_{11}:

Based on E_2 it is almost certain that H_{11} is true. [S8]

By composing the solutions [S7] and [S8] (e.g., through a "max" function) Disciple-LTA assesses the inferential force/weight of the favoring evidence (i.e., E_1 and E_2) on H_{11}:

Based on the favoring evidence it is almost certain that H_{11} is true.

Through a similar process Disciple-LTA assesses the disfavoring evidence for H_{11}:

Based on the disfavoring evidence it is unlikely that H_{11} is false.

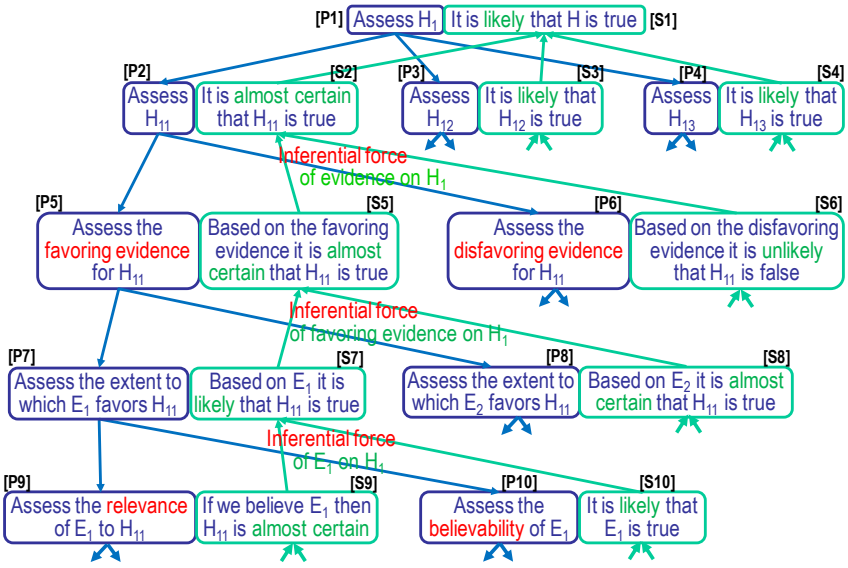

Fig. 1. Wigmorean inference network for hypothesis assessment generated by Disciple-LTA.

Because there is very strong evidence favoring H_{11} and there is weak evidence disfavoring H_{11}, Disciple-LTA concludes:

It is almost certain that H_{11} is true.

The sub-hypotheses H_{12} and H_{13} are assessed in a similar way:

It is likely that H_{12} is true. It is likely that H_{13} is true.

The solutions of H_{11}, H_{12} and H_{13} are composed (e.g., through "average") into the evidence-based assessment of H_1:

It is likely that H_1 is true.

A concrete example of such a Wigmorean network generated by Disciple-LTA is shown in Fig. 4.

III. Classification of Evidence Based on Believability

In the previous section we have discussed the process of evidence-based hypothesis assessment down to the level where one has to assess the relevance and the believability of an item of evidence. In this section we discuss how Disciple-LTA and its user assess the believability of an item of evidence by using a substance-blind classification of evidence.

Here is an important question we are asked to answer regarding the individual kinds of evidence we have: *How do you, the analyst, stand in relation to this item of evidence?* Can you examine it for yourself to see what events it might reveal? If you can, we say that the evidence is *tangible* in nature. But suppose instead you must rely upon other persons, assets, or informants, to tell you about events of interest. Their reports to you about these events are examples of *testimonial evidence*. Fig. 2 shows a substance-blind classification of evidence based on its believability credentials.

A. Tangible Evidence

There is an assortment of tangible items we might encounter and that could be examined by an intelligence analyst. Both IMINT and SIGINT provide various kinds of sensor records and images that can be examined. MASINT and TECHINT provide various objects such as soil samples and weapons that can be examined. COMINT can provide audio recordings of communications that can be overheard and translated if the communication has occurred in a foreign language. Documents, tabled measurements, charts, maps and diagrams or plans of various kinds are also tangible evidence.

There are two different kinds of tangible evidence: *real tangible evidence* and *demonstrative tangible evidence* Real tangible evidence is a thing itself and has only one major believability attribute: *authenticity. Is this object what it is represented as being or is claimed to be?* There are as many ways of generating deceptive and inauthentic evidence as there are persons wishing to generate it. Documents or written communications may be faked, captured weapons may have been altered, and photographs may have been altered in various ways. One problem is that it usually requires considerable expertise to detect inauthentic evidence.

Demonstrative tangible evidence does not concern things themselves but only representations or illustrations of these things. Ex-

amples include diagrams, maps, scale models, statistical or other tabled measurements, and sensor images or records of various sorts such as IMINT, SIGINT, and COMINT. Demonstrative tangible evidence has three believability attributes. The first concerns its *authenticity*. For example, suppose we obtain a hand drawn map from a captured insurgent showing the locations of various groups in his insurgency organization. Has this map been deliberately contrived to mislead our military forces or is it a genuine representation of the location of these insurgency groups?

The second believability attribute is *accuracy* of the representation provided by the demonstrative tangible item. The *accuracy question* concerns the extent to which the device that produced the representation of the real tangible item had a degree of sensitivity (resolving power or accuracy) that allows us to tell what events were observed. We would be as concerned about the accuracy of the hand-drawn map allegedly showing insurgent groups locations as we would about the accuracy of a sensor in detecting traces of some physical occurrence. Different sensors have different resolving power that also depends on various settings of their physical parameters (e.g., the settings of a camera).

The third major attribute, *reliability*, is especially relevant to various forms of sensors that provide us with many forms of demonstrative tangible evidence. *A system, sensor, or test of any kind is reliable to the extent that the results it provides are repeatable or consistent.* You say that a sensing device is reliable if it would provide the same image or report on successive occasions on which this device is used.

B. Testimonial Evidence

For *testimonial evidence* we have two basic sources of uncertainty: *competence* and *credibility*. This is one reason why it is more appropriate to talk about the *believability* of testimonial evidence which is a broader concept that includes both competence and credibility considerations. The first question to ask related to competence is whether this source actually made the observation he claims to have made or had access to the information he reports. The second competence question concerns whether this source understood what was being observed well enough to provide us with an intelligible account of what was observed. Thus competence involves *access* and *understandability*.

Assessments of human source credibility require consideration of entirely different attributes: *veracity* (or *truthfulness*), *objectivity*, and *observational sensitivity under the conditions of observation*. Here is an account of why these are the major attributes of testimonial credibility. First, is this source telling us about an event he/she believes to have occurred? This source would be untruthful if he/she did not believe the reported event actually occurred. So, this question involves the source's *veracity*. The second question involves the source's *objectivity*. The question is: Did this source base a belief on sensory evidence received during an observation, or did this source believe the reported event occurred either because this source expected or wished it to occur? An objective observer is one who bases a belief on the basis of sensory evidence instead of desires or expectations. Finally, if the source did base a belief on sensory evidence, how good was this evidence? This involves information about the source's relevant *sensory capabilities and the conditions under which a relevant observation was made*.

As indicated in Fig. 2, there are several types of testimonial evidence. If the source does not hedge or equivocate about what he/she observed (i.e., the source reports that he/she is certain that the event did occur), then we have *unequivocal testimonial evidence.* If, however, the source hedges or equivocate in any way (e.g., "I'm fairly sure that E occurred") then we have *equivocal testimonial evidence*. The first question we would ask this source of *unequivocal testimonial evidence* is: *How did you obtain information about what you have just reported?* It seems that this source has three possible answers to this question. The first answer is: "I made a *direct observation* myself. In this case we have *unequivocal testimonial evidence based upon direct observation.* The second possible answer is: "I did not observe this event myself but heard about its occurrence (or nonoccurrence) from another person". Here we have a case of secondhand or hearsay evidence, called *unequivocal testimonial evidence obtained at second hand*. A third answer is possible: "I did not observe event E myself nor did I hear about it from another source. But I did observe events C and D and inferred from them that event E definitely occurred". This is called *testimonial evidence based on opinion* and it requires some very difficult questions. The first concerns the source's credibility as far as his/her observation of event C and D; the second involves our examination of whether we ourselves would infer E based on events C and D. This matter involves our assessment of the source's *reasoning ability*. It might well be the case that we do not question this source's credibility in observing events C and D, but we question the conclusion that event E occurred the source has drawn from his observations. We would also question the certainty with which the source has reported an opinion that E occurred. Despite the source's conclusion that "event E definitely occurred", we should consider that *testimonial evidence based on opinion* is a type of *equivocal testimonial evidence*.

There are two other types of equivocal testimonial evidence. The first we call *completely equivocal testimonial evidence*. Asked whether event E occurred or did not, our source says: "I don't know", or "I can't remember".

But there is another way a source of HUMINT can equivocate; the source can provide *probabilistically equivocal testimonial evidence* in various ways: "I'm 60 percent sure that event E happened"; or "I'm fairly sure that E occurred"; or "It is very unlikely that E occurred". We could look upon this

Fig. 2. Evidence classification based on believability.

particular probabilistic equivocation as an assessment by the source of his own observational sensitivity.

C. Missing Evidence

To say that evidence is missing entails that we must have had some basis for expecting we could obtain it. There are some important sources of uncertainty as far as missing evidence is concerned. In certain situations missing evidence can itself be evidence. Consider some form of tangible evidence, such as a document, that we have been unable to obtain. There are several reasons for our inability to find it, some of which are more important than others. First, it is possible that this tangible item never existed in the first place; our expectation that it existed was wrong. Second, the tangible item exists but we have simply been looking in the wrong places for it. Third, the tangible item existed at one time but has been destroyed or misplaced. Fourth, the tangible item exists but someone is keeping it from us. This fourth consideration has some very important inferential implications including denial and possibly deception. An adverse inference can be drawn from someone's failure to produce evidence.

D. Accepted Facts

There is one final category of evidence about which we would never be obliged to assess its believability. Tabled information of various sorts such as tide table, celestial tables, tables of physical or mathematical results such as probabilities associated with statistical calculations, and many other tables of information we would accept as being believable provided that we used these tables correctly. For example, an analyst would not be obliged to prove that temperatures in Iraq can be around 120 degrees Fahrenheit in summer months, or that the population of Baghdad is greater than that of Basra.

E. Mixed Evidence

We have just considered a categorization of individual items of evidence but there are situations in which individual items can reveal various mixtures of these types of evidence. One example involves a tangible document containing a testimonial assertion based on other alleged tangible evidence. Thus these forms of evidence are not mutually exclusive; they can occur together in a single item of evidence.

F. Believability Assessment with Disciple-LTA

Disciple-LTA knows about the types of evidence shown in Fig. 2 and how their believability should be evaluated. For example, Fig. 3 shows the reasoning tree automatically generated by Disciple-LTA for solving the problem: "Assess the extent to which one can believe Osama bin Laden as the source of EVD-Dawn-Mir01-01c." Notice that, in accordance with the above discussion, Disciple-LTA reduces the believability of this testimony of Osama bin Laden to two simpler problems, one for assessing the competence of Osama bin Laden, and the other for assessing his credibility. This second problem is further reduced to assessing bin Laden's veracity, objectivity and observational sensitivity.

Disciple-LTA may have knowledge about these believability characteristics of Osama bin Laden (e.g., that his veracity is an even chance). Alternatively, the analyst may make assumptions with respect to the values of these characteristics. In any case, once the solutions of the simplest problems are obtained, they are combined, from bottom up, to assess the believability of Osama bin Laden. For example, the probabilistic estimates of bin Laden's veracity, objectivity and observational sensitivity (i.e., an even chance, almost certain, and almost certain, respectively) are combined (through a min function) to obtain a probabilistic estimate of his credibility (i.e., an even chance). Then, bin Laden's credibility is automatically combined with his competence (again through a min function), to estimate bin Laden's believability as the source of EVD-Dawn-Mir01-01c.

Disciple-LTA also allows the analysts to assess these believability characteristics by developing Wigmorean networks, as illustrated in Fig. 4 where Disciple-LTA reduces the problem of assessing the veracity of bin Laden to simpler problems, and then assesses the simplest problems based on the available evidence [7]. As one can see, the Wigmorean network in Fig. 4 has the general structure shown in Fig. 1.

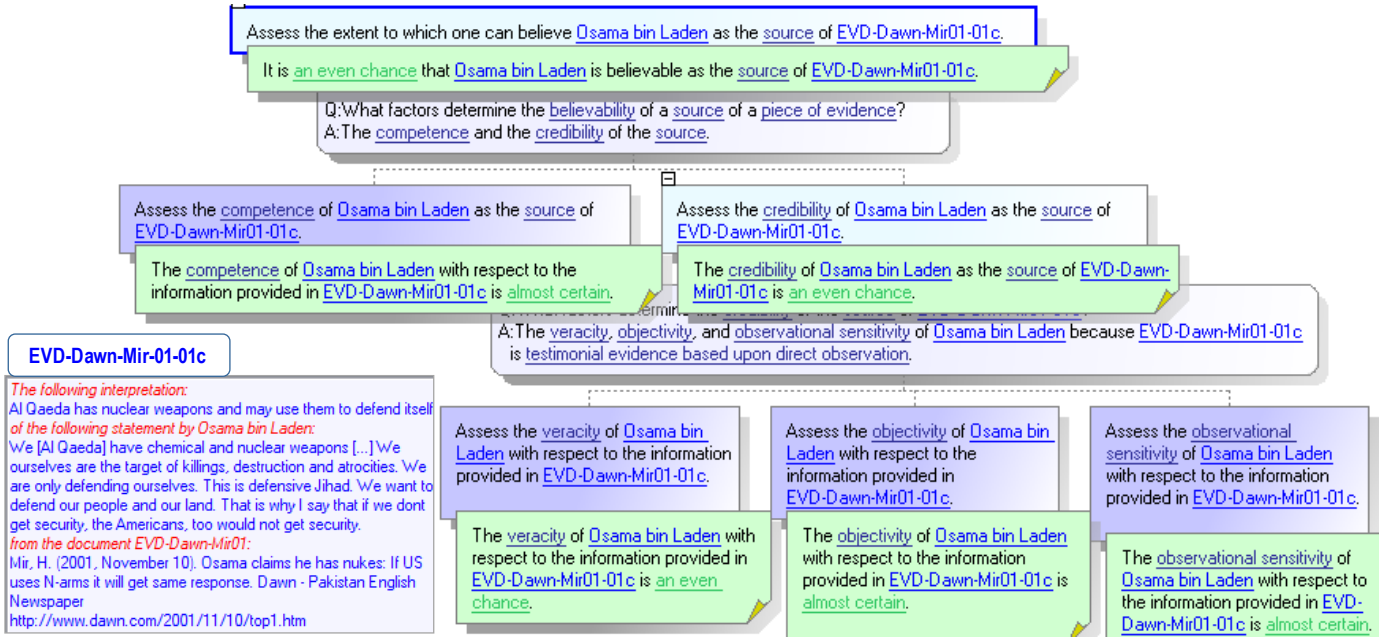

Fig. 3. Source believability assessment.

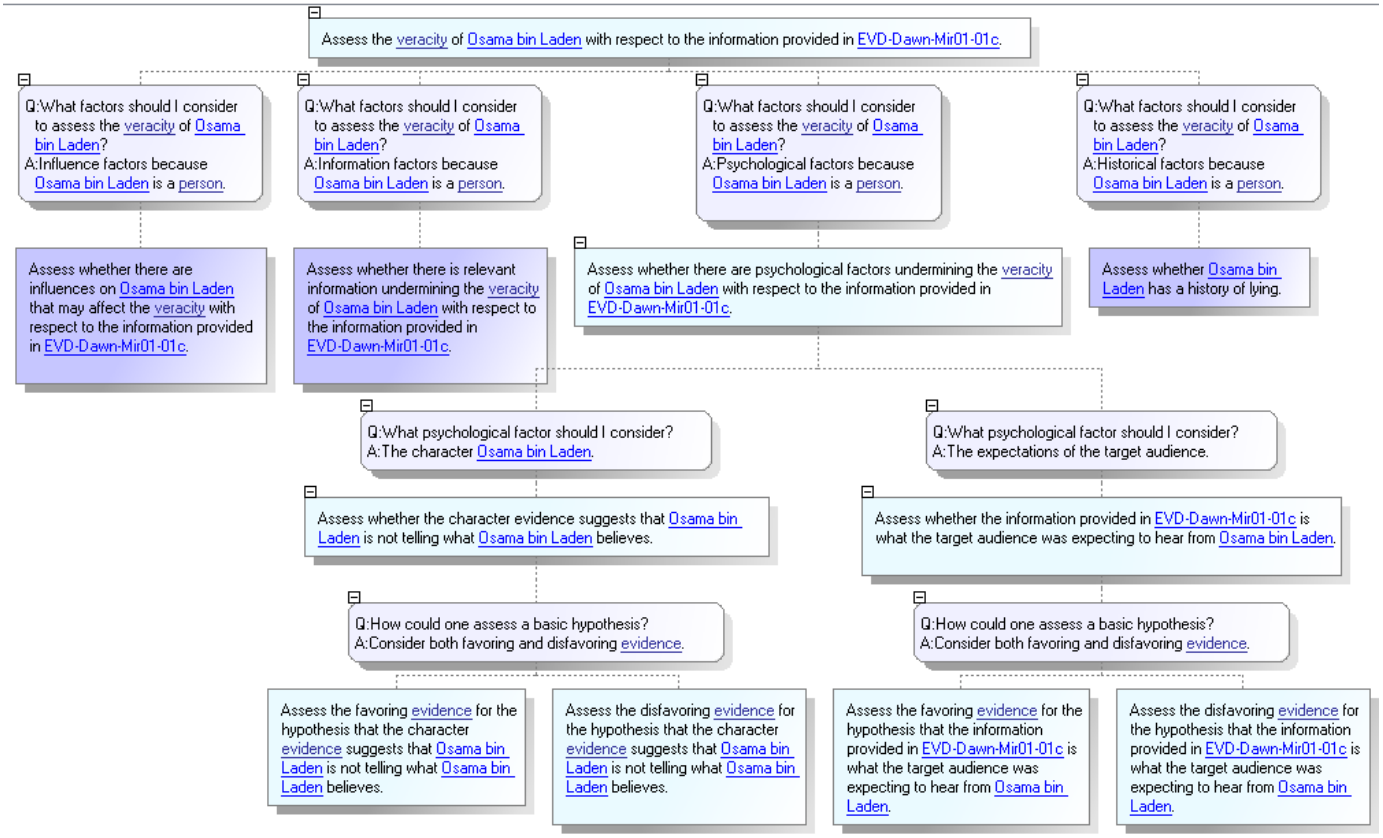

Fig. 4. Wigmorean network for veracity assessment.

IV. Classification of Evidence Based on Relevance

Here is an important *relevance question* we are asked to answer regarding the individual kinds of evidence we have: *How does this item of evidence stand in relation to what you, the analyst, are trying to prove or disprove from it?*

There are two species of relevant evidence. Some evidence may be *directly relevant* if you can form a defensible chain of reasoning from this item of evidence to hypotheses you are considering. For example, E_1 and E_2 in Fig. 1 are directly relevant items of evidence.

Other evidence may be *indirectly relevant*, or *ancillary evidence* if it bears upon the strength or weakness in chains of reasoning set up by directly relevant evidence. Consider, for example, the problem "Assess the believability of E_1" from the bottom right side of Fig. 1. Any item of evidence that might be used in solving this problem would be indirectly relevant evidence. Indirectly relevant evidence is also any evidence used in solving the problem "Assess the extent to which one can believe Osama bin Laden as the source of EVD-Dawn-Mir01-01c" from Fig. 3, such as any evidence from the reasoning tree in Fig. 4. The term *meta-evidence* is also appropriate since ancillary evidence is evidence about other evidence. Fig. 5 shows this relevance-based classification of evidence.

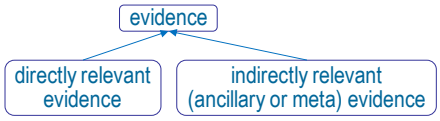

Fig. 5. Evidence classification based on relevance.

V. Classification Of Evidence Based on Inferential Force or Weight

Here is an *inferential force or weight question* we are asked to answer regarding the individual kinds of evidence we have: *How does this item of evidence changes your belief in the truthfulness in what you are trying to assess?* If the item of evidence increases our belief in the truthfulness of the hypothesis we are analyzing, we call it *favoring evidence*. Otherwise, we call it *disfavoring evidence*. For example, both E_1 and E_2 in Fig.1 are examples of favoring evidence with respect to the hypothesis H_{11}.

As shown in Fig. 1, one also has to assess the inferential force of a combination of two or more individual items of evidence. These combinations of evidence are also recurrent and do not involve the substance or content of the evidence. One reason for carefully considering these combinations of evidence is that they are often confused or incorrectly identified leading to mistakes in how the evidence is described in an analysis. But perhaps the most important reason is that *there are very important sources of uncertainty lurking in these evidential combinations.* As shown in Fig. 6, there are three main classes of evidence combinations.

A. Harmonious Evidence

Two or more items of evidence are harmonious if they are *directionally consistent* in the sense that they all favor the same hypothesis. There are two basic forms of harmonious evidence, *corroborative evidence* and *convergent evidence.* In the case of *corroborative evidence* we have two or more

sources telling us that the same event E has occurred. This form of corroboration often allows us to have greater confidence that the event in question did occur. In such cases we would say that one source has verified what the other source has told us. The exception involves instances where we have other evidence suggesting that two or more HUMINT sources collaborated in deciding what to tell us, or that one source influenced or coerced another source to report the same event.

In the case of *convergent evidence* we have two or more evidence items that concern *different events all of which point toward or favor the same hypothesis*. *Convergent evidence can exhibit evidential synergism. In many situations two or more evidence items, considered jointly, have greater inferential force or weight than they would have if we considered them separately or independently.* Another way to characterize evidential synergism is to say that one item of evidence can have greater force if we consider it in light of other evidence.

B. Dissonant Evidence

Dissonant evidence involves combinations of two or more items that are directionally inconsistent; they can point us in different inferential directions or toward different hypotheses. There are two basic forms of evidential dissonance; the first involves *contradictory evidence*. Contradictory evidence always involves events that are mutually *exclusive*, they cannot have occurred jointly. From one source we learn that event E occurred; but from another source we learn that this same event did not occur. The dissonance seems obvious in this case since event E cannot have occurred and not have occurred at the same time. Evidential contradictions are always resolved on credibility grounds. As an example, suppose we have three HUMINT sources who tell us that event E occurred, and one HUMINT source who tells us that event E did not occur. In the not so distant past, it was believed that we should always resolve the contradiction by counting heads; i.e. majority rules. So, on this basis we would side with the three sources who tell us that event E did occur. The trouble here is that counting heads assumes that all of the four sources involved in this episode of contradictory evidence have equal credibility. This may be a very bad assumption since, on ancillary evidence about these four sources, we may well believe that the one source telling us that E did not occur has greater credibility than does the aggregate credibility of the three sources who tell us that event E did occur. So, what matters in resolving evidential contradictions is the aggregate credibility of the sources on either side of this contradiction.

There is another form of dissonant evidence called *divergent evidence*. This pattern of dissonance differs from contradictory evidence in the following way. A contradiction always involves whether one event occurred or did not occur. But divergent evidence involves entirely different events; the directional inconsistency here means that these events point us toward different hypotheses. In one case, suppose credible evidence about event E would favor hypothesis H, but credible evidence about event F would favor hypothesis not-H.

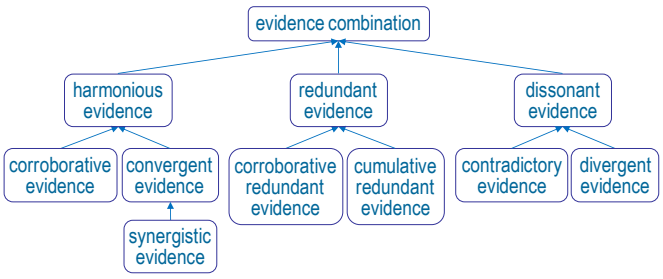

Fig. 6. Recurrent substance-blind combinations of evidence.

C. Evidential Redundance

We often encounter two or more items of evidence in which the first item acts to reduce the force of subsequent items of evidence. Stated another way, the first item acts to make subsequent items *redundant* to some degree. There are two ways this can happen. The first form of evidential redundance involves the corroborative evidence we discussed above. In this case we have repeated evidence of the same events. Although having corroborative evidence does add to our confidence that an event of interest did occur, each additional item adds less and less to our confidence. We refer to this situation as *corroborative redundance*.

The second form of redundancy involves different events in which evidence about one event, if credible, takes something off the inferential force of evidence about another event. We have called this *cumulative redundance*. The word "cumulative" is an expression used in law to refer to evidence that does not add anything to what we already know.

It is very important to consider these two forms of evidential redundancy. In the case of corroborative redundance we risk *double counting* evidence about the same event and ascribing additional weight the evidence does not always have. For cumulative redundance we risk getting more inferential mileage out of the evidence than can be justified.

VI. Conclusions

We have discussed several substance-blind forms and combinations of evidence, each raising uncertainty issues that cannot be ignored in any intelligence analysis. Disciple-LTA knows about several of them and takes them into account for evidence-based hypothesis assessment, but much more work remains to be done, especially concerning the inferential force.

References

[1] Twining W., Evidence as a Multi-Disciplinary Subject. *Law, Probability & Risk: A Journal of Reasoning Under Uncertainty*. Vol. 2, No. 2, June 2003.

[2] Wigmore J.H., *The Science of Judicial Proof,* Boston: Little, Brown, 1937.

[3] Schum D.A., The Evidential Foundations of Probabilistic Reasoning, Northwestern University Press, 2001.

[4] Tecuci G., Boicu M., Marcu D., Boicu C., Barbulescu M., Disciple-LTA: Learning, Tutoring and Analytic Assistance, Journal of Intelligence Community Research and Development, 2008.

[5] Schum D.A., Tecuci G., Boicu M., "Analyzing Evidence and its Chain of Custody: A Mixed-Initiative Computational Approach," *International Journal of Intelligence and Counterintelligence,* 22:2, 2009, pp.298-319.

[6] Boicu M., Tecuci G., Schum D., Intelligence Analysis Ontology for Cognitive Assistants, in *Proc OIC-08,* George Mason University.

[7] Schum D.A. and Morris J., Assessing the Competence and Credibility of Human Sources of Intelligence Evidence: Contributions from Law and Probability, *Law, Probability, & Risk,* Vol. 6, pp. 247–274, 2007.