Research Report #7, Learning Agents Center, George Mason University,
August 2009, updated October 2009, March 2010.

GHEORGHE TECUCI, MIHAI BOICU, DAVID SCHUM and DORIN MARCU [1]

Coping with the Complexity of Intelligence Analysis: Cognitive Assistants for Evidence-Based Reasoning

ABSTRACT: This paper presents a computational approach to intelligence analysis which is viewed as ceaseless discovery in a non-stationary world involving concurrent processes of evidence in search of hypotheses, hypothesis in search of evidence, and evidential tests of hypotheses. This approach is at the basis of Disciple-LTA, a cognitive assistant that helps intelligence analysts evaluate the likelihood of hypotheses by developing Wigmorean probabilistic inference networks that link evidence to hypotheses in argumentation structures that establish the relevance, believability and inferential force or weight of evidence. The paper also shows how the intelligence analysis concepts and methods embedded into Disciple-LTA, which are based on the Science of Evidence and Artificial Intelligence, can be used to improve other structured analytic methods, using Analysis of Competing Hypothesis as an example.

KEY WORDS: Science of Evidence, Artificial Intelligence, Discovery, cognitive assistant Wigmorean networks, abduction, deduction, induction, substance-blind classification of evidence, relevance, believability, inferential force or weight, assumption, abstraction, analysis of competing hypotheses

---

[1] ***Dr. Gheorghe Tecuci*** *is Professor of Computer Science in the Volgenau School of Information Technology and Engineering and Director of the Learning Agents Center at George Mason University (GMU), and Visiting Professor and former Chair of Artificial Intelligence at the US Army War College.*

Dr. Mihai Boicu *is Assistant Professor of Applied Information Technology and Associate Director of the Learning Agents Center in the Volgenau School of Information Technology and Engineering, George Mason University.*

Dr. David Schum *is Professor in the Volgenau School of Information Technology and Engineering and in the School of Law, at George Mason University. He is also Honorary Professor of Evidence Science, University College London.*

Dr. Dorin Marcu *is Research Assistant Professor in the GMU Learning Agents Center.*

1. INTRODUCTION

Intelligence analysts and others face the difficult task of drawing defensible and persuasive conclusions from masses of information of all kinds that come from a variety of different sources. Many books and papers have been written on the obvious complexity of such tasks.[1] The mass of evidence upon which conclusions eventually rest has five major characteristics that make conclusions drawn from evidence necessarily probabilistic in nature. Our evidence is always *incomplete* no matter how much we have and is commonly *inconclusive* in the sense that it is consistent with the truth of more than one hypothesis or possible explanation. Further, the evidence is frequently *ambiguous*; we cannot always determine exactly what the evidence is telling us. A mass of evidence is in most situations *dissonant* to some degree; some of it favors one hypothesis or possible explanation but other evidence favors other hypotheses. Finally, all of our intelligence evidence comes from sources having any possible gradation of *believability or credibility* shy of perfection. Arguments, often stunningly complex, are necessary in order to establish and defend the three major credentials of evidence: its *relevance*, *believability or credibility*, and *inferential force or weight*. These arguments rest upon both *imaginative* and *critical reasoning* on the part of intelligence analysts.

But these assorted evidential characteristics are not the only elements of the complexity of intelligence analysis tasks. A major objective of intelligence analysis is to help insure that the policies and decisions reached by the governmental and military leaders, at all levels, are well informed. The policy-relevance of analytic "products" is a goal routinely kept in mind. Analysts face different requirements in their efforts to serve these policy and decision-making "customers". In some cases current analyses are required to answer questions that are of immediate interest and that do not allow analysts time for extensive research and deliberation on available evidence regarding the questions being asked. In other cases, teams of analysts participate in more lengthy analyses that combine evidence from every available source to make long-term assessments on matters of current and abiding interest.

Identifying the complexities of intelligence analysis is actually the easy part. What is not so easy are efforts to assist analysts in coping with the complexities of the evidential reasoning tasks they routinely face.

This paper presents a systematic approach to hypotheses analysis which is based on the solid theoretical foundations of the emerging Science of Evidence[2] and uses Artificial

Intelligence[3] methods to automate significant portions of the hypotheses analysis process, helping intelligence analysts to overcome many analytic complexities. This approach is implemented in an intelligent cognitive assistant called Disciple-LTA[4]. Disciple-LTA is a new type of analytic tool that integrates three complex capabilities. It can rapidly *learn*, directly from an expert analyst, the analytic expertise which currently takes years to establish, is lost when analysts separate from service, and is costly to replace. It can *tutor* new intelligence analysts how to systematically analyze complex hypotheses. Finally, it can *assist* the analysts to analyze complex hypotheses, collaborate, and share information.

In the next section we discuss the critical role of discovery in intelligence analysis. Then, in Section 3, we present a view of intelligence analysis as a process of ceaseless discovery in a non-stationary world, process involving evidence in search of hypotheses (through abductive reasoning), hypotheses in search of evidence (through deductive reasoning), and evidential tests of hypotheses (through inductive reasoning), all going on at the same time. In Section 4 we show how Disciple-LTA performs this process by employing a general divide and conquer reasoning strategy called problem reduction and solution synthesis. Section 5 defines the major credentials of evidence (relevance, believability and inferential force or weight) and how they are represented in Disciple-LTA. After that, Section 6 presents the structure of the Wigmorean probabilistic inference networks[5] generated by Disciple-LTA to assess the likelihood of hypotheses. Section 7 presents how an analyst can use assumptions in an analysis developed with Disciple-LTA, to deal with lack of evidence or analysis time, and to investigate what-if scenarios. Section 8 presents a substance-blind classification of evidence and how it is used in assessing the believability or credibility of evidence. Because the analysis of complex hypotheses from masses of evidence generally result in very large reasoning trees, Section 9 presents the abstractions used by Disciple-LTA to facilitate the browsing and understanding of these trees.

An additional claim with respect to Disciple-LTA is that the intelligence analysis concepts and methods embedded into it, which are based on the Science of Evidence and Artificial Intelligence, particularly the systematic approach to the development of argumentation structures, the substance-blind classification of evidence and the associated procedure for assessing the believability of evidence, the drill-down analysis and assumptions-based reasoning, may help the analysts perform better analyses, no matter what analysis methods they use. To justify this claim, Section 10 describes what is probably the most popular structured analytic method, Richards J. Heuer's *Analysis of Competing Hypothesis* [ACH],[6] and show how it can be improved by employing the concepts and methods embedded in Disciple-LTA.

We conclude this paper with a brief discussion on how Disciple-LTA can be used to teach intelligence analysts to perform theoretically-sound evidence-based hypothesis analysis, through a hands-on, learning by doing approach. We also mention future research directions aiming at further facilitating the complex evidential reasoning tasks faced by the intelligence analysts.

2. DISCOVERY: GENERATING HYPOTHESES, EVIDENCE, AND ARGUMENTS

All intelligence analyses, in common with analytic activities in any other context, begin with the asking of questions about matters of interest. These questions can arise from the analysts themselves or from other persons, such as the policy or decision makers, who are being served by intelligence analysts. These questions can concern possible explanations for events or situations in the past or possible predictions about events or situations in the future. In many cases these questions are bound together. In order to predict possible events in the future we need accurate explanations for related events in the past. The field of intelligence analysis has many inherent difficulties, but none seem more difficult than the fact that analysts must provide their explanations or predictions in a non-stationary world. In short, the world keeps changing as analysts are trying their best to understand it well enough to provide explanations or to make predictions. One consequence is that we have continuing streams of new information, some items of which we will assess as being relevant evidence regarding our explanations or predictions. An explanation for some pattern of past events analysts have previously regarded as correct may now seem incorrect in light of new evidence just discovered today. A prediction regarded as highly likely today may be overtaken by events we will learn about tomorrow. In fact, the very questions we have asked yesterday may need to be revised or may even seem unimportant in light of what we learn today. One consequence of all of this is that the process of discovery or investigation in intelligence analysis is a ceaseless activity. It would be a drastic mistake to view discovery in intelligence analysis as being a stationary activity in a non-stationary world.

What exactly does discovery involve, or what needs to be discovered in intelligence analysis? The answer is: *hypotheses*, *evidence*, and *arguments linking hypotheses and* *evidence*. From observations we make, or questions we ask, we generate alternative hypotheses or propositions offered in explanation for past events or possible predictions about future events. In the continual streams of data or information provided to intelligence analysts only a minute fraction of these data are justified as being termed evidence. *Data or items of information only become evidence when their relevance to hypotheses being considered is established by defensible and persuasive arguments.* What is true is that establishing these three ingredients of all intelligence analysis is a very complex activity involving *imaginative* as well as *critical reasoning*. Discovery in intelligence analysis involves mixtures of all three forms of reasoning that have been identified: *abduction*, *deduction*, and *induction*. As we know, deduction shows that something is *necessarily* true, induction shows that something is *probably* true, and abduction shows that something is *possibly* true. The identification of abductive reasoning was first made by the American philosopher Charles S. Peirce, who argued that we will not generate any new ideas, in the form of hypotheses, by deductive or inductive reasoning. He identified abductive reasoning as being associated with imaginative, creative, or insightful reasoning.[7]

But now we must return to intelligence analysis being a ceaseless discovery-related activity performed in a non-stationary world. On at least some accounts it may appear that the generation of a productive hypothesis occurs as a result of a single glorious episode of abductive or imaginative reasoning on the part of a particular intelligence analyst. Barring clairvoyance or divine intervention, this seems quite unlikely. Tying discovery to just abductive reasoning overlooks the true complexity of discovery in intelligence analysis and in many other contexts. Remember that we have three things to be discovered in intelligence analysis: hypotheses, evidence, and arguments linking evidence to hypotheses. The fact that the world is changing all the time we are trying to understand it means that we have evidence in search of hypotheses, hypotheses in search of evidence, and evidential tests of hypotheses *all going on at the same time*. What this means is that discovery in intelligence analysis involves mixtures of abductive, deductive, and inductive reasoning. By means of abductive reasoning we generate hypotheses from evidence we gather; by deductive reasoning, we make use of our hypotheses to generate new lines of inquiry and evidence; and by inductive reasoning we test hypotheses on the basis of the evidence we are discovering. Such testing depends on the relevance and believability of our evidence. These factors combine in further complex ways to allow us to assess the inferential force or weight of the evidence we are considering.

There is one further matter of interest here. What is termed abductive, imaginative, or insightful reasoning is not perfectly understood. There are many accounts of this reasoning, what it entails, how it arises, and how it can be enhanced. About the only point of agreement among most persons devoted to the study of this form of reasoning is that it cannot be performed alone by a computer. Analysts may, however, be assisted in performing abductive reasoning. As we have shown elsewhere, there are many species of abductive reasoning, depending both on how imaginative a new generated idea is and what form the new idea takes.[8] This account also shows how these different species of abductive reasoning are always interspersed with deductive and inductive reasoning steps in any form of complex analysis. Although this account was given in the context of law, the same ideas apply to intelligence analysis.

3. EVIDENCE IN SEARCH OF HYPOTHESES, HYPOTHESES IN SEARCH OF EVIDENCE, AND EVIDENTIAL TESTING OF HYPOTHESES

Figure 1 represents the process of ceaseless discovery in a non-stationary world, a process viewed as evidence in search of hypotheses (through abductive reasoning), hypotheses in search of evidence (through deductive reasoning), and evidential tests of hypotheses (through inductive reasoning), all going on at the same time. To illustrate this process, let us consider an analyst, Mavis, who reads today's Washington Post and comes upon an article that concerns how safely radioactive materials are stored in this general area. The investigative reporter and author of this piece begins by noting how the storage of nuclear and radioactive materials is so frequently haphazard in other countries and wonders how carefully these materials are guarded here in the USA, particularly in this general area. In the process of his investigations the reporter notes his discovery that a canister containing cesium-137 has gone missing from the XYZ Company in MD, just three days ago. The XYZ Company manufactures devices for sterilizing medical equipment and uses cesium-137 in these devices along with other radioactive materials. This piece arouses Mavis' curiosity because of her concern about terrorists planting dirty bombs in our cities.

The bottom left-hand of Figure 1 shows an item of evidence (E^*_i) that leads Mavis to abductively leap to the hypothesis H_k shown at the top of Figure 1, that a dirty bomb will be set off in the Washington DC area. In this case we have *evidence in search of hypotheses* where Mavis may experience a flash of insight allowing her to generate the hypothesis H_k. Asked to indicate why this hypothesis explains the evidence, Mavis generates a series of propositions that can logically link the evidence and the hypothesis, as shown in Table 2. These interim propositions, in a logical sequence, are sources of doubt or uncertainty about the linkage between the evidence E^*_i and the hypothesis H_k. So, in this case, we have evidence in search of hypotheses where new items of evidence "search" for hypotheses that explain them.

The diagram in the middle of Figure 1 illustrates the deductive processes involved when we have *hypotheses in search of evidence*. Once the new hypothesis H_k has been generated, Mavis has to assess it. The reasoning might start as follows. If H_k were true, there are sub-hypotheses, listed as H_d and H_e, that would be necessary and sufficient to make H_k true. In turn, each of these sub-hypotheses allows Mavis to deduce potential items of evidence (shown as the shaded circles) that bear upon them. Notice that the path from the hypothesis H_k to the evidence E^*_i is the reverse of the abductive reasoning path from the left-hand side of Figure 1. So here we have hypotheses in search of evidence that may favor or disfavor them.

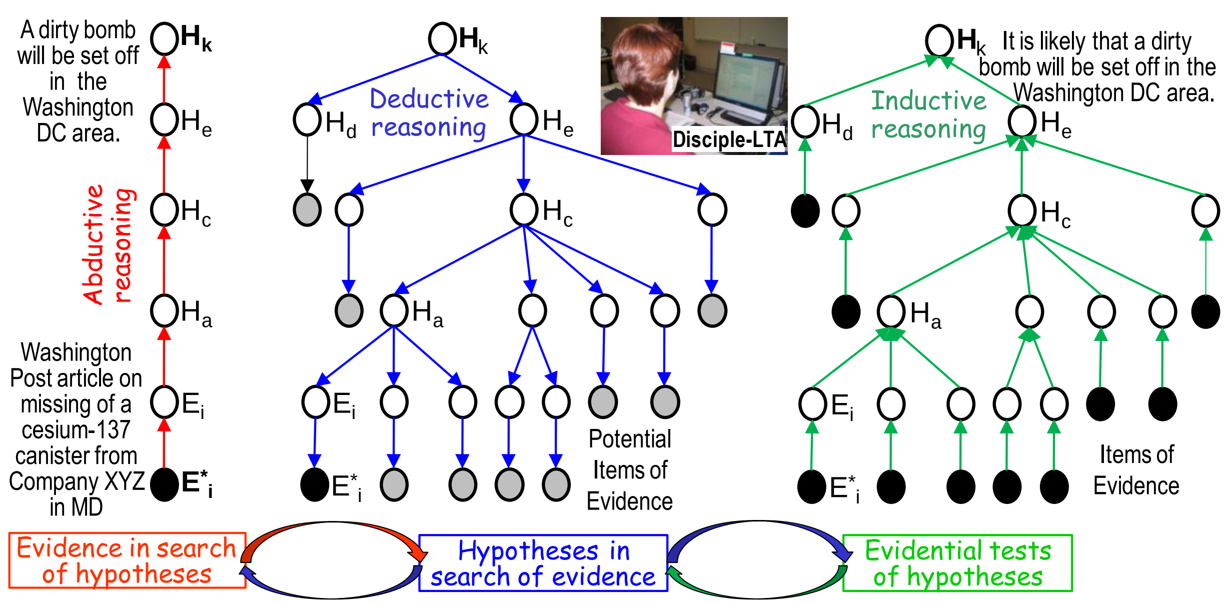

Figure 1. Evidence-Hypotheses Relations.

Table 2. An illustration of the abductive reasoning from Figure 1.

Evidence E^*_i:	Washington Post article on the missing of a cesium-137 canister from Company XYZ in MD.
E_i:	A cesium-137 canister is missing from Company XYZ in MD.
H_a:	The canister containing cesium-137 was stolen.
H_c:	The cesium-137 was stolen by someone associated with a terrorist organization.
H_e:	The cesium-137 will be used by this terrorist organization to construct a dirty bomb.
Hypothesis H_k:	A dirty bomb will be set off in the Washington DC area.

Now, some of the newly discovered items of evidence may trigger new hypotheses (or the refinement of the current hypothesis). So, as indicated at the bottom left of Figure 1, the processes of evidence in search of hypotheses and hypotheses in search of evidence take place at the same time, and in response to one another.

This combination of evidence in search of hypotheses and hypotheses in search of evidence results in a hypothesis which has to be tested, through inductive reasoning, based on the discovered items of evidence, as shown in the right-hand side of Figure 1. The result of the testing process is the likelihood of the considered hypothesis (e.g., H_k: It is likely that a dirty bomb will be set off in the Washington DC area). If the testing of the hypothesis renders it unlikely, then new hypotheses are searched for through the other two processes.

4. COMPUTATIONAL APPROACH TO INTELLIGENCE ANALYSIS: PROBLEM REDUCTION AND SOLUTION SYNTHESIS

Once a hypothesis has been formulated by the analyst (e.g., H_k in the upper left of Figure 1), Disciple-LTA can help develop the complex structures from the middle and right-hand side of Figure 1. It does this by employing a general divide-and-conquer approach to problem solving, called *problem-reduction / solution-synthesis*, which has a grounding in the problem reduction representations developed in Artificial Intelligence,[9] and in the argument construction methods provided by the noted jurist John H. Wigmore,[10] the philosopher of science Stephen Toulmin,[11] and the evidence professor David Schum.[12] This approach uses expert knowledge and ancillary evidence to successively reduce a complex problem to simpler and simpler problems, to find the solutions of the simplest problems, and to compose these solutions, from bottom-up, to obtain the solution of the initial problem. For example, Figure 2 shows the step-by-step, top-down, reduction of the problem [P1] to simpler problems, while Figure 3 shows the step-by-step, bottom-up composition of the solutions of the simpler problems into the solution of the initial problem. Notice that this and all the other examples from this paper are only illustrative examples that should not be utilized in real-world analytic products. First, Disciple-LTA reduces the initial problem [P1] to three simpler problems, [P2], [P3], and [P4], guided by a question and its answer, as also shown in Table 3. Problem [P2] is further reduced to 5 simpler problems, again guided by a question and its answer, as shown in Figure 2.

Table 3. The top problem reduction step from Figure 2.

I have to	
Assess whether Al Qaeda has nuclear weapons based on the characteristics associated with the possession of nuclear weapons.	[P1]
What are the characteristics associated with possession of nuclear weapons?	
Reasons, desire, and ability to obtain nuclear weapons.	
Therefore I have to	
Assess whether Al Qaeda has reasons to obtain nuclear weapons.	[P2]
Assess whether Al Qaeda has desire to obtain nuclear weapons.	[P3]
Assess whether Al Qaeda has the ability to obtain nuclear weapons.	[P4]

Let us now consider the 5 leaf problems from the bottom of Figures 2 and 3. Disciple-LTA uses a six point symbolic probabilities scale (no evidence, a remote possibility, unlikely, an even chance, likely, almost certain) to express the probabilistic solutions of these problems, solutions shown at the bottom of Figure 3. These symbolic probabilities correspond to the US National Intelligence Council’s Standard Estimative Language. However, the language can easily be changed to consider more or fewer symbolic probabilities and to associate specific probability intervals with each of them.[13] The probabilistic solutions from the bottom part of Figure 3 are combined, through a “max” function, to obtain the solution [S2] shown both in Figure 3 and in Table 4. The probabilistic solutions [S3] and [S4] are obtained, in a similar way, from simpler solutions. Then the solutions [S2], [S3], and [S4] are combined, through a “min” function, into the solution [S1] of the problem [P1] from the top of Figure 3. Notice that some words from Figures 2 and 3 are underlined (e.g., Al Qaeda, nuclear weapons).

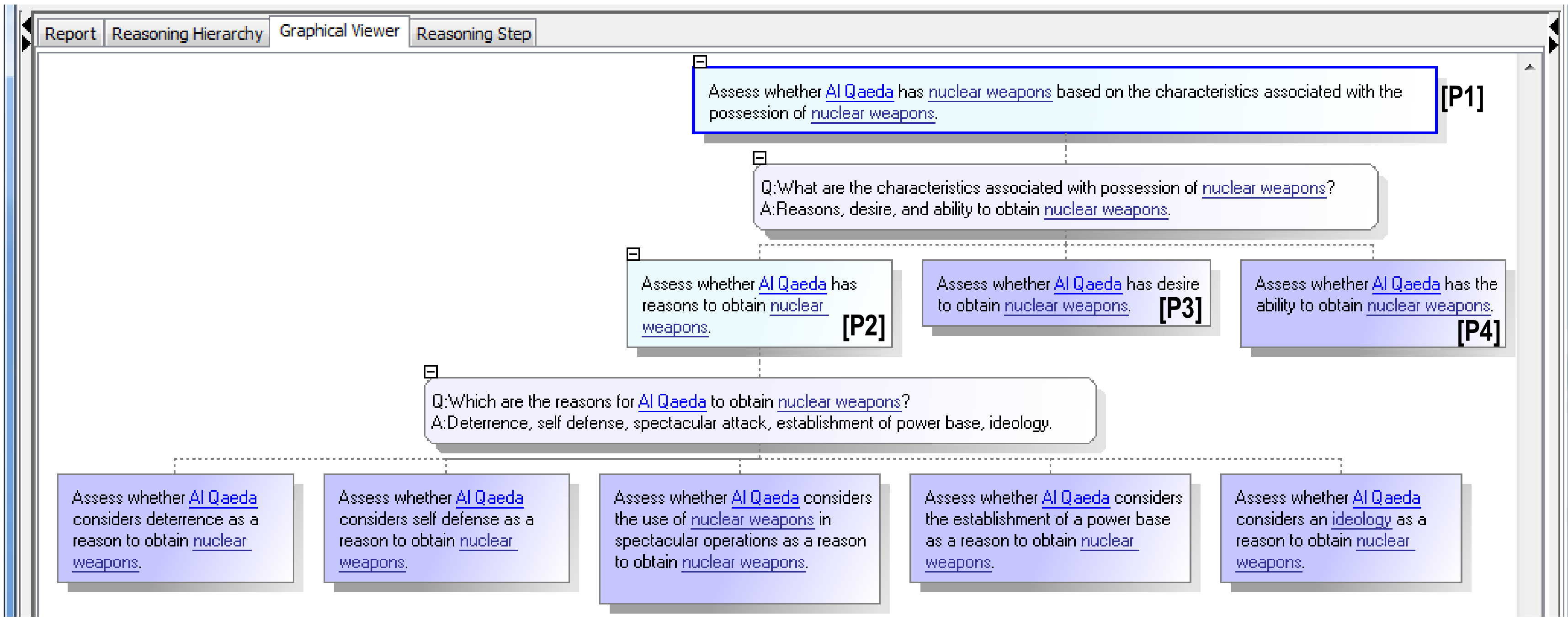

Figure 2. Reduction of a complex hypothesis analysis problem to simpler problems.

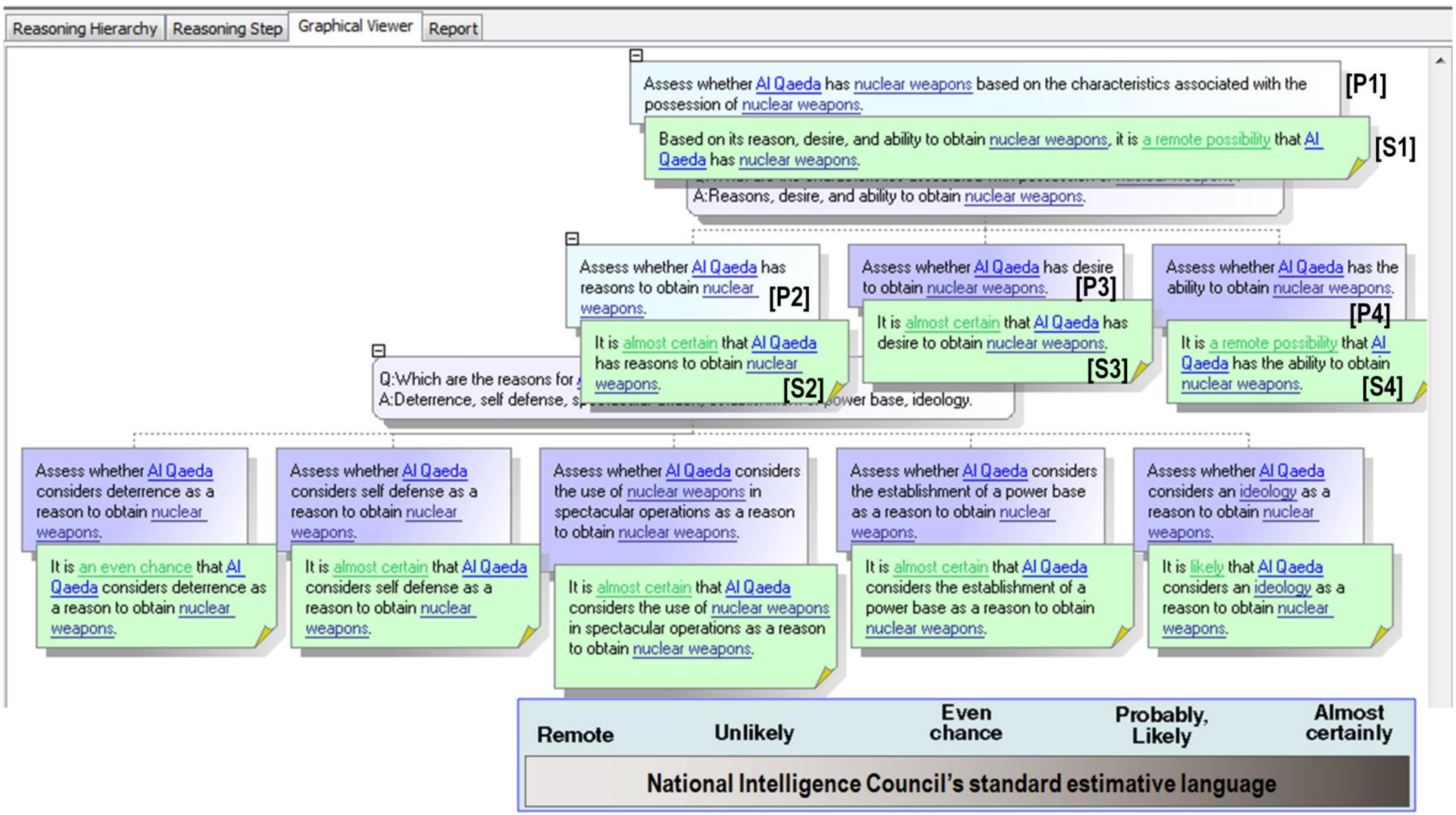

Figure 3. Hypothesis analysis through problem reduction and solution synthesis.

They correspond to entities that are represented into the knowledge base of Disciple-LTA.[14] Disciple-LTA displays their descriptions when the analyst clicks on their names.

What Disciple-LTA does is to reduce a complex hypothesis to hypotheses that are simple enough to be reliably assessed based on the available evidence. For example, it is easier to "Assess whether Al Qaeda considers self-defense as a reason to obtain nuclear weapons," based on the available evidence (as discussed in the next section), than it is to "Assess whether Al Qaeda has nuclear weapons based on the characteristics associated with the possession of nuclear weapons."

Table 4. The top solution synthesis step from Figure 3.

I have determined that	
It is almost certain that Al Qaeda has reasons to obtain nuclear weapons.	[S2]
It is an even chance that Al Qaeda has desire to obtain nuclear weapons.	[S3]
It is a remote possibility that Al Qaeda has the ability to obtain nuclear weapons.	[S4]
Therefore I conclude that	
Based on its reason, desire, and ability to obtain nuclear weapons, it is a remote possibility that Al Qaeda has nuclear weapons.	[S1]

5. EVIDENCE CREDENTIALS

In the previous section we have shown how Disciple-LTA reduces a complex hypothesis analysis problem to simpler hypothesis analysis problems. In this section we will show how the simplest hypothesis analysis problems are solved based on the available evidence. This requires the development of often stunningly complex arguments that link evidence to hypotheses by establishing the three major credentials of evidence: *relevance*, *believability*, and *inferential force or weight*.

The *relevance* answers the question: *So what? How does this datum or item of information, whatever it is, bear on what an analyst is trying to prove or disprove?* The *believability* answers the question: *Can we believe what this item of intelligence information is telling us?* The *inferential force or weight* answers the question: *How strong is this item or body of relevant evidence in favoring or disfavoring various alternative hypotheses or possible conclusions being entertained?*[15] Figure 4 provides a simple illustration of these evidence credentials, as discussed below.

In an interview with Hamid Mir, published in Dawn, a Pakistani English newspaper, Osama bin Laden made the claim that Al Qaeda has nuclear weapons to defend itself. We refer to this item of information as EVD-Dawn-Mir01-01c (see Figure 4). Our problem, shown at the top of Figure 4, is:

Assess to what extent the item of evidence EVD-Dawn-Mir01-01c favors the hypothesis that Al Qaeda considers self-defense as a reason to obtain nuclear weapons. [P5]

We can solve the problem [P5] by reducing it to two simpler problems:

Assess to what extent EVD-Dawn-Mir01-01c favors the hypothesis that Al Qaeda considers self-defense as a reason to obtain nuclear weapons, assuming that EVD-Dawn-Mir01-01c is believable. [P6]

Assess the extent to which EVD-Dawn-Mir01-01c is believable. [P7]

[P6] is the problem of determining the degree of *relevance* of an item of evidence (i.e., EVD-Dawn-Mir01-01c) to a hypothesis. In general, this is a complex problem of developing a relevance argument linking the content or substance of the item of evidence to that of the hypothesis. The relevance argument may be developed by employing the problem reduction and solution synthesis approach discussed in the previous section. However, because the relevance argument depends on the substance or content of the item of evidence of which there is a near infinite variety, there will be no book or other reference source which, in any situation, will tell an intelligence analyst what the links in a relevance argument should be. The analyst must imagine these links based on her experience and stock of knowledge. This is where the analyst's imaginative reasoning becomes so important. However, Disciple-LTA can learn from a specific relevance argument developed by the analyst and may help develop similar arguments in the future.[16] In our example, however, the relevance problem is very simple. Indeed, according to EVD-Dawn-Mir01-01c, Osama bin Laden claimed that Al Qaeda has nuclear weapons for self-defense and this item of evidence is obviously very relevant to the problem of assessing whether Al Qaeda considers self-defense as a reason to obtain nuclear weapons.

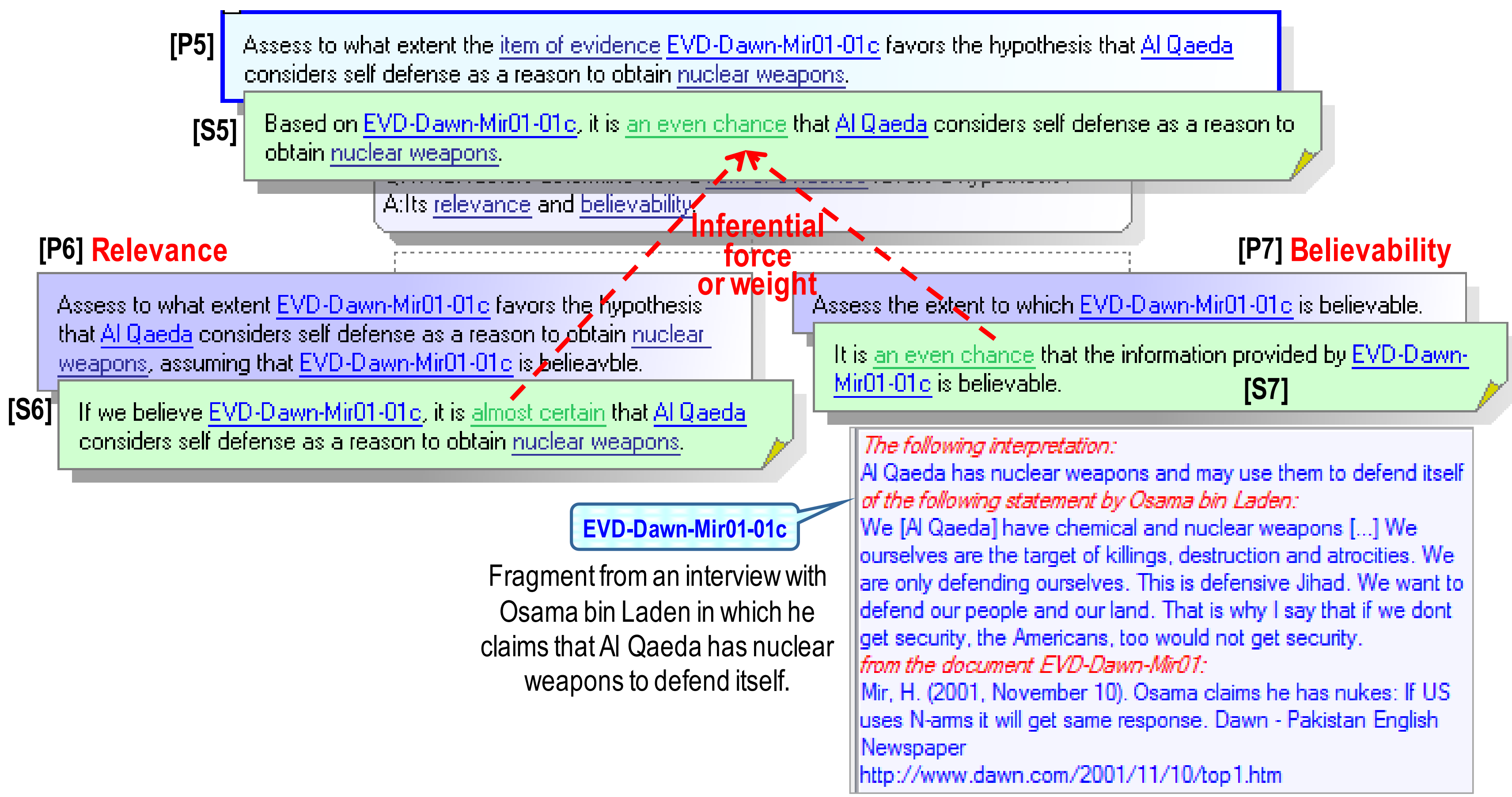

Figure 4. A simple illustration of evidence credentials.

However, just because we have evidence about an event does not entail that the corresponding event did occur. Therefore our empirical testing involves inference about whether the event did occur, which is the problem [P7]. Solving [P7] supplies the *believability*-related foundation for inferences about the degree to which evidence EVD-Dawn-Mir01-01c favors the considered hypothesis. This involves a significant amount of critical reasoning on the part of Disciple-LTA, which will significantly support the analyst, as discussed in the Section 8.

Let us now assume that we have obtained the solutions [S6] and [S7] of the problems [P6] and [P7], respectively, as indicated in Figure 4:

If we believe EVD-Dawn-Mir01-01c, it is almost certain that Al Qaeda considers self defense as a reason to obtain nuclear weapons. [S6]

It is an even chance that the information provided by EVD-Dawn-Mir01-01c is believable. [S7]

These probabilistic estimates (i.e. “almost certain” and “an even chance”) are combined (through a “min” function) to determine the *inferential force or weight* of EVD-Dawn-Mir01-01c on the considered hypothesis:

Based on EVD-Dawn-Mir01-01c, it is an even chance that Al Qaeda considers self defense as a reason to obtain nuclear weapons. [S5]

Disciple-LTA’s use of the *min* function to combine the *relevance* of EVD-Dawn-Mir01-01c with its *believability*, to estimate its *inferential force or weight* on the considered hypothesis is reasonable. Indeed, consider a highly relevant item of evidence E which is not believable. This item of evidence will not influence us much in accepting the hypothesis H. The same is true for a believable item of evidence which is not relevant. Therefore, in both cases, the inferential force of E on H is very small, which is consistent with using the min function.

In its current implementation, Disciple-LTA uses an approach to combining probabilistic estimates based on Fuzzy probabilities.[17] However, one can also define other types of synthesis functions.

6. EVIDENCE-BASED HYPOTHESIS ASSESSMENT

Disciple-LTA may develop very complex arguments for hypothesis assessment by employing the general problem-reduction/solution-synthesis approach discussed in Section 4. Figure 5 illustrates the process of assessing the hypothesis H_1 (problem [P8]) which is first reduced to three simpler hypotheses, H_{11}, H_{12}, and H_{13} (problems [P9], [P10] and [P11], respectively). Each of these hypotheses is assessed by considering both *favoring evidence* and *disfavoring evidence* (e.g., problems [P12] and [P13]). 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., [P14]). This requires assessing both the *relevance* of E_1 to H_{11} (problem [P16]) and the *believability* of E_1 (problem [P17]). 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. [S16]
It is likely that E_1 is true. [S17]

By compositing the solutions [S16] and [S17] (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. [S14]

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. [S15]

By composing the solutions [S14] and [S15] (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. [S12]

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. [S13]

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. [S9]

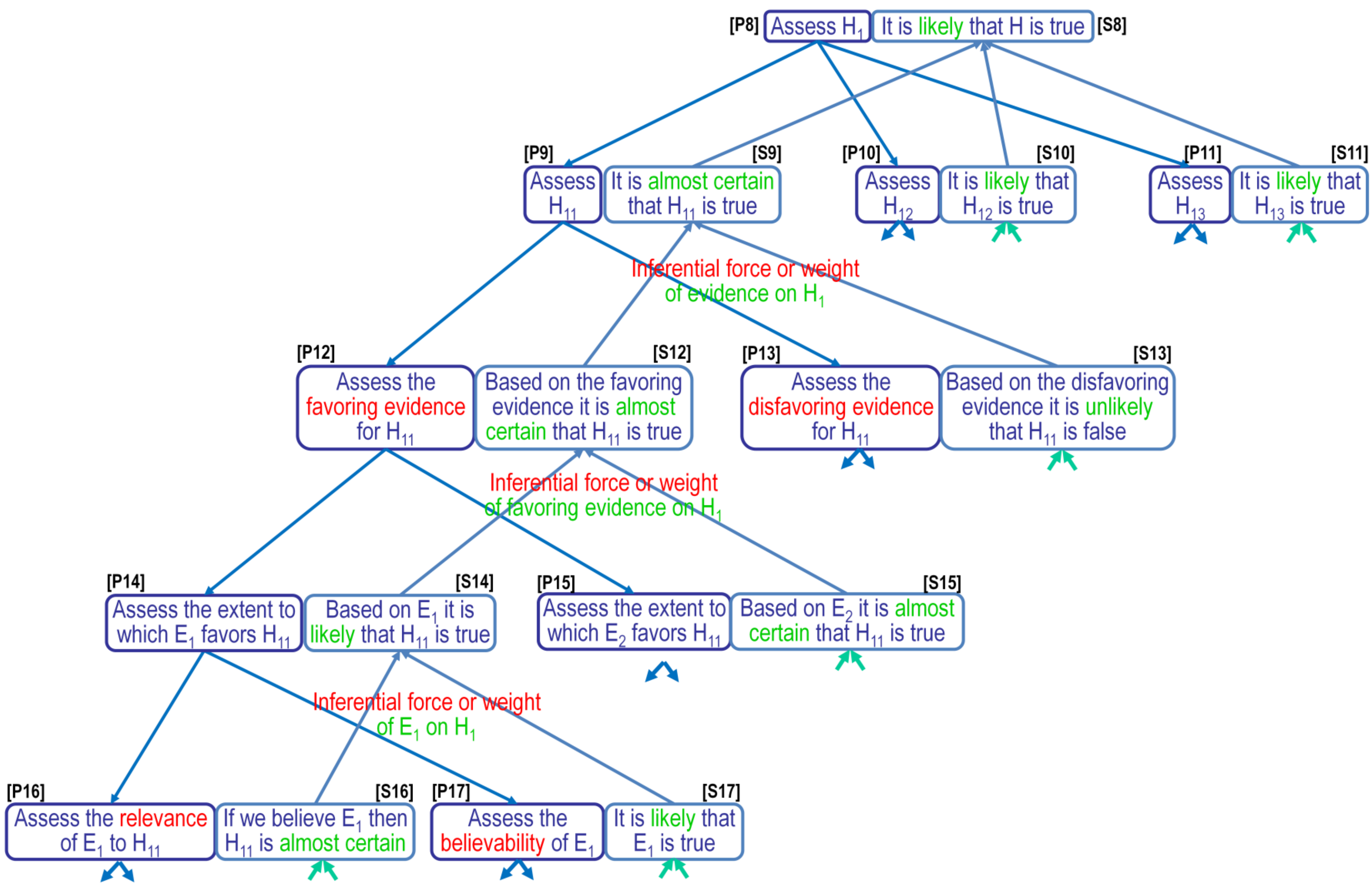

Figure 5. Wigmorean network for hypothesis assessment generated by Disciple-LTA.

The sub-hypotheses H_{12} and H_{13} are assessed through a similar process:

It is likely that H_{12} is true. [S10]

It is likely that H_{13} is true. [S11]

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. [S8]

7. ASSUMPTION-BASED REASONING AND WHAT-IF SCENARIOS

Disciple-LTA allows an analyst to select any problem from an analysis tree and provide its solution in the form of an assumption and an optional justification.[18] To illustrate the use of assumptions, let us consider the analysis tree from Figure 6 corresponding to the following problem:

Assess whether Al Qaeda has the ability to obtain nuclear weapons. [P18]

This problem is reduced to three simpler problems:

Assess whether Al Qaeda might receive nuclear weapons. [P19]

Assess whether Al Qaeda has the ability to buy nuclear weapons. [P20]

Assess whether Al Qaeda has the ability to make nuclear weapons. [P21]

For the first and the third of these sub-problems the analyst provided solutions in the form of assumptions which appear with yellow background in Figure 6, to distinguish them from the regular solutions. For example, the analyst made the assumption that the solution of [P19] is [S19]:

It is a remote possibility that Al Qaeda will receive nuclear weapons. [S19]

Problem [P20] is reduced to two simpler problems, [P22] and [P23]. [P22] is further reduced to the problems [P23] and [P24]. [P23] is solved by making the assumption [S23] which is shown with yellow background in the bottom left of Figure 6. The solution of [P24] is [S24], obtained through a reasoning which is not shown in Figure 6.

As illustrated in this example, an assumption can be made at any level of an analysis tree. That is, through the use of assumptions, Disciple-LTA allows the analysis to drill-

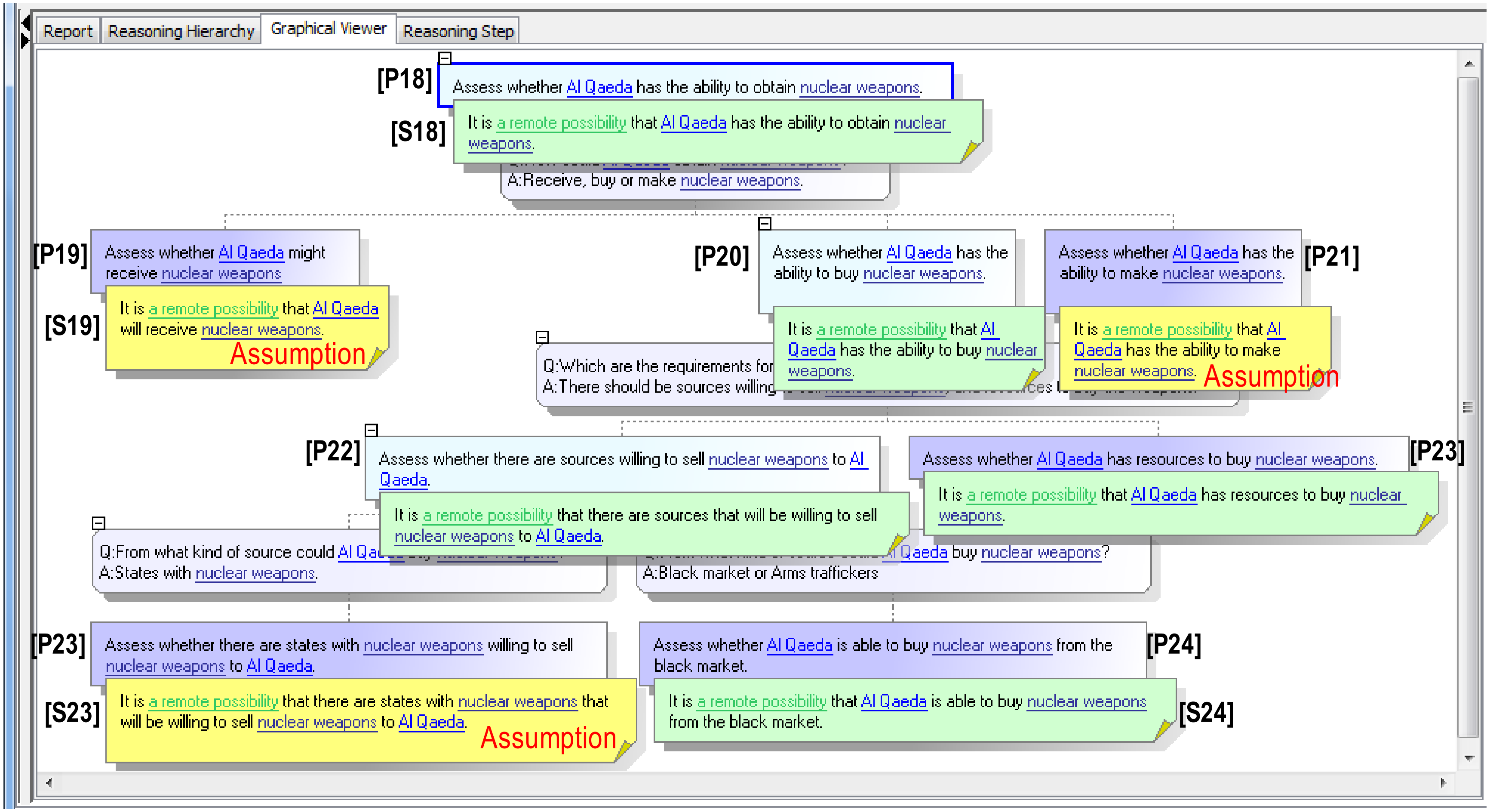

Figure 6. Assumption-based reasoning.

down in the analysis tree as much as desired. This is particularly useful when the analyst does not have time to analyze all the sub-hypotheses of an investigated hypothesis or does not have evidence to perform the analysis of some of the sub-hypotheses. In each of these cases, the analyst can provide solutions for these sub-hypotheses in the form of assumptions.

By defining, enabling and disabling such assumptions, the analyst may study various What-if scenarios. For example, the analyst may consider alternative solutions to [P19] or [P23] and study how each of them changes the solution of the top level problem.

8. BELIEVABILITY ASSESSMENTS BASED ON A SUBSTANCE-BLIND CLASSIFICATION OF EVIDENCE

Attempts to categorize evidence in terms of its substance or content would be a fruitless task, the essential reason being that the substance or content of evidence is virtually unlimited. What we have termed a *substance-blind* classification of evidence refers to a classification of recurrent forms and combinations of evidence based, not on substance or content, but on the inferential properties of evidence.[19] One major reason we have had for dwelling upon the three credentials of evidence (its relevance, believability, and inferential force or weight) is that these three credentials supply a very useful basis for categorizing the *individual items of evidence* we have in any intelligence analysis. These classifications guide the process of building arguments that link evidence to hypotheses. It happens that there are two forms of relevance, direct and indirect. *Directly relevant evidence* is that which can be linked directly to hypotheses being considered by a defensible chain of reasoning or argument. For example, both E_1 and E_2 in Figure 5 and EVD-Dawn-01-01c in Figure 4 are directly relevant items of evidence. *Indirectly relevant evidence* has no such direct linkage but bears upon the strength or weakness of links in chains of reasoning set up by directly relevant evidence. Consider, for example, the problem "Assess the believability of E_1" from the bottom of Figure 5. Any item of evidence that might be used in solving this problem would be indirectly relevant evidence. Indirectly relevant evidence would also be any evidence used in solving the problem "Assess the extent to which EVD-Dawn-Mir01-01c is believable", from the bottom right of Figure 4. The term *meta-evidence* is also appropriate since ancillary evidence is evidence about other evidence.

In what follows, we focus on the believability credential and the recurrent forms of individual items of evidence it suggests. It is here that we identify specific attributes of the believability of various recurrent types of evidence without regard to their substance or content.

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*. Figure 7 shows a substance-blind classification of evidence based on its believability credentials. This classification is discussed in the following sections.

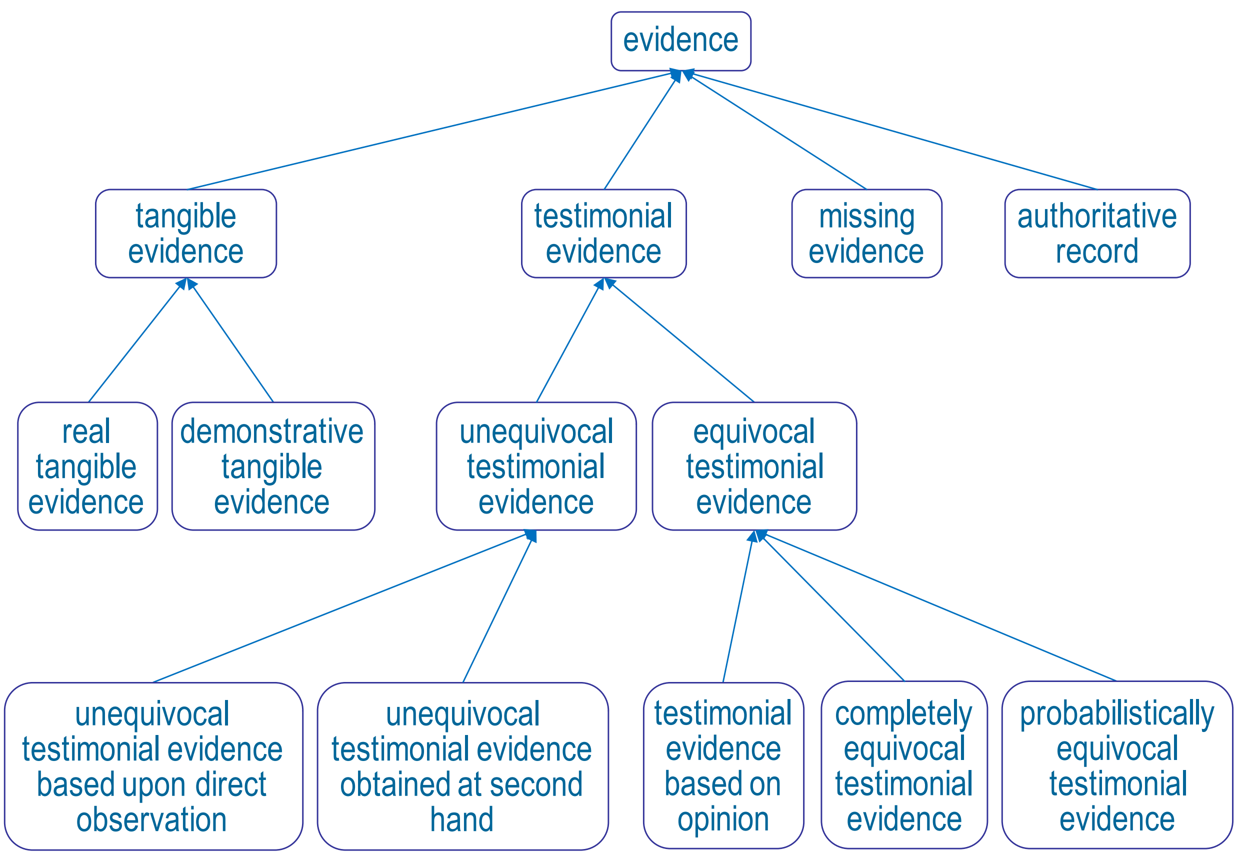

Figure 7. Substance-blind classification of evidence.

8.1 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.*[20] 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. Examples 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.

8.2 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*.[21] 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 Figure 7, 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", and because of many sources of uncertainty, 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 particular probabilistic equivocation as an assessment by the source of his own observational sensitivity.

8.3 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.

8.4 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.

8.5 Believability Assessment with Disciple-LTA

Disciple-LTA knows about the types of evidence shown in Figure 7 and how their believability should be evaluated. For example, Figure 8 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 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.

Based on system’s knowledge base, all these problems can be further reduced to even simpler problems. Alternatively, the system may have general knowledge in its knowledge base about these believability characteristics of Osama bin Laden. Yet another possibility is for the analyst to provide solutions for these problems in the form of assumptions. Once the solutions of the simplest problems are obtained, they are

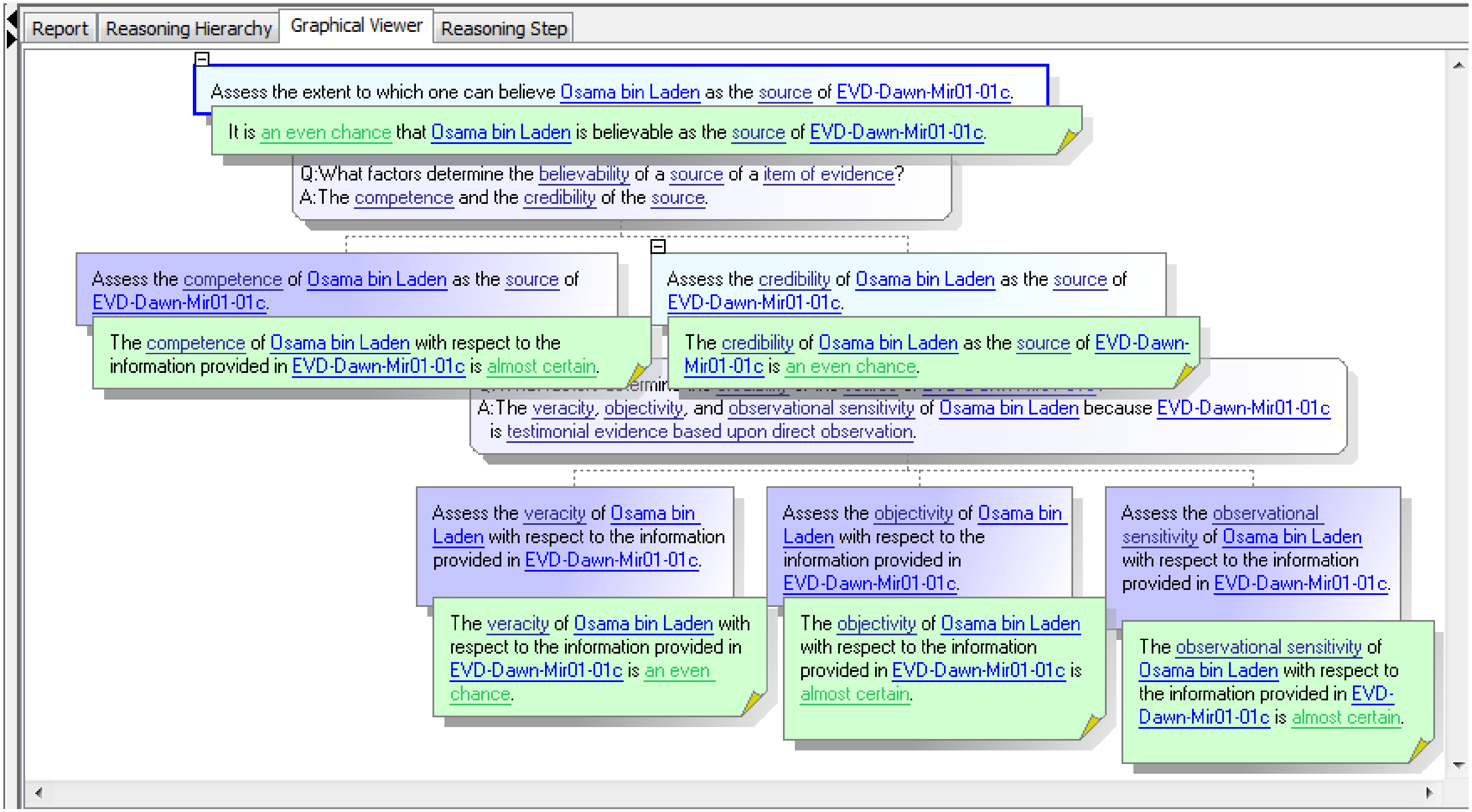

Figure 8. Source believability assessment.

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. These computations are automatically performed by Disciple-LTA. But this is only one component in the more complex reasoning of assessing the believability of EVD-Dawn-Mir01-01c, as will be discussed in Section 9.

8.6 Chains of Custody

In the previous sections we have discussed the different types of evidence (such as testimonial or tangible), and the ingredients of their believability assessment. However, very rarely, if ever, has the analyst access to the original evidence. Most often, what is being analyzed is a piece of evidence that has undergone a serious of transformations through a *chain of custody*. Here we have borrowed an important concept from the field of law where a chain of custody refers to the persons or devices having access to the original source evidence, the time at which they had such access, and what they did to the original evidence when they had access to it. The important point here is to consider the extent to which what the analyst finally receives is an *authentic* and *complete* account of what an original source provided. Uncertainties arising in chains of custody of intelligence evidence are not always taken into account. One result is that analysts can often mislead themselves about what the evidence is telling them. The original evidence may be altered in various ways at various links in chains of custody. Consider, for example, the situation where our analyst, Clyde, receives an item of testimonial evidence from a source code-named Wallflower who reports that five days ago he saw a member of the government of Iraq, Emir Z., leaving a building in Ahwaz, Iran in which the Iranian Islamic Revolutionary Guards Corp has offices.[22] Wallflower's original testimony is first *recorded* by an intelligence professional and then it is *translated* from Farsi into English by a paid translator. This translation is then *edited* by another intelligence professional; and then the edited version of this translation is *transmitted* to an intelligence analyst. There are four links in this conjectural chain of custody of this original testimonial item: recording, translation, editing, and transmission. Various things can happen at each one of these links that can prevent the analyst from having an authentic account of what our source originally provided. For instance, we should be concerned both by the *competence* of the translator (in Farsi, in English, and in the subject matter), and by his *credibility*.

There are many possible chains of custody, for different types of evidence, as illustrated in Figure 9. However, they can all be characterized by a chain of basic evidence transformation processes (such as translation, editing, or transmission). Moreover, for each such process, one can identify the ingredients and the arguments of its believability assessment, just as for the different types of the evidence. Disciple-LTA employs a systematic approach to the assessment of the believability of items of evidence obtained through a chain of custody.[23]

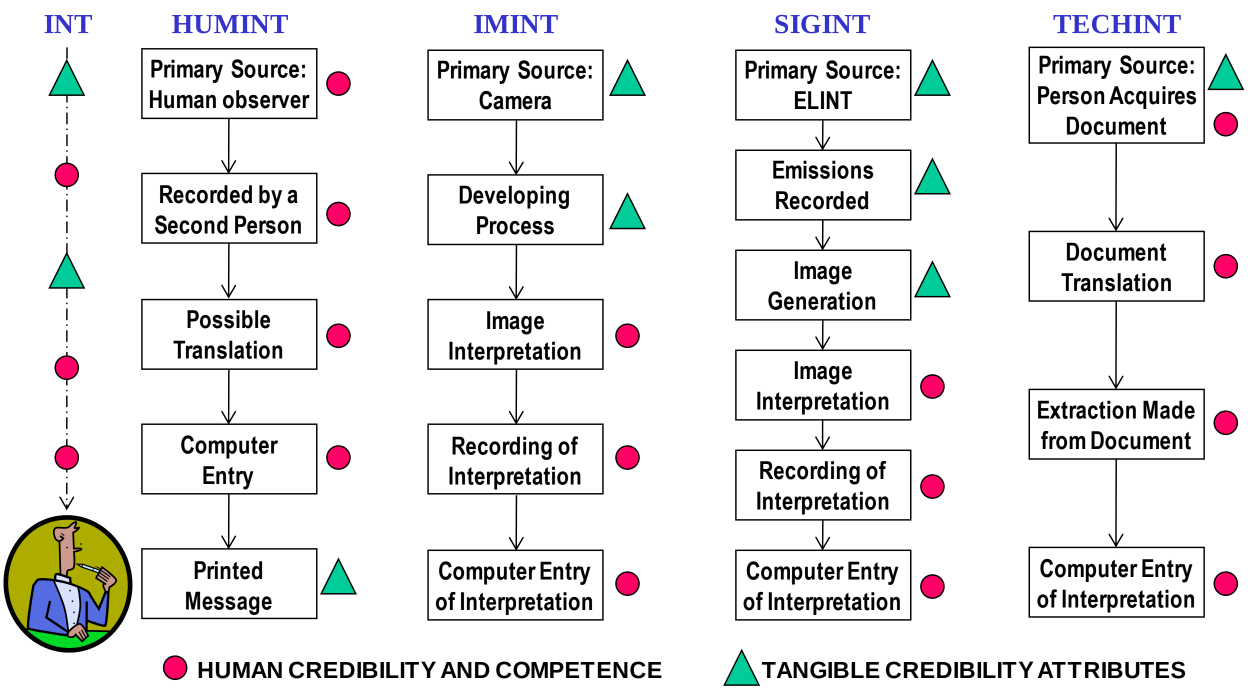

Figure 9. Typical chains of custody for different INTs.

In this section we have shown that we can classify all evidence, regardless of its substance or content, into just a few categories of recurrent forms and combinations of evidence. That is why this classification is called *"substance-blind"*. This classification of evidence is based on its inferential properties rather than upon any feature of its substance or content. Knowledge of these substance-blind forms and combinations of evidence pays great dividends. Such knowledge informs us and Disciple-LTA how to evaluate the believability of evidence, based on its type. It allows us and Disciple-LTA to more easily assess evidence coming from different sources and to compare the evidence and conclusions reached from it in different intelligence analyses and at different times.

9. ABSTRACTION OF REASONING

Analyses of complex hypotheses from masses of evidence result in the generation of very large reasoning trees, some with thousands of nodes. To help browse and understand such a complex analysis, Disciple-LTA will display an abstraction of it which only shows abstractions of the main sub-problems considered in the analysis. This is illustrated in the left-hand side of Figure 10. The top line is the analyzed problem and its solution: "Assess whether Al Qaeda has nuclear weapons: likely." This problem is reduced to the three simpler problems of assessing whether Al Qaeda has reasons, has desire, and has the capability to obtain nuclear weapons. The left hand side of Figure 10 shows the abstractions of these three sub-problems and their solutions: "Reasons: almost certain", "Desire: almost certain" and "Capability: a remote possibility." Each of these abstract problems can be expanded to browse its abstract sub-problems and their solutions. For example, to assess whether Al Qaeda has reasons to obtain nuclear weapons, Disciple-LTA considers all the likely reasons (deterrence, self-defense, spectacular operations, etc.), attempting to assess each of them by considering both favoring and disfavoring evidence. As shown in the left-hand side of Figure 10, there are two favoring items of evidence for the hypothesis that Al Qaeda considers self defense as a reason to obtain nuclear weapons. The second one, "EVD-Dawn-Mir01-01c," is also shown in the bottom right-side of Figure 10. It is a fragment from an interview taken by Hamid Mir to Osama bin Laden in which bin Laden stated that "Al Qaeda has nuclear weapons and may use them to defend itself." Disciple-LTA will assess to what extent EVD-Dawn-Mir01-01c favors the hypothesis that Al Qaeda considers self defense as a reason to obtain nuclear weapons by considering both its *relevance* and its *believability*. The believability is a function of the believability of Hamid Mir and that of Osama bin Laden because this item of evidence is unequivocal testimonial evidence obtained at second hand (see Figure 7). Indeed, here Hamid Mir is telling us what (presumably) Osama bin Laden has told him. Further on, the believability of Osama bin Laden is a function of his *competence* and *credibility* (as discussed in Section 8.5). His credibility is a function of his *veracity*, *objectivity* and *observational sensitivity*. In this example it is assumed that the

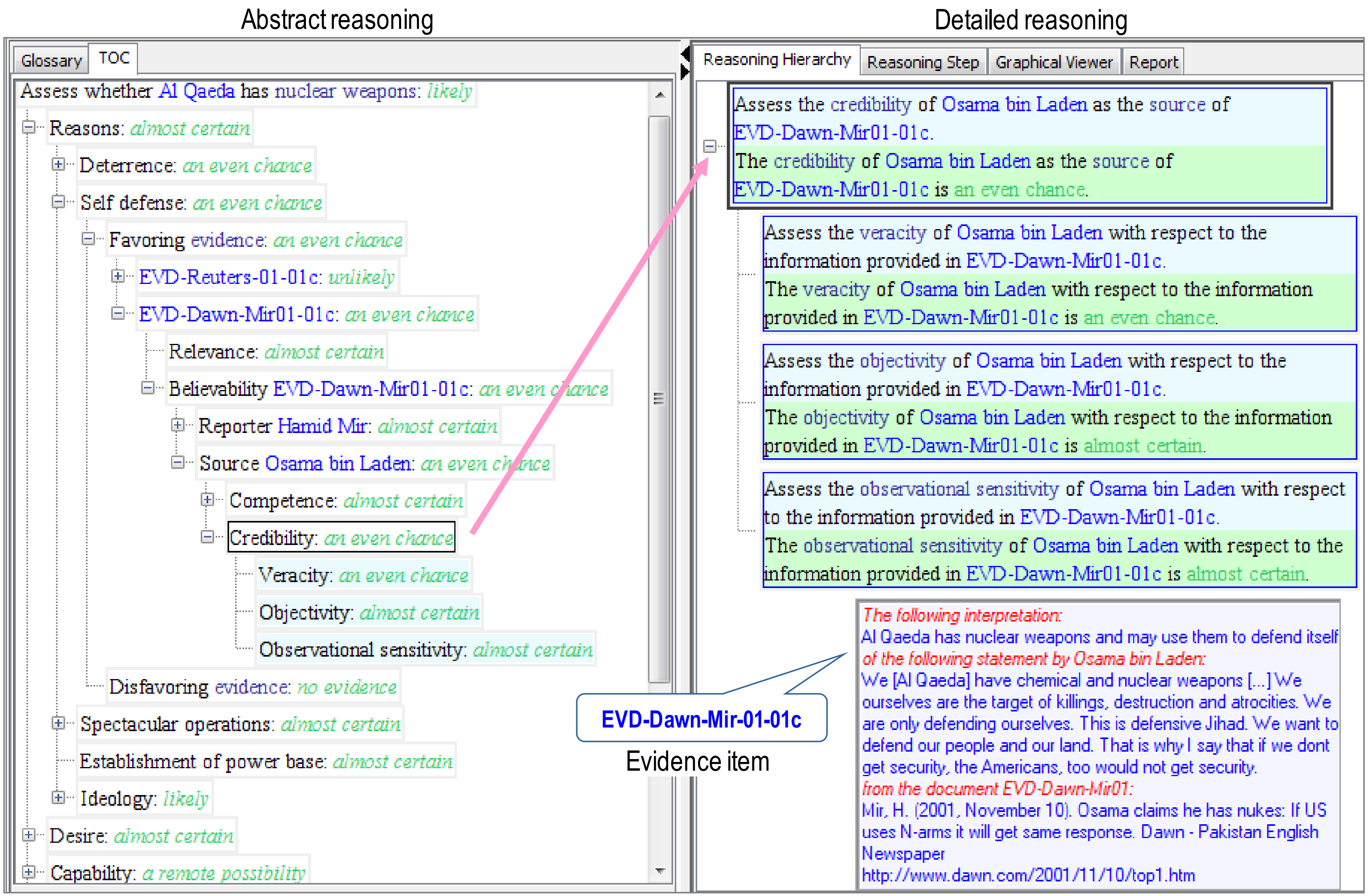

Figure 10. Abstract reasoning and detailed reasoning.

knowledge base of Disciple-LTA contains some probabilistic estimates for these credibility factors: “an even chance” for veracity, “almost certain” for objectivity, and “almost certain” for observational sensitivity. They are shown in green in the left-hand side of Figure 10. Alternatively, Disciple-LTA may estimate these values by using the problem reduction / solution synthesis approach, based on other items of evidence. Yet another possibility is for the analyst to provide these values as assumptions.

The left-hand side of Figure 10 presents only an abstract, simplified view of the reasoning process that even omits some intermediary reasoning steps. However, when the user clicks on a reasoning step in this abstract view (e.g. “Credibility: an even chance”), the right-hand side shows a more detailed reasoning for that step. In this case it shows the actual problems and their solutions for assessing the credibility of Osama bin Laden based on his veracity, objectivity, and observational sensitivity (by using “min” as the synthesis function).

Let us notice that Disciple-LTA knows how to assess the believability of EVD-Dawn-Mir01-01c based on its type in the substance-blind classification from Figure 7. Notice how many reasoning steps are performed by Disciple-LTA in order to determine that the believability of EVD-Dawn-Mir01-01c is an even chance. Notice also that there are many intermediate reasoning steps linking an item of evidence to the top level hypothesis, not all of them shown in Figure 10. EVD-Dawn-Mir01-01c is favoring evidence for the hypothesis that self-defense is a reason for Al Qaeda to obtain nuclear weapons. This and other potential reasons are analyzed to determine whether Al Qaeda, in general, has reasons to obtain nuclear weapons. It is also analyzed whether Al Qaeda has desire to obtain nuclear weapons and whether it has the ability to obtain nuclear weapons. Based on the evaluation of its reasons, desire and ability, Disciple-LTA assesses the likelihood that Al Qaeda has nuclear weapons. It is through such complex arguments that EVD-Dawn-Mir01-01c has a certain inferential force on the hypothesis that Al Qaeda has nuclear weapons.

10. IMPROVING STRUCTURED ANALYTIC METHODS WITH DISCIPLE-LTA: THE CASE OF THE ANALYSIS OF COMPETING HYPOTHESES

The intelligence analysis concepts and methods embedded in Disciple-LTA, which are based on the Science of Evidence and Artificial Intelligence, particularly the systematic approach to the development of argumentation structures, the substance-blind classification of evidence and the associated procedure for assessing the believability of evidence, the drill-down analysis and assumptions-based reasoning, may help the analysts perform better analyses, no matter what analysis methods they use. To justify this claim, we consider in this section what is probably the most popular structured analytic method, Richards J. Heuer's *Analysis of Competing Hypothesis* [ACH].[24] We show how ACH, which has found favor among many intelligence analysts and is used in many advanced analysis courses, can be significantly improved, by employing some of the concepts and methods embedded in Disciple-LTA. Our present comments are based upon a very recent account of a system being developed to implement the ACH approach.[25]

10.1 Using the Substance-blind Classification of Evidence

The basis of ACH consists of a matrix in which various items of interest in an intelligence analysis are recorded, as illustrated in the abstract example from Table 1. In this two-dimensional matrix, analysts first list the substance or content of the evidence in the first column. Then, in the second column, analysts list what Heuer calls “source type,” which should guide them in evaluating the credibility and relevance of evidence (columns 3 and 4).

Table 1. An illustration of Heuer’s Analysis of Competing Hypotheses.

Evidence	Source Type	Credibility	Relevance	H_1	H_2	H_3
E_1	Inference	medium	high	C	C	I
E_2	Assumption	high	low	C	I	C
E_3	Intel Reporting	low	high	I	C	I
E_4	HUMINT	medium	medium	C	C	C
E_5	Liaison	high	low	C	C	I
E_6	Lack of Intel Reporting despite vigorous search	low	medium	I	C	C
E_7	Contrarian hypothesis	high	high	C	I	I

Here are the actual examples Heuer provides of source types: Inference, Assumption, Intel Reporting, HUMINT, Liaison, Lack of intelligence reporting despite vigorous search, Contrarian hypothesis. One problem with this classification is that the believability/credibility of evidence in the same category (e.g. Liaison) is evaluated based on certain credentials if it is tangible evidence (e.g., authenticity) and on other credentials if it is testimonial evidence (e.g., veracity, objectivity, etc.). Thus this classification does not help with this evaluation.

As discussed in Section 8, there is a “substance-blind” classification of evidence that emerges precisely from the fact that entirely different believability or credibility questions must be asked of tangible and testimonial evidence. Therefore, an improvement of the ACH method is to use the forms of evidence shown in Figure 7, which will guide the analyst in assessing its believability. In fact, several of Heuer’s types can easily be mapped to these forms. For example, HUMINT is a species of testimonial evidence; Intel Reporting may either involve testimonial or tangible evidence; Liaison evidence (obtained from contacts with representatives of friendly or neutral governments) may be either tangible or testimonial in nature. Heuer's "Lack of Intell Reporting despite vigorous search" qualifies as "missing evidence" having potential inferential value, as discussed in Section 8.3.

Heuer uses a very broad interpretation of evidence as “all the factors that influence an analyst’s judgment about the relative likelihood of the hypotheses.”[26] However, according to the Science of Evidence[27] and as discussed in Section 5, all evidence, regardless of its substance or content, has three credentials that must be established by defensible arguments: *relevance*, *believability or credibility*, and *inferential force or weight*. From this point of view, three of the examples provided by Heuer (inference, assumption, and contrarian hypothesis) do not qualify as evidence. We agree with Heuer that they play an important role in evidential reasoning, but they should be accounted for not as evidence (how do we ever establish the credibility of an assumption or a hypothesis?), but as components of arguments. For example, assumptions could be used to assess the relevance or the believability of evidence, as illustrated in Section 7 and discussed below.

10.2 Assessing the Believability of Evidence

In the third column ACH requires the analyst to rate the credibility of the "source type" of an item of intelligence evidence as high, medium or low. First, as discussed in Section 8, we think that it is better to talk about the "believability" of evidence which may also include "competence" considerations in addition to "credibility" ones.

As discussed in Sections 8.5, 8.6 and 9, believability assessments for some items of evidence may be very complex, especially if these items have been obtained through chains of custody.[28] Disciple-LTA has a lot of knowledge about the believability of evidence and its constituents and supports the analyst in making these assessments. For example, it knows about the necessity for determining the *authenticity*, *accuracy*, and *reliability* of the *demonstrative tangible evidence.* It knows that it has to establish both the competence and the credibility of the human sources of testimony. As discussed in Section 8.2, source credibility and source competence are entirely different characteristics, each with its own ingredients. For example, in order to determine the credibility one has to determine the source's *veracity*, *objectivity*, and *observational sensitivity.* On the other hand, in order to determine the competence one would need to determine the source's *access* and *understandability*. As shown by Schum and Morris, each of these assessments may be a very complex.[29] It is therefore important to assist the analysts in performing these assessments, for instance, by incorporating into ACH the Disciple-LTA procedures for evaluating the believability of evidence which are discussed in Section 8. In particular, the arguments developed with Disciple-LTA for establishing the believability of evidence may include the use of assumptions.

10.3 Assessing the Relevance of Evidence

In the fourth column of the ACH table the analyst has to rate the relevance of an item of evidence as high, medium, and low. However, if the relevance arguments are not specifically constructed they can never be subjected to any form of critical reasoning. Disciple-LTA can help with this issue because it involves both the top-down and bottom-up argument-structuring methods discussed in Section 4, and draws upon, and even extends, Wigmore's concern and methods for assessing the relevance of evidence.[30]

As discussed at the beginning of Section 8, there are two forms of relevance, direct and indirect. *Directly relevant evidence* is that which can be linked directly to hypotheses being considered by a defensible chain of reasoning or argument. *Indirectly relevant evidence* has no such direct linkage but bears upon the strength or weakness of links in chains of reasoning set up by directly relevant evidence. Consider, for example, an item of evidence that says nothing about the hypotheses being considered in the

ACH method but would allow us to infer that the source of a relevant evidence item is not credible. To use such indirectly relevant evidence, the ACH method would need to be extended to allow the development of arguments for both the believability and the relevance credentials, arguments that could be developed with Disciple-LTA.

10.4 Assessing the Likelihood of Hypotheses

The last columns in the ACH table correspond to the hypotheses being considered in the analysis at hand. A significant advancement of ACH over the conventional intuitive analysis approach is precisely the requirement to look at several competing hypotheses. In contrast, conventional intuitive analysis focuses on what is suspected to be the most likely hypothesis and then assesses whether or not the available evidence supports it. This may lead to wrong conclusions because the same evidence may also support other hypotheses.

In the column corresponding to a hypothesis, the analyst grades the bearing of an item of evidence on that hypothesis as either *consistent* [C] or *inconsistent* [I]. Then the most likely hypothesis is the one with the least evidence against it, that is, the hypothesis with the least number of Is. But there is no indication of how relatively strong any of the Is are. Suppose we have ten items of evidence for which H_1 and H_2 have the same number of Is. How do we decide which hypothesis to accept, given the fact that the evidence items assessed as I under H_1 might be different from the evidence items assessed as I under H_2? In their extension of the ACH method, Good and his colleagues attempted to address this issue by associating numbers to the high, medium, and low gradations of credibility and relevance, and scorings the competing hypotheses.[31] The problem with this approach is that numbers applied to hypotheses will have little meaning in the absence of any specific relevance arguments, considerations of credibility and competence attributes for different sources of evidence, and characteristics of the evidence itself. This also applies to any ordinary probability assessments under alternative hypotheses that will have little meaning either in the absence of specific arguments justifying them. In that sense, Good's extension of ACH may do more harm than help because it may provide the analysts with a false sense of confidence rather than encouraging them to give more careful attention to the arguments necessary to justify their conclusions regarding the competing hypotheses.

An additional difficulty with the ACH method is that it requires that we begin with what Heuer calls a full set of hypotheses;[32] presumably this means that the hypotheses are *mutually exclusive* and *exhaustive*. In some cases, such as in the example Heuer provides, we may consider a set of hypotheses that occur in response to a specific question we have been asked. The analysis example Heuer provides is in answer to the question: *What is the status of Iraq's nuclear weapons program?* The three hypotheses he lists as being a full set are: *H_1: Dormant or shut down*; *H_2: Has been started up again*; *H_3: Weapon available within this decade*. It could of course be argued about whether the hypotheses on this list are in fact either exhaustive or mutually exclusive. For example, H_3 and H_2 are not mutually exclusive. If the weapons program has been started up again (H_2) then we might infer that there might be *at least* one weapon available within this decade (H_3). Conversely, for a weapon to be available Iraq must have started-up its weapons program. What this shows is that it may be difficult to assure that we have a complete set of mutually exclusive hypotheses. However, if the set of hypotheses is not complete it may just be the case that the most likely hypothesis is among the missing ones. Disciple-LTA may help with this issue by estimating the likelihood of each of the competing hypotheses considered or, at least, the one selected through the ACH method. If the ACH-selected hypothesis does not have a high enough likelihood, then this is an indication that additional hypotheses should be considered.

A simplification made by the ACH method is to consider that both the credibility/believability and the relevance of an item of evidence are independent of the particular hypothesis being considered. Let us consider, for example, an item of evidence revealing the number of years needed by North Korea to develop its nuclear program. This item of evidence is relevant to *H_3: Weapon available to Iraq within this decade*, but it is not at all relevant to the other two hypotheses, *H_1: Iraqi nuclear program is dormant or shut down*; *H_2: Iraqi nuclear program has been started up again*. One way to address this issue is to simply estimate a different believability and relevance for each hypothesis.

James Bruce, who is well-known for his valuable work on the importance of epistemology in intelligence analysis, discusses reasons why the ACH method does represent a significant advance over analytic methods that are entirely unsystematic and have so often resulted in a favored hypothesis being uncritically endorsed on a very shaky evidential foundation.[33] He also mentions various reasons why the ACH method enjoys current popularity among many intelligence analysts. However, the example he provides illustrating the virtues of ACH also illustrates one of its most severe limitations.

He mentions the unjustified conclusions reached about Saddam's alleged possession and development of WMDs based on the reports provided by "Curveball". Bruce argues that had these reports been subjected to analysis using ACH, a possibly different conclusion would have been reached, especially regarding bioweapons. There are, however, some good reasons why ACH might not have helped regarding this conclusion. The trouble here is that the ACH method says nothing about the attributes of the competence and credibility of HUMINT or the attributes of the credibility of various forms of tangible evidence such as the diagrams of bioweapons facilities that Curveball provided. We are just as concerned as James Bruce about the epistemology of intelligence analysis but we are especially concerned that intelligence analysts be provided with appropriate background knowledge regarding such tasks as assessing the credibility of sources of evidence and establishing the relevance of evidence on alternative hypotheses. A system developed by one of us for CIA, called MACE (Method for Assessing the Credibility of Evidence), shows the specific competence and credibility attributes we must consider for HUMINT sources.[34] This system would have been especially useful in assessing the competence and credibility of Curveball. Analysts would have been prompted to ask questions they did not ask about Curveball, but for which we did have answers. And our system Disciple-LTA has significant knowledge about the properties, uses, discovery and marshaling of evidence that it can share with the intelligence analysts who use it. It also knows about the necessary credibility-related questions that form the basis for MACE. This knowledge can be integrated into the ACH method, as suggested above.

There is problem that seems endemic in intelligence analysis that the ACH method does not address. The problem is that, in so many situations of interest to the Intelligence Community, we have a seamless activity in which we have evidence in search of hypotheses *at the same time* with hypotheses in search of evidence. Suppose we wish to consider hypothesis H_2, that Iraq's weapons program has been started up again. There is no mechanism in ACH for putting this hypothesis to use in generating new lines of evidence and inquiry. This mechanism should address the question: *What things need to be tested by what evidence in order to sustain this hypothesis?* What this amounts to is generating main lines of argument under H_2, showing what evidence would be necessary to prove or disprove the hypothesis that the Iraqis have started up their weapons program. Many possibilities come to mind such as the acquisition of necessary materials, the bringing together of necessary talented scientific and technical people, the development of facilities necessary in the development of weapons of various sorts. You recognize here that this is what we described in Section 3 as hypotheses in search of evidence. To put some hypothesis to use requires us to generate arguments from it that will eventually identify classes of observable evidence necessary to sustain this hypothesis. But the world continues to change as we are attempting to understand events in it. The result is that we must continually generate new hypotheses or revise the ones we have constructed. Thus, a major item left out in ACH is the crucial importance of the discovery process in which we have evidence in search of hypotheses at the same time with hypotheses in search of evidence. As discussed in Sections 2 and 3, Disciple-LTA promotes a systematic approach to this complex issue, although the evidence in search of hypothesis part needs further development.

A very good feature of the ACH method is that it shows how individual items of evidence relate to the competing hypotheses. This suggests an improvement of Disciple-LTA with a module that will automatically compare the analyses of competing hypotheses, to reveal differences in the evidence used and the assumptions made, including a focus on areas with less evidential support.

Heuer has conceived ACH as a manual method that can be easily used by the analysts and has therefore made many simplifications. The Disciple-inspired improvements suggested above will complicate the original ACH method, but the added complexity will not create any problem if one can use the corresponding components of Disciple-LTA. For example, assessing the believability of some item of evidence could easily be done with Disciple-LTA, as discussed in Section 8.5.

Finally, let us notice that many of the improvements suggested above for ACH may be applicable to any other evidence-based analytic method, such as the use of the substance-blind classification of evidence and the Disciple-LTA methods for assessing the believability of evidence based on its credentials. This suggests that Disciple-LTA may be an excellent tool for teaching intelligence analysts because the concepts and method for evidence-based reasoning that would be learned with it would help the analysts no matter what specific evidence-based analytic methods they would use.

11. CONCLUSIONS

Intelligence analysts face the highly complex task of drawing defensible and persuasive conclusions from masses of evidence of all kinds from a variety of different sources. Arguments, often stunningly complex, requiring both imaginative and critical reasoning, are necessary in order to establish and defend the three major credentials of evidence: its relevance, believability, and inferential force or weight. Additionally, the analysts may be required to answer questions that are of immediate interest and that do not allow time for extensive research and deliberation. Given this complexity, there is a strong emphasis currently placed in the Intelligence Community on developing structured analytic techniques and computer-based tools to assist analysts.[35]

This paper presented Disciple-LTA, an intelligent agent that incorporates a lot of knowledge from the Science of Evidence[36] and uses it in the analysis. Disciple-LTA knows about the substance-blind classification of evidence and about the ingredients of believability assessments for tangible as well as testimonial evidence, knowledge which allows it to develop theoretically-justified argumentation structures for believability assessments. Disciple-LTA supports the development of relevance arguments linking evidence to hypotheses, and it uses a general probabilistic approach to the evaluation of the inferential force of evidence on the considered hypotheses. It also knows how to analyze various types of hypotheses and enforces necessary conditions for sound analysis, such as considering both favoring and disfavoring evidence for each analyzed hypothesis, or qualifying each analytic conclusion with the assumptions made.

Disciple-LTA is also concerned about the many demands placed on analysts and does allow for particular simplification methods. However, these simplifications are not mandated but chosen by the analyst. In particular, Disciple-LTA allows the analyst to drill-down to various levels in the analysis at hand, to make assumptions concerning various verbal assessments of uncertainty, and to revise these assumptions in light of new evidence. And it also alerts the analysts to matters that cannot be overlooked.

But Disciple-LTA has many other (current or under-development) capabilities that have not been presented in this paper. First of all, it is a learning agent that can learn problem solving knowledge directly from an expert analyst, with assistance from a knowledge engineer. This allows Disciple-LTA to continuously improve its knowledge and provide better analytic assistance.

Disciple-LTA can be used to teach intelligence analysts how to perform theoretically-sound evidence-based hypothesis analysis, through a hands-on, learning by doing approach[37] which is much more effective than learning by listening to someone discuss his/her own analyses, or reading papers on these topics. For example, Disciple-LTA can help analysts understand critical concepts, such as types of evidence, relevance, believability and inferential force, and how to use them in constructing arguments in the form of Wigmorean networks. As demonstrated by the analysis of the ACH method in Section 10, mastering these concepts will help the analysts perform better analyses no matter what evidence-based methods they use. This makes Disciple-LTA a particularly useful teaching tool.

Although Disciple-LTA significantly assists the analysts with performing complex evidence-based probabilistic reasoning, it can also be improved along several dimensions. For example, as discussed in Section 10, a good feature of the ACH method is that it shows how individual items of evidence relate to the competing hypotheses. A future module of Disciple-LTA will automatically compare the analyses of competing hypotheses, to reveal differences in the evidence used and the assumptions made, including a focus on areas with less evidential support. Another future module of Disciple-LTA will compare two analyses of the same hypothesis, both generated with Disciple-LTA by two different users. This comparison will reveal differences in the evidence used and assumptions made to uncover cognitive biases. This, for instance, will reveal situations where two analysts disagree with respect to the credibility of a specific item of evidence. Additional future work for improving the Disciple-LTA approach also includes the development of computational models for evidence-based hypothesis generation, for the detection and mitigation of cognitive biases, for deception detection, for collaborative analysis, for evidence monitoring, and for narrative generation at multiple levels of abstraction. And, of course, continuous efforts have to be devoted to developing knowledge bases for a wide range of analytical problems, to simplifying the interfaces of Disciple-LTA, and to facilitating its use by the analysts.

ACKNOWLEDGEMENTS

This material is based on research performed in the Learning Agents Center and was partially sponsored by several U.S. Government organizations, including the Air Force Office of Scientific Research (FA9550-07-1-0268), the National Science Foundation (0750461), and the National Geospatial-Intelligence Agency. The US Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of any agency of the U.S. government. Examples of analysis performed within this article are only examples. They should not be utilized in real-world analytic products as they are based only on very limited and dated open source information. Assumptions made within the analysis are not reflective of the position of any U.S. government entity.

Vu Le, Cristina Boicu, Marcel Barbulescu and other members of the Learning Agents Center have contributed to the development of Disciple-LTA.

We are very grateful to Joan McIntyre, David Luginbuhl, Phillip Hwang, Ben Hamilton, Don Kerr, Susan Durham, Cindy Ayers, William Nolte, Sharon Hamilton, Ben Wible, Kelcy Allwein, George Stemler, Keith Anthony and Jim Homer for their suggestions and support.

REFERENCES

[1] David A. Schum, *Evidence and Inference for the Intelligence Analyst* (2 Vols), University Press of America, MD: Lanham, 1987; David A. Schum, *The Evidential Foundations of Probabilistic Reasoning*, Northwestern University Press, 2001; Cooper, J.R., *Curing Analytic Pathologies,* Center for the Study of Intelligence, Central Intelligence Agency, 2007-2008, https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/curing-analytic-pathologies-pathways-to-improved-intelligence-analysis-1/index.html.

[2] David A. Schum, "Classifying Forms and Combinations of Evidence: Necessary in a Science of Evidence," in *Evidence, Inference and Inquiry,* (The British Academy, Oxford University Press, 2009).

[3] Tecuci, G., *Building Intelligent Agents: An Apprenticeship Multistrategy Learning Theory, Methodology, Tool and Case Studies*, Academic Press, 1998. http://lac.gmu.edu/publications/1998/TecuciG_Building_Intelligent_Agents/default.htm

[4] Gheorghe Tecuci, Mihai Boicu, Cindy Ayers, and David Cammons, ''Personal Cognitive Assistants for Military Intelligence Analysis: Mixed-Initiative Learning, Tutoring, and Problem Solving,'' *Proceedings of the First International Conference on Intelligence Analysis,* McLean, VA, 2–6 May 2005,

https://analysis.mitre.org/proceedings/Final_Papers_Files/86_Camera_Ready_Paper.pdf; Gheorghe Tecuci, Mihai Boicu, Dorin Marcu, Cristina Boicu, Marcel Barbulescu, Cindy Ayers, and David Cammons, ‘‘Cognitive Assistants for Analysts,’’ *Journal of Intelligence Community Research and Development,* 2007, http://lac.gmu.edu/publications/2007/TecuciG_Cognitive_Assistants.pdf; Gheorghe Tecuci, Mihai Boicu, Dorin Marcu, Vu Le, and Cristina Boicu, ‘‘Disciple-LTA: Learning, Tutoring and Analytic Assistance,’’ *Journal of Intelligence Community Research and Development,* 2008, http://lac.gmu.edu/publications/2008/Disciple-LTA08.pdf; and David A. Schum, Gheorghe Tecuci, and Mihai Boicu, “Analyzing Evidence and its Chain of Custody: A Mixed-Initiative Computational Approach,” *International Journal of Intelligence and Counterintelligence,* Vol. 22, No. 2, Summer 2009, pp. 298-319.

[5] John H. Wigmore, *The Science of Judicial Proof* (Boston: Little, Brown, 1937); David A. Schum, *The Evidential Foundations of Probablistic Reasoning*.

[6] Richards J. Heuer, *Psychology of Intelligence Analysis* (Washington, D.C., Center for the Study of Intelligence, Central Intelligence Agency, 1999), pp. 77-88; and Richards J. Heuer, “Computer-Aided Analysis of Competing Hypotheses,” in Roger Z. George and James B. Bruce, eds., *Analyzing Intelligence: Origins, Obstacles, and Innovations* (Georgetown University Press, Washington DC, 2008), pp. 251 - 265.

[7] Charles S. Peirce (1898), *Reasoning and the Logic of Things* (Kenneth L. Ketner, ed., Harvard University Press, Cambridge, MA, 1992); and Charles S. Peirce (1901), “Abduction and Induction,” in Justus Buchler, ed., *Philosophical Writings of Peirce* (Dover, New York, NY, 1955), pp. 150-156.

[8] David A. Schum, Species of abductive reasoning in fact investigation in law, *Cardozo Law Review*, Vol. 22, No. 5-6, 2001, pp 1645–1681.

[9] Nils J. Nilsson, *Problem Solving Methods in Artificial Intelligence* (New York: McGraw-Hill, 1971); Gheorghe Tecuci, DISCIPLE: A Theory, Methodology and System for Learning Expert Knowledge, Thèse de Docteur en Science [Doctoral dissertation] University of Paris-South, France, 1988; and Gheorghe Tecuci, *Building Intelligent Agents: An Apprenticeship Multistrategy Learning Theory, Methodology, Tool and Case Studies* (San Diego: Academic Press, 1998), http://lac.gmu.edu/publications/1998/TecuciG_Building_Intelligent_Agents/default.htm

[10] John H. Wigmore, *The Science of Judicial Proof*.

[11] Toulmin, S. E., *The Uses of Argument* (Cambridge University Press, 1963).

[12] David A. Schum, *Evidence and Inference for the Intelligence Analyst;* and David A. Schum, *The Evidential Foundations of Probablistic Reasoning*.

[13] Charless Weiss, ‘‘Communicating Uncertainty in Intelligence and Other Professions,’’ *International Journal of Intelligence and CounterIntelligence,* Vol. 21, No. 1, Spring 2008, pp 57–85; and Sherman Kent, ‘‘Words of Estimated Probability,’’ in Donald P. Steury, ed., *Sherman Kent and the Board of National Estimates: Collected Essays* (Washington, DC: CIA, Center for the Study of Intelligence, 1994).

[14] Gheorghe Tecuci, Mihai Boicu, Dorin Marcu, Cristina Boicu, Marcel Barbulescu, Cindy Ayers, and David Cammons, ‘‘Cognitive Assistants for Analysts.’’

[15] David A. Schum, “Classifying Forms and Combinations of Evidence: Necessary in a Science of Evidence.”

[16] Gheorghe Tecuci, Mihai Boicu, Dorin Marcu, Vu Le, and Cristina Boicu, ‘‘Disciple-LTA: Learning, Tutoring and Analytic Assistance’’; Gheorghe Tecuci, Mihai Boicu, Dorin Marcu, Cristina Boicu, Marcel Barbulescu, Cindy Ayers, and David Cammons, ‘‘Cognitive Assistants for Analysts;” and Gheorghe Tecuci, Mihai Boicu, Cindy Ayers, and David Cammons, ‘‘Personal Cognitive Assistants for Military Intelligence Analysis: Mixed-Initiative Learning, Tutoring, and Problem Solving.’’

[17] David A. Schum, *The Evidential Foundations of Probablistic Reasoning*.

[18] Gheorghe Tecuci, Mihai Boicu, Dorin Marcu, Vu Le, and Cristina Boicu, ‘‘Disciple-LTA: Learning, Tutoring and Analytic Assistance’’; and David A. Schum, Gheorghe Tecuci, and Mihai Boicu, “Analyzing Evidence and its Chain of Custody: A Mixed-Initiative Computational Approach.”

[19] David A. Schum, *The Evidential Foundations of Probablistic Reasoning*, pp. 114 - 130.

[20] Richard O. Lempert, Samuel R. Gross, and J.S. Liebman, *A Modern Approach to Evidence*, 3rd ed. (St. Paul, MN: West Publishing, 2000), pp. 1146–1148.

[21] David A. Schum, "Knowledge, Probability, and Credibility," *Journal of Behavioral Decision Making*, Vol. 2, 1989, pp. 39-62.

[22] David A. Schum, Gheorghe Tecuci, and Mihai Boicu, "Analyzing Evidence and its Chain of Custody: A Mixed-Initiative Computational Approach."

[23] Ibid.

[24] Richards J. Heuer, *Psychology of Intelligence Analysis* (Washington, D.C., Center for the Study of Intelligence, Central Intelligence Agency, 1999), pp. 77-88;

[25] Richards J. Heuer, "Computer-Aided Analysis of Competing Hypotheses" in Roger Z. George and James B. Bruce, eds., *Analyzing Intelligence: Origins, Obstacles, and Innovations*, pp. 251 - 265.

[26] Richards J. Heuer, "Computer-Aided Analysis of Competing Hypotheses," p. 253.

[27] David A. Schum, "Classifying Forms and Combinations of Evidence: Necessary in a Science of Evidence," in *Evidence, Inference and Inquiry.*

[28] David A. Schum, Gheorghe Tecuci, and Mihai Boicu, "Analyzing Evidence and its Chain of Custody: A Mixed-Initiative Computational Approach."

[29] David A. Schum, and Jon Morris, "Assessing the Competence and Credibility of Human Sources of Intelligence Evidence: Contributions from Law and Probability," *Law, Probability and Risk*. Vol. 6, 2007, pp. 247 – 274; and David A. Schum, Lessons and Stories about Concepts Encountered in Disciple-LTA, (Fairfax, VA: George Mason University, Learning Agents Center, 2008), http://lac.gmu.edu/publications/2007/PRP_GMU_070821_EvidenceConcepts.pdf.

[30] John H. Wigmore, *The Science of Judicial Proof*.

[31] Lance Good, Jeff Shrager, Mark Stefik, Peter Pirolli, Stuart Card, and Richards J. Heuer, *ACH1.1: A Tool for Analyzing Competing Hypotheses*, p. 1, http://www.pherson.org/PDFFiles/ACHTechnicalDescription.pdf

[32] Richards J. Heuer, "Computer-Aided Analysis of Competing Hypotheses," p. 256.

[33] James Bruce, “Making Analysis More Reliable: Why Epistemology Matters to Intelligence,” in Roger Z. George and James B. Bruce, eds., *Analyzing Intelligence: Origins, Obstacles, and Innovations*, pp. 184-185.

[34] David A. Schum, and Jon Morris, “Assessing the Competence and Credibility of Human Sources of Intelligence Evidence: Contributions from Law and Probability”; and David A. Schum, Lessons and Stories about Concepts Encountered in Disciple-LTA.

[35] Jack Davis, “Why Bad Things Happen to Good Analysts,” in Roger Z. George and James B. Bruce, eds., *Analyzing Intelligence: Origins, Obstacles, and Innovations* (Georgetown University Press, Washington DC, 2008), pp. 157 – 170.

[36] David A. Schum, “Classifying Forms and Combinations of Evidence: Necessary in a Science of Evidence,” in *Evidence, Inference and Inquiry;* David A. Schum, “A Science of Evidence: Contributions from Law and Probability,” *Law, Probability and Risk,* in press, 2009.

[37] Gheorghe Tecuci, Mihai Boicu, Dorin Marcu, Vu Le, and Cristina Boicu, “Disciple-LTA: Learning, Tutoring and Analytic Assistance.”