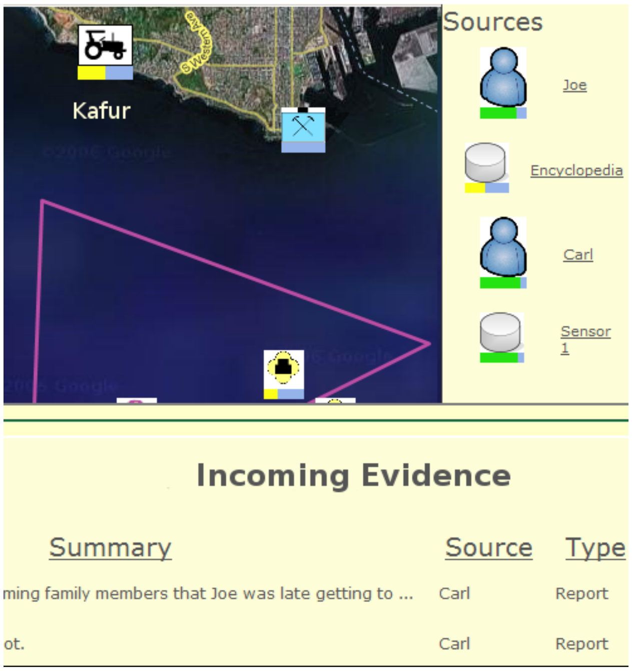
Credibility Models

Charles R. Twardy, Edward J. Wright, Stephen J. Canon & Masami Takikawa[*]
Information Extraction & Transport, Inc.
1911 N. Fort Myer Dr., Suite 600
Arlington, VA 20190
`{ctwardy,ewright,scannon,mtakikawa}@iet.com`

Revision : 1.6

Abstract

We present a general hierarchical Bayesian model where Intelligence Sources make Reports about events or states in the world, which we call Hypotheses. The underlying multi-entity Bayes net for even a simple scenario has hundreds of nodes. We hide the details via Wigmore diagrams and a Google Maps GUI. Our application domain is Intelligence data fusion in asymmetrical warfare (terrorism). Some Hypotheses – like whether a village is a threat – may be abstract or unobservable. For these, we define Indicators – more observable Hypotheses whose value has some bearing on the target Hypothesis. The hierarchy can be arbitrarily deep, and Reports can provide evidence at any level. Furthermore, all Sources have credibility models. Traditional Sources are physical sensors with well-known error models. Non-traditional Sources include humans, websites, news, etc. For these Sources, our credibility models include Hypotheses about unknown factors like objectivity, competence, accuracy, reliability, and veracity. Every Report by a Source provides evidence about those factors. So, for example, successful *ad hominem* attacks against one Source can undermine his assurances that a village is safe, and lead us to believe it is hostile after all.



Incoming Evidence

Summary	Source	Type
ming family members that Joe was late getting to ...	Carl	Report
ot.	Carl	Report

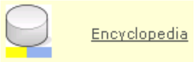
Figure 1: Initial screenshot showing the map window, the Sources dock, and incoming evidence. The hypothetical village of Kafur is shown with the tractor logo, in the upper left. The screenshot has been altered to fit.

1 INTRODUCTION

Our domain is structured evidential reasoning, especially Intelligence analysis. Our task is to reason scientifically about the credibility of Intelligence sources, so that we may properly weigh conflicting evidence for and against various hypotheses. We do that via a complex, hierarchical Bayesian network. However, because Intelligence Analysts do not wish to encounter the full computational complexity of the resulting multi-entity model, we provide simplified views via Wigmore diagrams centered on specific hypotheses. Even so, there are too many of these, so we attach them to specific actors in the world. These actors are presented in a map-based GUI like that shown in Figure 1. In our model, there are three main classes: Hypotheses, Sources, and Reports. When capitalized, these refer to classes or

[*] Now with Cleverset, Inc., Dr. Takikawa can also be reached at `takikawa@cleverset.com`.

objects in our model.[1]

We begin with a use case involving a fictional farming and fishing village named Kafur, a few Intelligence sources, and reports from those sources. This morning, electronic Signals Intelligence (SIGINT) reported that Kafur received a suspiciously large shipment of fertilizer. We are concerned that it may be used for explosives. Figure 1 shows our initial state. The belief bar under Kafur shows its perceived threat level – a default 50% for an unknown place. Our Intelligence encyclopedia reports that Kafur's allegiance is 80% Blue – which means favorable to us. We drag that Report from Incoming Evidence onto Kafur's icon. A dialog box lets us confirm the Source, set the precise Hypothesis to Kafur.allegiance, and the value to "Blue". Encyclopedia has medium credibility , enough to move Kafur's threat level comfortably into the green .

Now we apply this morning's SIGINT Report from "Sensor 1" onto Kafur. For dramatic effect, we consider it direct evidence that Kafur is a threat, so we apply it to Kafur.threat, with the value True. Because Sensor 1 is a reliable source 🟩, Kafur's threat probability moves into the red zone.
We begin talking to Sources.

Joe reports that the fertilizer is in fact going to farmers who live outside the village. This drops the threat probability into the yellow zone, but Joe is a relatively new source. We're waiting to hear from our agent Carl. While we wait, we view the Wigmore[2] diagram showing the lines of evidence for Kafur.threat. (See Figure 2.)

Reports from Joe and SIGINT directly influence Kafur.threat, while the Encyclopedia report contributes via Allegiance. Different sources, in turn, may have different credibility models. For example, the credibility model for SIGINT looks like Figure 3. The sensor attributes TPrate, FPrate, and Reliability (second row) determine probabilities for Boolean report attributes TP, FP, and Working, respectively (third row). These in turn influence the report's status (the bottom red circle).

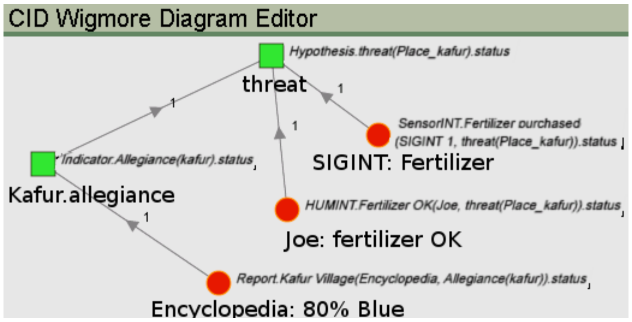

CID Wigmore Diagram Editor
Hypothesis.threat(Place_kafur).status
threat
1
1
1
SensorINT.Fertilizer purchased
(SIGINT 1, threat(Place_kafur)).status
SIGINT: Fertilizer
Indicator.Allegiance(kafur).status
Kafur.allegiance
HUMINT.Fertilizer OK(Joe, threat(Place_kafur)).status
Joe: fertilizer OK
1
Report.Kafur Village(Encyclopedia, Allegiance(kafur)).status
Encyclopedia: 80% Blue

Figure 2: Wigmore diagram showing lines of evidence for Kafur.threat (top), after Joe's HUMINT report. For clarity, we have added short names in large bold, and italicized the unique ID's.

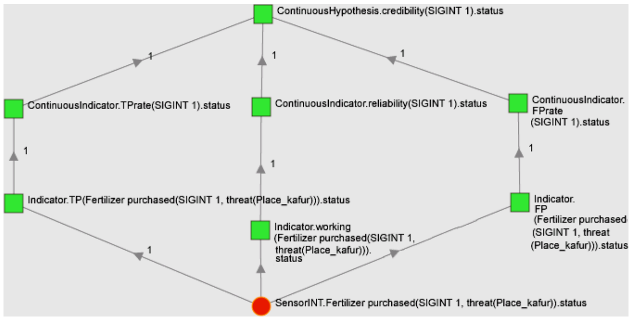

ContinuousHypothesis.credibility(SIGINT 1).status
1
1
1
ContinuousIndicator.TPrate(SIGINT 1).status
ContinuousIndicator.reliability(SIGINT 1).status
ContinuousIndicator.
FPrate
(SIGINT 1).status
1
1
1
Indicator.TP(Fertilizer purchased(SIGINT 1, threat(Place_kafur))).status
Indicator.working
(Fertilizer purchased(SIGINT 1,
threat(Place_kafur))).
status
Indicator.
FP
(Fertilizer purchased
(SIGINT 1, threat
(Place_kafur))).status
1
SensorINT.Fertilizer purchased(SIGINT 1, threat(Place_kafur)).status

Figure 3: Wigmore diagram for SIGINT credibility, after filing one Report. (Some text rearranged)

The credibility models for HUMINT sources (like Joe) track objectivity, competence, accuracy, and veracity, following [Schum, 1997].[3] Joe's initial credibility model is shown in Figures 4 and 5.

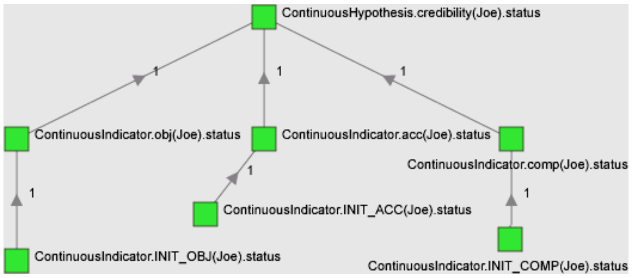

ContinuousHypothesis.credibility(Joe).status
1
1
1
ContinuousIndicator.obj(Joe).status
ContinuousIndicator.acc(Joe).status
ContinuousIndicator.comp(Joe).status
1
1
1
ContinuousIndicator.INIT_ACC(Joe).status
ContinuousIndicator.INIT_OBJ(Joe).status
ContinuousIndicator.INIT_COMP(Joe).status

Figure 4: Wigmore diagram of Joe's initial credibility model. (Some text rearranged)

At this point, Carl reports two things. First, that he has independent reason to think Joe is lying about the fertilizer. Depending on whether we take Carl to mean Joe is lying just on this occasion or habitually, we can apply this to Joe's report (via the tellingTruth) node, or directly to Joe's credibility model (via the veracity node). Here we take the second (more serious) case

[1] We capitalize Intelligence to distinguish the government activity from the human attribute. Attributes appear like so: accuracy.

[2] In a Wigmore diagram, arrows show the direction of inference, from evidence to hypothesis. [Wigmore, 1931] BN arrows are often in the opposite direction.

[3] We apply them concurrently; Schum uses a chain.

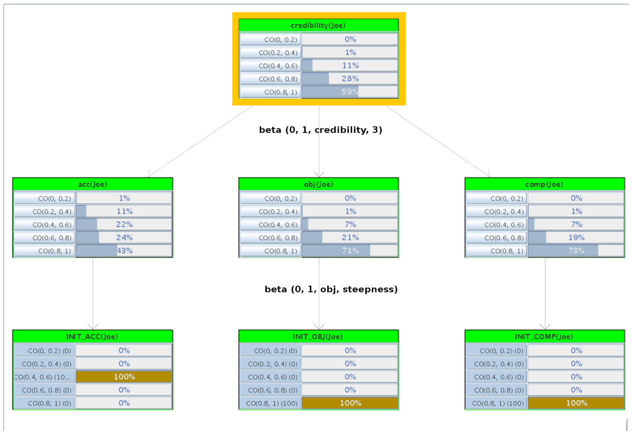


Figure 5: The BN fragment for Joe's initial credibility

– Carl isn't just saying the coin came up tails. He's saying it's weighted to favor tails. Because Carl has a very high credibility, his report casts serious doubt on Joe's report. Consequently, Kafur's threat level increases 🟥🟦.

Carl also says that Joe has been attending several subversive meetings. We might take this as evidence that Joe is a DoubleAgent, but instead we apply it to his objectivity, which affects his credibility. Again, we drag Carl's Report directly onto Joe. By now, Joe's credibility is down to about 50% 🟨🟦. The Wigmore diagram (Figure 6) shows all the evidence influencing Joe's credibility. Kafur's threat level is still in the red zone, though down slightly.

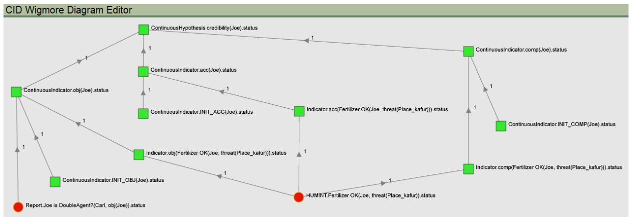

CID Wigmore Diagram Editor
ContinuousHypothesis.credibility(Joe).status
ContinuousIndicator.comp(Joe).status
ContinuousIndicator.acc(Joe).status
ContinuousIndicator.obj(Joe).status
ContinuousIndicator.INIT_ACC(Joe).status
Indicator.acc(Fertilizer OK(Joe, threat(Place_kafur))).status
ContinuousIndicator.INIT_COMP(Joe).status
Indicator.obj(Fertilizer OK(Joe, threat(Place_kafur))).status
Indicator.comp(Fertilizer OK(Joe, threat(Place_kafur))).status
ContinuousIndicator.INIT_OBJ(Joe).status
HUMINT.Fertilizer OK(Joe, threat(Place_kafur)).status
Report.Joe is DoubleAgent?(Carl, obj(Joe)).status
1
1
1
1
1
1
1
1
1
1
1
1
1

Figure 6: Wigmore view of Joe's final credibility model

Compare that Wigmore diagram (Figure 6) to the underlying Bayes net fragment, shown in Figure 7. As we connect reports, the BN becomes complicated. And recall that this is just the fragment dedicated to Joe's credibility, which is not our main concern. The entire BN has several hundred nodes, as shown in Figure 8. As we discuss in Section 2, the model needs all that machinery to "do the right thing", but the Wigmore diagrams provide a much more accessible view of the underlying model. The analyst is seldom concerned with, nor prepared to encounter, the full machinery.

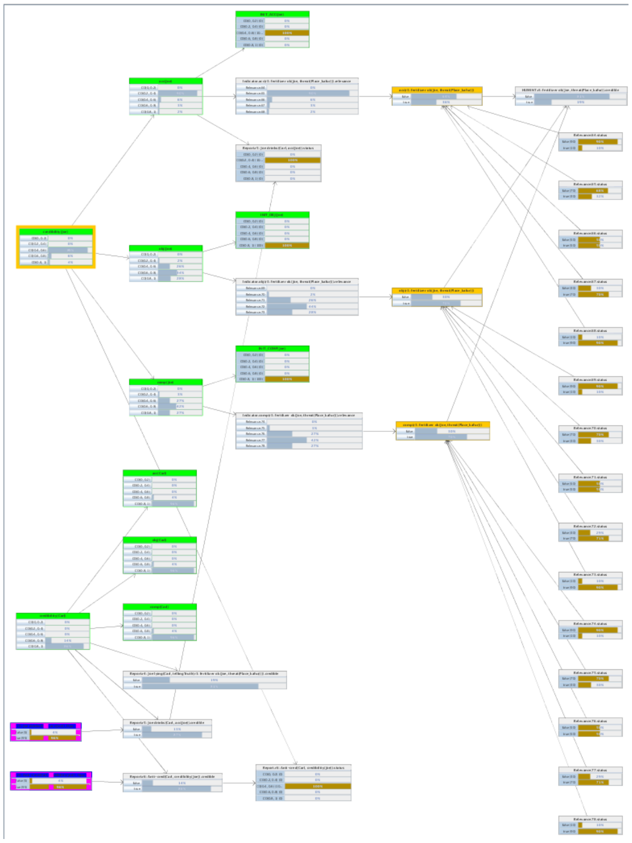


Figure 7: BN fragment for Joe's final credibility model

2 MODEL STRUCTURE

Fundamentally, we have a system for reasoning from observations to hypotheses, accounting for the credibility of sources and the relevance of observations for different hypotheses. It extends and generalizes [Wright and Laskey, 2006] by making report credibilities derive from source credibilities, and by defining general multi-entity Bayesian network (MEBN) fragments[4] that serve for both the scenario and the credibility model. There are three main classes: Hypotheses, Sources, and Reports.

Each class includes several Bayesian network nodes, and references other classes. For example, Reports necessarily have a Source and a target Hypothesis. (These typed links give a MEBN system the expressive power of first-order logic, and thereby support dynamic construction of the network.)

One key to our model's success is that nearly every attribute is a Hypothesis, and therefore nearly every claim can become the center of evidence and reasoning – the top node in a Wigmore diagram. We were especially interested in being able to reason

[4] For MEBNs and fragments, see [Laskey, 2006].

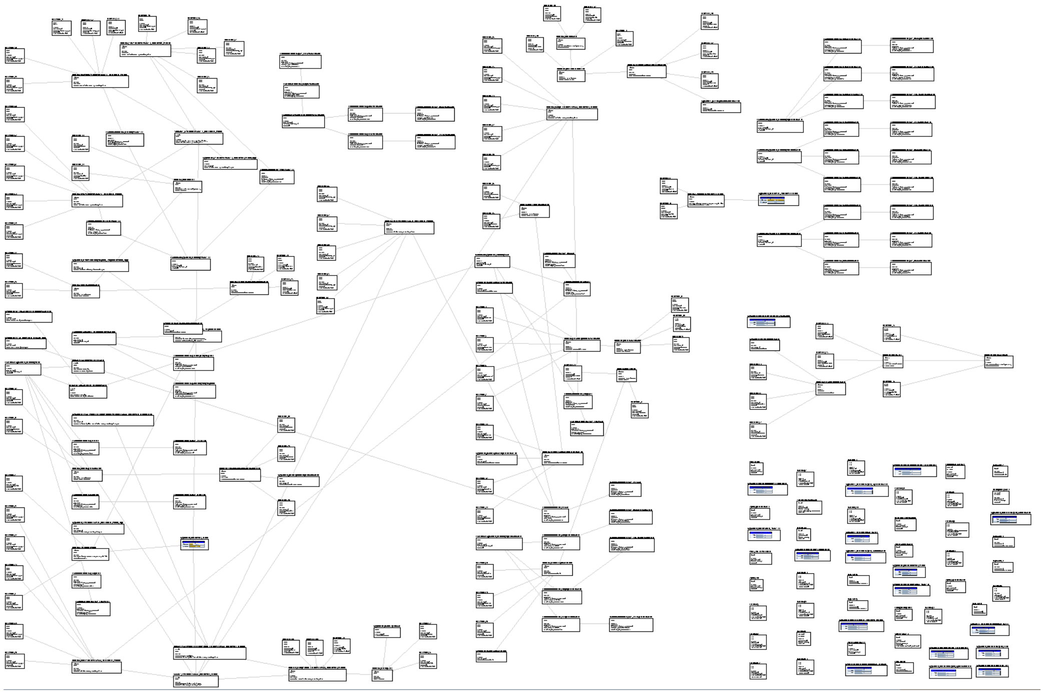


Figure 8: The whole model view: each box is a *frame*, with one or more BN nodes inside. This frame view and the BN views come from our network visualizer.

about the credibility of Reports and Sources. Therefore, key attributes such as Report.opportunity and Source.accuracy are themselves Hypotheses, for which we can define further Indicators, and to which we can apply evidence (i.e. Reports).

2.1 HYPOTHESES

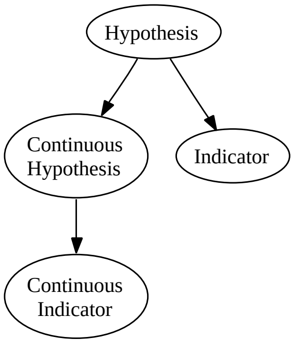


Figure 9: Hypothesis hierarchy

In our scenario, the main hypothesis was whether the village of Kafur posed a threat. In our system, this Hypothesis was one of the predefined attributes of all entities, Entity.threat. Insofar as possible, *every* observable is a Hypothesis, or subclass thereof. Therefore, everything can be observed and reasoned about, including key attributes of our credibility model.

A quick word about our modeling language. The models are defined in in **Quiddity*Script** (**Q*S**), IET's own modeling language for MEBNs. Although **Q*S** has traditional classes, it uses a frame system for MEBN fragments: fragments are "frames" and BN nodes are one kind of "slot" that can inhabit a frame. So for example[5]:

```
frame Hypothesis
  slot status
    domain = Object
    distribution = UniformDiscreteDistribution
```

By this definition, every Hypothesis (and descendant) has a status node. Because status has a distribution, it will become a BN node. For example, Kafur.threat is a default Boolean Hypothesis. Subclasses may redefine status. For example, ContinuousHypothesis sets domain = Continuous and distribution to a function defined in a fn slot.

2.1.1 Relevances

[6] Now, often we cannot apply evidence directly to our main Hypothesis. Instead, we define *Indicators*. For example, Kafur.Allegiance is an Indicator for Ka-

[5] The examples clean up the syntax for presentation.

[6] This section supplies technical detail that may be skipped. It shows how we have made Hypotheses as general as they are.

fur.threat. We might define others, such as poverty level, economic instability, youth bulge, presence of militias or weapons, etc. Similarly, we might argue for the credibility of a sensor by reference to sensitivity, specificity, and reliability. An Indicator need not have the same domain as the Hypothesis it indicates, though it may. Later we will see how we use Indicators in our agent credibility model.

Some indicators are better than others, so we need a measure of strength. We call this the *Relevance*. Then, Indicator I is a function of Hypothesis H and Relevance R: $I = f(H, R)$. For example, if I is normally distributed around H, with a standard deviation given by R, we would say $I = N(H, R)$. Graphically, $H \rightarrow I \leftarrow R$. This could even be $I = H + R$, where $R = N(0, x)$. But unless we need to vary the variance (etc.) on the fly, we do not need actually need to create separate nodes for R (So far, our continuous indicators have not needed to do so.)

However, in the general discrete case, the indicator may well have a different CPT for each hypothesis state. We need to be able to use the proper one depending on our current beliefs about the hypothesis. We also wish to define a single class regardless how many states our Indicator and Hypothesis have. We can do so using *reference uncertainty.* Our Relevance R is actually a set of nodes – one per hypothesis state – defining the the indicator's CPT for that hypothesis state.

```
frame Indicator isa Hypothesis
  slot hypothesis
    domain = Hypothesis
  slot relevance
    domain = Relevance
    parents = [hypothesis.status,
       relevance.hypothesisValue]  # h, r
    distribution = function h, r
       { if h == r then 1 else 0 end }
  slot status
    # domain is inferred
    parents = [relevance.status]
    distribution = function r { r }
```

During inference, the correct $r \in R$ is chosen by the distribution:

```
  function h, r {if h == r then 1 else 0 }
```

In effect, we change the CPT of status on the fly, according to our beliefs about the current hypothesis.

Having defined a structure that can handle such a general case, we next seek to avoid having to fill in all those tables. So we have many special-case constructors which make stronger independence assumptions, and fill in the various R tables for us. In many cases, we can define our relevance with a single number – a notional *strength* for the Indicator, even for discrete CPTs. For example, when H and I have the same domain, we can use a single number to represent $\Pr(I = x | H = x)$, assuming the remainder is spread uniformly among the other states.

2.2 SOURCES

We spend more time on Sources when discussing credibility models in Section 3. Now, it is sufficient to know that all sources have a credibility, which is a ContinuousHypothesis ranging from [0..1], with 1 being perfectly credible. Specific kinds of sources have specific attributes that determine the overall credibility. These attributes are themselves Hypotheses, usually ContinuousHypotheses. All agents have accuracy, objectivity, and competence. In addition, since people can lie, sources of type Person have veracity, which is a function of whether they are a DoubleAgent.

2.3 REPORTS

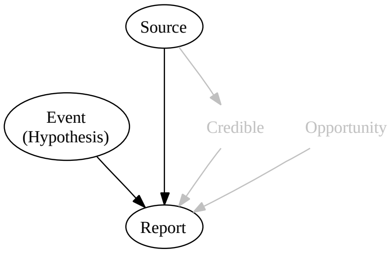


Figure 10: Report schema

Every Report has opportunity and credible Boolean attributes. A generic Report does not know what kind of source it has, so its credible is a direct function of the source's credibility, as well as opportunity:

```
slot credible
  parents = [source.credibility.status,
                  opportunity.status]
  domain = BooleanDomain
  distribution = function cred, opp {
    if opp == false then return [1, 0];
    else {
      c = cred->getMid();
      return [1-c, c];   # [false, true]
    };
    end
  }
```

In contrast, a HUMINT report (see Figure 11) knows that the source is a Person, and defines the additional properties accurate, objective, and competent. Then, credible is a function of all these. (For now we simply AND them together.)

```
slot credible
  parents = [opportunity.status,
     competent.status, objective.status,
     accurate.status]
  domain = BooleanDomain
  distribution = function opp, comp, obj, acc
    { return opp && comp && obj && acc; }
```

So far, we have not considered lying, because our model separates credibility from telling the truth. *Credibility* refers only to the source's ability to *know* the situation. Whether they are *lying* is a separate matter, tracked by HUMINT.tellingTruth, itself determined in part by Source.veracity, which defines their general tendency to tell the truth (to us, anyway).

The status of a Report reflects both credible and tellingTruth. The more credible the report is, the more a lie will mislead. Conversely, a Report with credible=0 has no impact on status, regardless of whether the source is lying: they're simply not in a position to know one way or another.

The Report properties credible and status are *not* Indicators, and so cannot be further observed by Indicators and Reports.[7] Nevertheless, it is possible to provide evidence for constituents like accurate and objective.

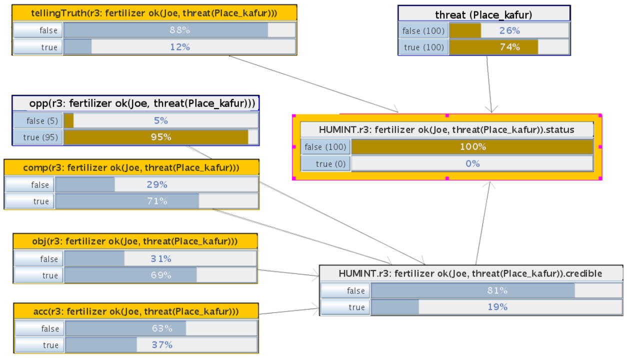

tellingTruth(r3: fertilizer ok(Joe, threat(Place_kafur)))
false
88%
true
12%
threat (Place_kafur)
false (100)
26%
true (100)
74%
opp(r3: fertilizer ok(Joe, threat(Place_kafur)))
false (5)
5%
true (95)
95%
HUMINT.r3: fertilizer ok(Joe, threat(Place_kafur)).status
false (100)
100%
true (0)
0%
comp(r3: fertilizer ok(Joe, threat(Place_kafur)))
false
29%
true
71%
obj(r3: fertilizer ok(Joe, threat(Place_kafur)))
false
31%
true
69%
HUMINT.r3: fertilizer ok(Joe, threat(Place_kafur)).credible
false
81%
true
19%
acc(r3: fertilizer ok(Joe, threat(Place_kafur)))
false
63%
true
37%

Figure 11: Fragment showing the key nodes for Joe's report suggesting that the fertilizer is OK.

Figure 11 shows a fragment of the Bayes net with the key nodes for Joe's report claiming the fertilizer shipment is OK (not a threat). TellingTruth, threat, and credible determine the report's status (right), which is known. Opportunity, competent, objective, and accurate determine credible (lower right). Not shown are Carl's reports disparaging Joe's credibility, nor Joe's intrinsic credibility attributes.

3 CREDIBILITY MODELS

The underlying credibility model tells us how much to believe the claim, when a source makes a report. With electronic sensors reporting on known events in known conditions, we usually have some information on accuracy, false positive rate, and of course, reliability. For human observers, David Schum [Schum, 1997, Schum, 1994] has created a detailed credibility model, which is summarized in Figure 12 by Peter Tillers.[8]

3.1 SCHUM

In short, to be credible, a human report must be: competent, accurate, objective, and truthful. As Figure 12 suggests, Schum has broken each of these down into further observables; our model allows users to add such details, but does not yet require them. Although it is not required, these may perhaps most naturally be thought of as propensities or relative frequencies for accurate, objective, etc. reports. Indeed, in our model, each Report becomes evidence for the credibility of the Source.

3.2 SOURCES

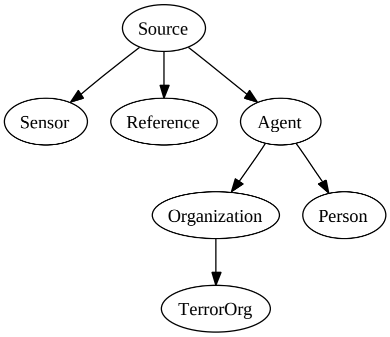


Figure 13: Source hierarchy

Figure 13 shows the basic kinds of sources which we have defined so far. The main classes are Sensors, References, and Agents. Figure 14 shows the credibility attributes defined for the various kinds of sources.

As we saw in Figure 5, the HUMINT credibility model has a Naïve Bayes structure: credibility is treated as an unknown common cause, and the attributes accuracy, objectivity, and competence are assumed to be linked only via the hidden factor, at least until we make reports. The underlying variables are continuous [0..1],

[7] The technical reason is that they are functions of more than one parent, and for now, Indicators indicate precisely one Hypothesis.

[8] The diagram emerged from a discussion among Peter Tillers, Tim van Gelder, and Dan Prager, on the Rationale "Google Group".

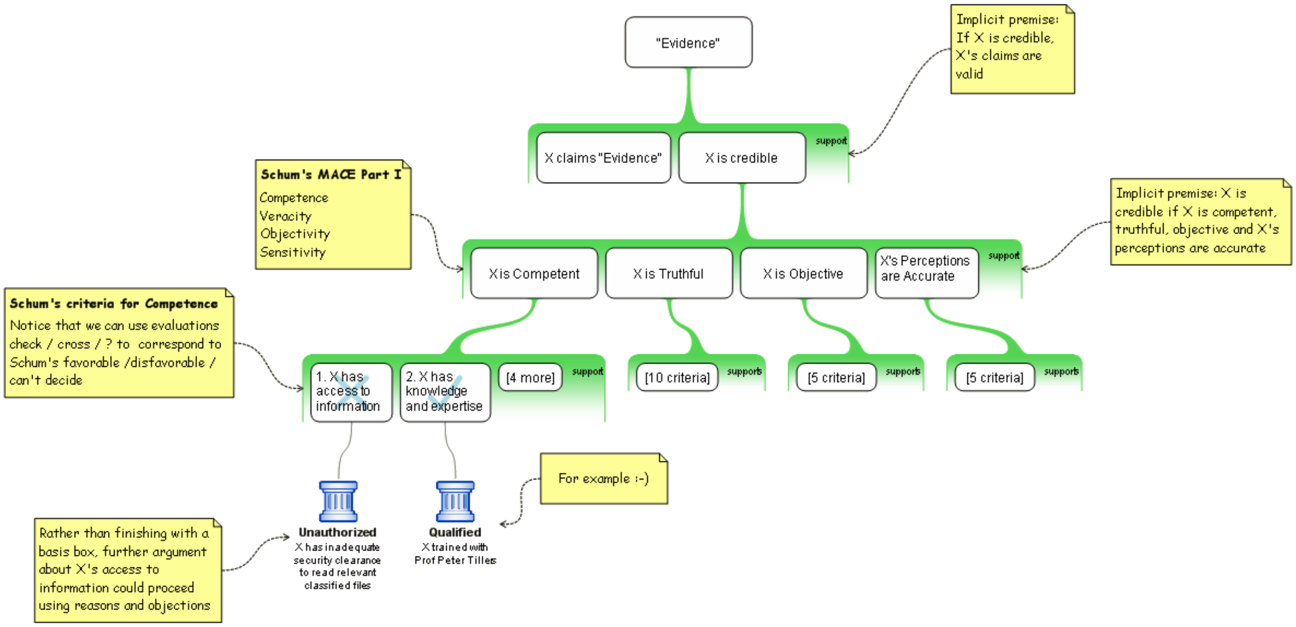

"Evidence"
X claims "Evidence"
X is credible
support
Implicit premise: If X is credible, X's claims are valid
Schum's MACE Part I
Competence
Veracity
Objectivity
Sensitivity
X is Competent
X is Truthful
X is Objective
X's Perceptions are Accurate
support
Implicit premise: X is credible if X is competent, truthful, objective and X's perceptions are accurate
Schum's criteria for Competence
Notice that we can use evaluations check / cross / ? to correspond to Schum's favorable /disfavorable / can't decide
1. X has access to information
2. X has knowledge and expertise
[4 more]
support
[10 criteria]
supports
[5 criteria]
supports
[5 criteria]
supports
Unauthorized
X has inadequate security clearance to read relevant classified files
Qualified
X trained with Prof Peter Tillers
For example :-)
Rather than finishing with a basis box, further argument about X's access to information could proceed using reasons and objections

Figure 12: Schum's credibility model as an argument map. (Peter Tillers)

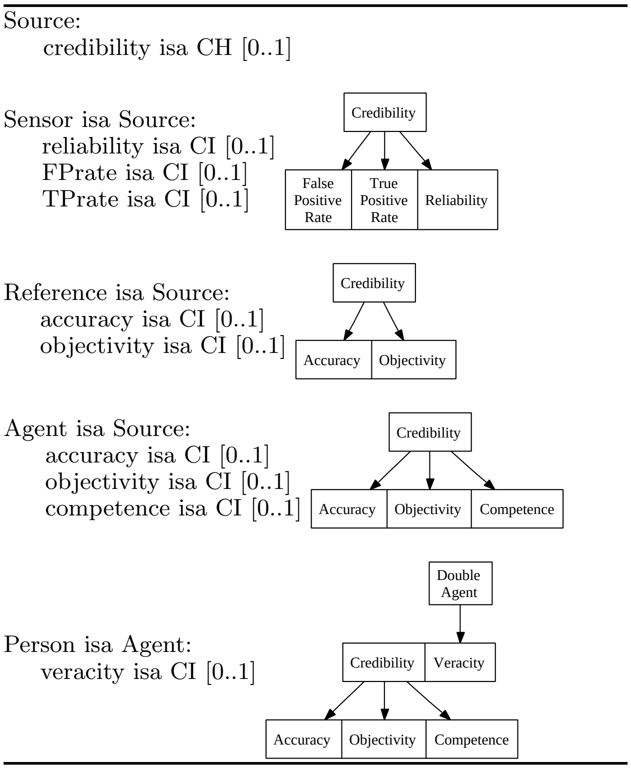


Figure 14: Credibility models for various sources; CH = ContinuousHypothesis; CI = ContinuousIndicator

and each attribute is Beta distributed around credibility.[9]

Using a Beta distribution allows a smooth, flexible distribution bounded on [0..1]. It can flex from U-shaped through flat to quite peaked. For example:

$$\mathsf{acc} = \beta(0, 1, \mathsf{credibility}, 3)$$

where the range is [0..1], the mean is credibility, and the final parameter (here, 3) is, roughly, the steepness of the peak. Because we never observe credibility itself, the other values are correlated.

Neither do we generally observe the attribute values. Doing so would render them insensitive to data in the form of the accuracy (etc.) of reports coming from that source. Instead, each report creates Boolean Indicators of the attribute in question. This is a binomial model: the parameter (say accuracy) has a prior distribution, and each Report gives us a Boolean value. We are, in essence determining the bias of a coin.

When we create a new source, we can make an initial observation of its attributes. As shown in Figure 5, these initial observations are implemented as ContinuousIndicators, themselves Beta distributed around the attribute value. Therefore, credibility begins its life with 3 observations. Every Report will generate more observations, especially as we discover whether

[9] The values are shown discretized. **Quiddity** can do exact inference by discretization, or approximate inference by particle filters. Our visualizer works only with discrete nodes.

the true status of the event in question. Furthermore, we may uncover Indicators that speak directly to these attributes, by performing a background check. For example, if Carl reports that Joe is habitually drunk, we might apply that directly to Joe's accuracy, at the very least!

4 CONCLUSION

We have developed a detailed, hierarchical Bayesian credibility model in the style of David Schum's work. Our low-level model allows very general control of continuous and discrete parameters, with many auxiliary nodes defining arbitrary relevance of indicators to hypotheses, using the relational power of multi-entity Bayesian networks. Then we provide constructors that make various kinds of independence assumptions, for example, allowing one to use a single measure of "strength", or mapping a continuous variable onto a Boolean indicator without further parameters, etc. Next, we hide the Bayes net behind a Wigmore diagram, stripping the view down to the essential flow of evidence among Reports and Hypotheses, hiding all the auxiliary machinery. Credibility models are built on the same Hypothesis—Indicator—Report architecture, and can be inspected and augmented in the same way as the scenario hypotheses. Finally, we subordinate the Wigmore diagrams to a map-based GUI. The prototype uses the Google Maps API to show drag-and-drop functionality from reports to entities on the map, or to sources.

The existing system is only a prototype. The GUI does not yet support all of the features of the underlying probabilistic model, and the Wigmore diagram component is still display-only. Neither can the GUI client get to the BN GUI, as we have yet to package the full visualization component into a web service. Similarly, there is no support for retracting or modifying existing observations. At a more fundamental level, the model itself does not yet make use of the dynamic Bayes net capabilities of the underlying engine, but it will have to: a deployed system would have to "roll up" observations older than some horizon. It may allow a scrolling time window, but at no time would it be able to do inference on all the reports and events over all time.

However, a great deal of work has gone into defining the basic Hypothesis–Source–Report architecture and the top-level HUMINT credibility model in a generic, extensible, and composable manner. The ubiquitous use of Hypotheses is only possible because of the object-oriented (or first-order logic) nature of multi-entity Bayesian networks, and relied on some advanced features of the underlying **Quiddity** engine to perform domain inference on the fly, and allow us to define likelihood observations over runtime-composed domains.

Acknowledgements

This work was funded by Naval Research Laboratory Contract N00173-06-C-4034.

References

[Laskey, 2006] Laskey, K. B. (2006). Mebn: A logic for open-world probabilistic reasoning. Technical Report C4I06-01, George Mason University C4I Center.

[Schum, 1994] Schum, D. A. (1994). *Evidential Foundations of Probabilistic Reasoning.* Wiley & Sons, hardback edition. Paperback 2001, Northwestern University Press.

[Schum, 1997] Schum, D. A. (1997). Pedigree: Credibility design. IET Draft Report v1.0, plus revised Appendix A.

[Wigmore, 1931] Wigmore, J. H. (1931). *The Principles of Judicial Proof, or the Process of Proof as given by Logic, Psychology, and General Experience and Illustrated in Judicial Trials.* Little Brown & Co., second edition. There are hardcovers in print of either the 1913, 1931, or 1937 editions.

[Wright and Laskey, 2006] Wright, E. J. and Laskey, K. B. (2006). Credibility models for multi-source fusion. In *Proceedings of the 9th International Conference on Information Fusion*, Florence, Italy.