BACH (Bayesian Analysis of Competing Hypotheses)
A Robust Decision Support and Analytical Tool
(Original March 2010, revised April 2015)
Dr. David G. Ullman
ullman@davidullman.com

In CIA analyst Richard Heuer's classic work, "The Psychology of Intelligence Analysis," he identified two primary issues that will undermine intelligence analysts' assessments if not accounted for and addressed: the need to account for uncertainty and the potential suppression of competing analytical frameworks.[1] Heuer attempted to instill in CIA analysts the idea that they needed to be confident in their assessments. They needed to know not only the assessments, but also how they arrived at them. The way to become confident was to know both that they had rejected competing analyses only after a rigorous evaluation of the evidence and that they could accurately estimate the composite level of uncertainty underlying the assessment.

These are not new issues, nor unrecognized ones, but they continue to bedevil the intelligence community. One of the key criticisms of the 2005 Weapons of Mass Destruction Commission, to give but one example, was the intelligence community's failure to identify and sustain alternative assessments and its inability to adequately address the uncertainty underlying the evidence in all of its assessments. The WMD Commission specifically highlighted that the inability to understand, track, and carry forward uncertainty "results in false impressions of certainty for analysts' ultimate judgment."[2]

To cope with these needs, Heuer developed a methodology, the Analysis of Competing Hypotheses (ACH), to help analysts guide the collection of evidence, maintain an awareness of uncertainty, and then help one select the most likely hypothesis. This paper examines Heuer's methodology and extends it through a logical model by which the authors have developed a supporting software capability.

Heuer's Analysis of Competing Hypotheses methodology essentially consists of five steps. These steps are:

    a. Identify the possible hypotheses to be considered,
    b. List the significant evidence and assumptions for and against each hypothesis,
    c. Draw tentative conclusions about the relative likelihood of each hypothesis,
    d. Analyze sensitivity of the conclusion to critical items of evidence, and
    e. Identify future observations that would confirm one of the hypotheses or eliminate others.

After executing these steps and selecting one preferred hypothesis, Heuer's method has analysts consider the uncertainty, in the evidence, in their analysis of the evidence, and in the final assessments. Heuer found this task to be difficult and, after great study into the cognitive sciences, postulated three primary challenges to analyzing evidence. They are:

---

[1] Most recently consolidated and updated in Heuer, Richard J., Psychology of Intelligence Analysis (New York: Novinka Books, 2006).

[2] Report to the President of the United States, The Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction, March 31, 2005, http://www.fas.org/irp/offdocs/wmd_report.pdf

a. The human mind is limited, it is poorly "wired," to deal effectively with uncertainty;
b. Even increased awareness of this cognitive limitation does little to help analysts effectively address uncertainty; and
c. Tools and techniques which train analysts' minds to apply higher levels of critical thinking can substantially improve analysis on complex issues, particularly those issues on which information is incomplete, ambiguous, and potentially deliberately distorted.

Unfortunately, while Heuer's work provides a strong strategic foundation, it has its limits. The Analysis of Competing Hypothesis methodology is not sufficient in itself to support the analysis of uncertain evidence, the fusion of all-source evidence, and to give clear guidance as to where one needs to focus one's analysis. And, while recognizing the limitations on the human mind's ability to deal effectively, and carry forward uncertainty throughout the analytical process, Heuer offers little for analysts and managers to do to overcome these challenges.

Modern computer-based and statistical analysis, however, can be employed to address these challenges. Our effort has combined proprietary Bayesian[3] methods with ACH, resulting in BACH (Bayesian Analysis of Competing Hypotheses), a robust analytical capability that has the following goals:
a. Support individual analysts and teams of analysts as they determine the most likely hypotheses and the risks associated with them;
b. Support the management of uncertain, temporal, incomplete and conflicting evidence;
c. Be easy to use and useful for analysts and team members;
d. Support building a shared understanding for an all-source team;
e. Flexibly support the addition of new evidence regardless of source, certainty or fidelity;
f. Help the enterprise determine what actions to take next to ensure the best hypothesis has been identified;
g. Support decision flow up and down the decision-making levels.;
h. Maintain the provenance of decisions so the rationale can be queried, vetted and studied;
i. Support sensitivity and contribution analysis.

The remainder of this paper outlines the underlying model used in BACH, develops details through a simple example that shows how these goals are being met, and concludes with progress made to date.

BACH is a refinement of traditional assessment based decision making methods commonly used in business, engineering, and science. These methods facilitate the evaluation of discrete alternatives relative to a set of measurable criteria. The goal during the

---

[3] Bayesian decision theory is a well-founded computational theory for applying general knowledge to individual situations characterized by uncertainty and risk. The Bayesian methods used here are a refinement of those developed and commercialized over the last ten years to support decision-making in product development and other business activities.

evaluation is to determine how well each alternative satisfies the goal (either stated or unstated) defined by each measure (i.e., the criterion). The overall satisfaction for a given alternative is usually the weighted sum of the individual criterion satisfactions. The alternative with the highest overall satisfaction thus becomes the best possible choice.

Applications based on traditional assessment based decision making methods go by many names, such as Multi-Attribute Utility Theory, but the basic structure is straightforward and can be shown as a table with the alternatives heading the columns and the criteria in the rows. The example at Figure 1 is based on the selection of a project. Here, there are four alternative projects with four criteria used to select the "best" project. Estimates for the value of the factors may be derived from any method, from detailed analysis to opinion. In this case, we selected a simple worse than average, average, and better than average (-1, 0, 1) valuation. The satisfaction level of Project Gamma, the highest of the four, is 80, implying that there is an 80% satisfaction in meeting the goals described by the criteria.

		Alternatives							
		Alpha		Beta		Gamma		Delta	
Criteria	Weight	Factor	Score	Factor	Score	Factor	Score	Factor	Score
Low Program Complexity	35	0	0	0	0	1	35	0	0
High Market Maturity	10	0	0	0	0	-1	-10	-1	-10
Strong 5-yr Cash Flow	20	1	20	1	20	1	20	1	20
Short Time to Start-up	35	1	35	1	35	1	35	1	35
	Total:		55		55		80		45

Figure 1: An Example Decision Matrix

In this example, there is a separate evaluation for each alternative/criterion pair, i.e., for each cell in the matrix. The five-year cash flow evaluation for Alpha is independent of the evaluation for Gamma and may have been estimated using a totally different method. The challenge of reducing these widely varying evaluation items to a common basis so that the satisfaction result has meaning will be discussed later in the paper.

To enable decision making techniques applicable to intelligence analysis based on evolving evidence and to support Heuer's approach, assessment methods of decision support need to be modified. As with assessment methods of decision making described above, evidence based decision making (EBDM) focuses on the resolution of a specific issue or question. In EBDM, the alternatives are a list of hypothetical explanations for the state of the situation or the courses of action. Thus, from here on, alternatives will be referred to as hypotheses, consistent with ACH. The goal is to support or deny each hypothesis with the evidence that is collected, and to determine what additional evidence may be required to refine the decision.

In EBDM, *evidence is any data or information, whether solid or weak, obtained through experience, direct observation or experimental work that can help resolve the issue*. But evidence is not the same as criteria. Evidence contributes to knowledge about some or all of the hypotheses rather than about a single alternative. Evidence does this through adding information to indicators as shown in Figure 2. *Indicators are the classes of collectable*

information that help discriminate amongst the hypotheses. An indicator is a measure of a specific attribute of the hypotheses and is generally the identification, state or activity of an object, organization, occurrence, or location. As each new piece of evidence is collected, it contributes knowledge to one or more indicators which in turn supports or denies some or all of the hypotheses in one or more issue as shown in the figure.
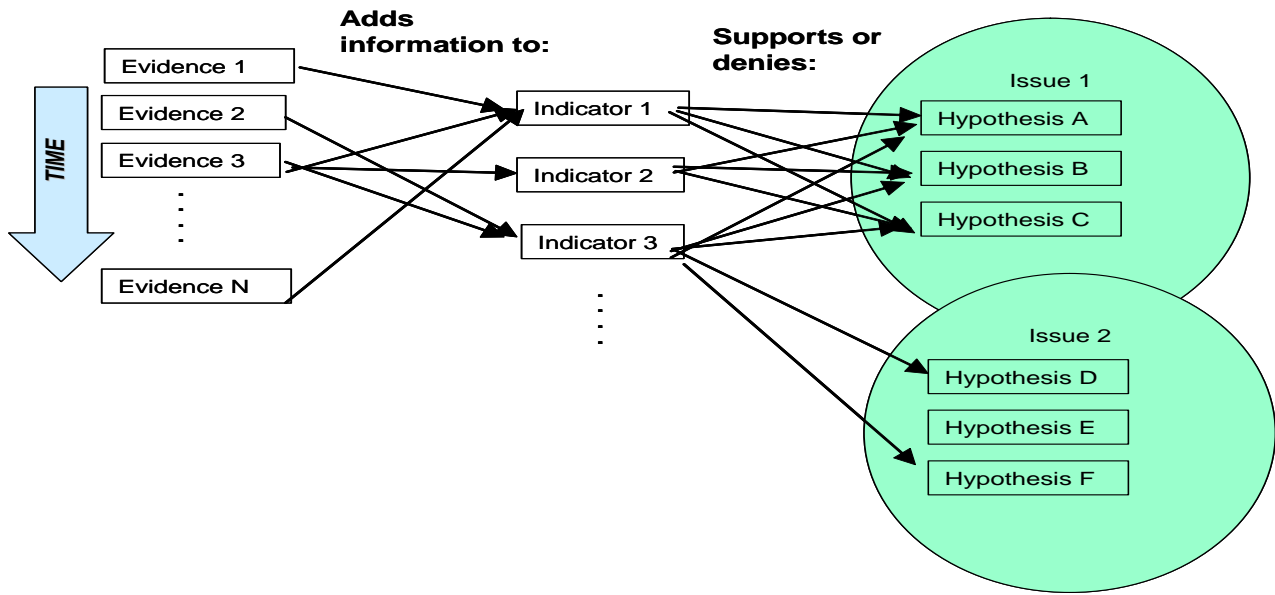

Figure 2: Evidence Influence on Hypotheses

In traditional assessment based decision making, a single piece of evidence is used to confirm or deny a single alternative. Where EBDM differs from this traditional methodology is that a single piece of evidence may contribute to the confirmation or denial of many hypotheses. This can be seen in Figure 3 where the multi-headed arrows show how the evaluation of a single piece of evidence contributes to knowledge about an indicator and thus, potentially, to knowledge about each of the hypotheses.
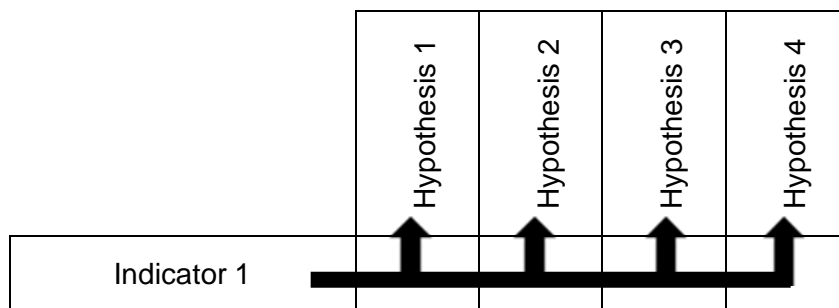

Figure 3: Evidence Based Decision Matrix

The implementation of an EBDM system requires two steps. First, the analyst must identify the historical or anticipated *expected behaviors* that support or deny the hypothesis. Second, the analyst must analyze each piece of evidence for its contribution to its associated indicators. As evidence is collected and analyzed, its support to or denial of each hypothesis

will be reflected through the expected behaviors. Thus, the assessment of the likelihood of any given hypothesis will evolve as the analyst evaluates each individual piece of evidence. As more evidence accumulates and expected behaviors mature, the analyst may update or create new hypotheses and, if warranted by the evidence, reject old hypotheses as not worthy of further consideration.

HOW BACH MAKES EBDM POSSIBLE

While the basics of EBDM are refinements of the Analysis of Competing Hypotheses and common assessment-based decision making, BACH integrates more sophisticated tools to develop a system that supports the logic, the uncertainty, and the evidentiary fusion of intelligence analysis. To understand our approach, reconsider Figure 1, a simple decision matrix. The scores in the factor cells reflect how well each alternative meets each criterion. Other scoring factors and methods can be used, but they all have the same problem in that they have no actual meaning - the results are not a rigorous measure of "satisfaction," as the factors have too wide a range of meaning. This limitation is often overcome by performing the evaluations as probabilities (e.g., Project Alpha has a 75% chance of having high market maturity). This gives mathematical meaning to the decision matrix and helps represent uncertainty, but probabilities have several other problems and it is well known that people have difficulties in consistently determining and applying probabilities. We have recognized these limitations and have worked to overcome them in BACH through the use of Bayesian decision methods.

Bayesian decision theory and methods are academically well founded and widely applied in commercial industry (e.g., most major e-mail spam tools use Bayesian methods, as do all speech recognition tools and, increasingly, medical diagnoses). Bayesian methods excel in situations where the available information is imprecise, incomplete, perhaps inconsistent, and in which outcomes may be uncertain and decision makers' attitudes toward them vary widely. Bayesian decision analysis can indicate not only the best alternative to pursue, but also whether a problem is yet ripe for deciding and, if not, how to proceed to reach the decision.

There have been previous efforts to apply Bayesian methods to ACH.[4] We have overcome the limitations experienced in these efforts by applying recently developed methodologies that require virtually no prior knowledge, can fuse inconsistent and contradictory information, and allows for multiple analysts. Current commercially marketed tools focus on assessment based decision making. BACH, by contrast, refines and extends these metrologies to achieve a true evidence-based decision making capability.

BACH EXAMPLE

To demonstrate the EBDM approach and the logic behind BACH, consider a simple example: a team needs to identify the most likely explanation for the actions of a terrorist group. These below steps follow those of ACH with modification. Even though the operation is presented sequentially, in reality (as more evidence is collected) an analyst can continually update and refine the problem.

---

[4] These efforts have relied upon so-called "Bayes Nets." While Bayes Nets are very good at modeling complex situations driven by uncertain information, they cannot fuse evaluations of multiple analysts, they require extensive model building, and one must estimate prior possibilities.

Step 1: Identify the possible hypotheses. For this example the hypotheses are:

Hypothesis 1: The terrorists are doing business as usual.
Hypothesis 2: The terrorists are planning an attack in the next two weeks.
Hypothesis 3: The terrorists are planning an attack in greater than two weeks.

In any ACH problem, the hypotheses must be mutually exclusive and are best if complete, although completeness is not necessary. In this case they are both independent and complete.

Step 2: Determine the indicators. As mentioned above, indicators are the classes of collectable information that help discriminate amongst the hypotheses; they are a measure of a specific attribute that can either support or deny the hypotheses. *Indicators are typically the identification, state or activity of an object, organization, occurrence, or location*. In this case, the team decided that there are three indicators:
Indicator 1: Level of activities.
Indicator 2: Presence of inflammatory rhetoric.
Indicator 3: Personnel count.

An additional part of this step is capturing the relative importance of the indicators. Importance may be inconsistent across analysts or organizations and will become part of the sensitivity analysis in Step 6. In this example there are two viewpoints to consider: one that emphasizes activity and downplays the importance of rhetoric and the other more balanced as shown in Figure 4, a graphical interface for capturing importance.

Criteria	Weight		Criteria	Weight	
① Level of activities		0.59	① Level of activities		0.32
② Presence of rhetoric		0.10	② Presence of rhetoric		0.46
③ Personnel count		0.31	③ Personnel count		0.23

Figure 4: Importance of indicators

Step 3: Capture the expected behavior. The expected behavior is a representation of how the indicators have historically or are anticipated to add support or denial to the hypotheses[5]. Each of the indicators in this example gives an instance of a different type of expected behavior. A strength of using BACH is that it can manage information that is evaluated at discrete normality levels (as in the case of Indicator 1 above), as a Yes/No choice (Indicator 2), or as a measured value (Indicator 3).

---

[5] Expected behavior is a form of prior knowledge, but by focusing on discreet ranges of behavior, it is simpler for the analyst to provide the range of expected behavior.

For Indicator 1, the *Level of activities* is measured as having one of three levels: normal, high, or very high. The expected behavior relative to this indicator is:

- If <Level of activities> is <Normal>, we expect <They are doing business as usual (Hypothesis 1)>.

- If <Level of activities> is <High or Very High>, we expect <They are planning an attack in the next two weeks (Hypothesis 2) or they are planning an attack in more than two weeks (Hypothesis 3) >.

For Indicator 2, the *Presence of inflammatory rhetoric*, the expected behavior is similar but it is measured by a Yes or No choice.

- If <Presence of inflammatory rhetoric> is <Yes>, we expect <They are planning an attack in the next two weeks (Hypothesis 2) or they are planning an attack in more than two weeks (Hypothesis 3) >.

- If <Presence of inflammatory rhetoric> is <No>, we expect <They are doing business as usual (Hypothesis 1)>.

Indicator 3, *Personnel count,* is measured by actual data and so the expected behavior statements are of the form:

- If <Personnel count> is between 30 and 40, we definitely expect <They are doing business as usual>.

- If <Personnel count> is between 20 and 50, we possibly expect <They are doing business as usual >.

- If <Personnel count> is between 40 and 60, we expect <They are planning an attack in more than two weeks>.

Each of these forms is modeled in BACH through a simple graphical user interface.

Step 4: Collect evidence, assumptions and how each piece contributes to the indicators.
The first three steps fully define the issue and now collected evidence can be added to the system. The following is a sample, in order, of collected evidence.

Evidence item 1: A UAV image at NIIRS 5 of group of people.
Analyst A interprets image as Personnel Count = 25 ± 5.
Analyst B interprets image as Personnel Count = 35, may be as low as 30 or as high as 45.

Evidence item 2: A new speech by the Divine Leader is received and translated.
Analyst C: Rhetoric in speech slightly more vitriolic than usual but has never heard a speech by the Divine Leader in this context before.
Analyst D: Rhetoric seems inflammatory but something else unknown was happening.

Evidence item 3: HUMINT reports 40-50 people are moving boxes.
Analyst B: Updates Personnel Count to 35-50 based this new evidence.
Analyst B: Interprets evidence as indicating high activity, but is unsure of this analysis.

As can be appreciated, regardless of form, each piece of evidence is:

a. Uncertain (e.g., 25 – 30 people, "may be high," "unsure"). All evidence has some level of uncertainty which may be stated or unstated.
b. May be changing with time (The Personnel Count was updated by Analyst B based on new HUMINT information).
c. May conflict with other evidence (Evidence in item 3 does not agree with that in item 1).
d. May be interpreted differently by different people (The first two evidence items are not evaluated the same by the analysts).
e. May fade with time (Analyst B updates the Personnel Count estimate based on the new evidence taking into account the age of Evidence 1) All evidence decays, some faster than others.

Figure 5: Analyst B - Evidence 1 input

One of the strengths of BACH is its use of graphical interfaces to aid in entering and interpreting information. In Figure 5, the uncertain evaluation is represented on a number line that allows capture of the estimated personnel count. Analyst B in this example would simply move the high, low and most likely icons to input her analysis and then type in her rationale for this evaluation.
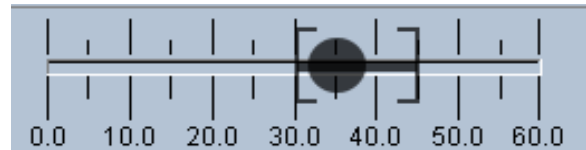
In Figure 6, the analysis of evidence item 2 is made on a Belief Map. Belief Maps are used to input qualitative evaluations. Here Analyst D believes the rhetoric is inflammatory (dot near top) but is not fully certain of this (dot not fully to the right).   Analyst C, evaluating independently of Analyst D, thought the rhetoric only slightly different than usual (dot slightly above the mid-point) but is unsure of this analysis (dot to the left). BACH fuses these two analyses to form a team assessment. This is not the average, but a true fusion where the whole is greater than the sum of the parts.
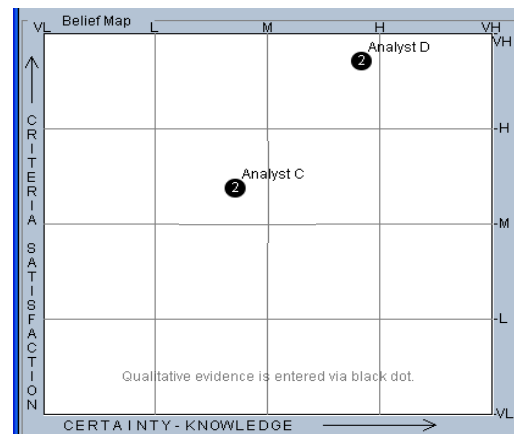

Figure 6: Evidence Item #2 input

Step 5: Draw tentative conclusions about the relative likelihood of each hypothesis.

The primary result of BACH analysis is the identification of the most likely hypothesis. Two other primary statistics are simultaneously developed, probability that each hypothesis is best and the risk that the evaluation is not correct. These statistics can be plotted over time to see the emergence of one hypothesis over the others. For the current example, the satisfaction of the three hypotheses is shown in Figure 7. The three bars on the left show the satisfaction for the three hypotheses with the importance weightings that emphasize the importance of the Level of activities (as defined in Step 2 (see Figure 4)). Those on the right show results for the more even weighting. These both use all the evidence evaluation as input in Step 4; just the relative importance is changed. As shown in the figure, from one viewpoint, it is business as usual (the first bar) and from the other they are planning to attack in the next two weeks (the middle bar). Neither case is the results overwhelmingly conclusive and thus the importance of Step 6.
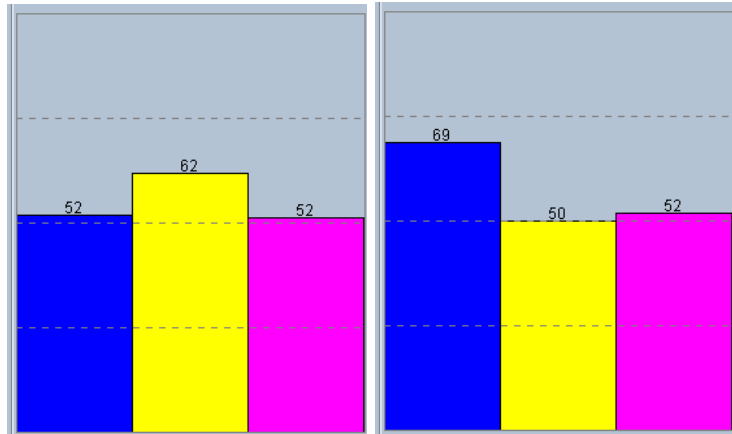

Figure 7: Satisfaction results from two viewpoints

Step 6: Analyze what to do next and the sensitivity of the results.

As important as the tentative conclusions are, the knowledge about what to do next and the ability to explore the sensitivity of information is of even greater value. BACH computes a what-to-do-next report and provides a dashboard to help managers answer the following questions:

- What is the **evidence contribution** to the results? If one piece of evidence is excluded, what is the effect on the results?
- What is the analyst contribution? If the input of one analyst is excluded, what is the effect on the results?
- What was the **inference chain** that led to the current recommendation?
- Do we have community **buy-in** to the process and the result?

BACH provides tools to explore these questions. A Bayesian method called a Value of Information (VOI) analysis provides a basis for recommending what to do next to help differentiate the hypotheses. It can help the team find where there is disagreement that is worth resolving (and where not to bother discussing) and where more evidence is needed (see Step 7).

Step 7: Identify future observations that could confirm one of the hypotheses or eliminate others.

The feedback in Step 6 identifies what information has the most value in resolving the issue. Using this information, decision managers can direct what additional information to collect and what new technologies should be developed. And, just as importantly the above analyses can help determine what type of evidence is not particularly helpful in confirming or denying a hypothesis. This allows analytical leaders to determine where *not* to expend resources in the attempt to gather additional evidence.

CONCLUSION
The merger of Bayesian computational algorithms with the intelligence community's Analysis of Competing Hypotheses methods has the potential to add significant rigor to the community's collective analytical work. By allowing analysts to capture and sustain uncertainty as evidence builds and analytical assessments emerge, the analyst achieves an increased confidence in his assessment, knowing that he has considered alternative hypotheses and only rejected those evidentially shown to be less likely, while still retaining a healthy appreciation for the uncertainty in any "final" analysis. The analyst may also then perform sensitivity analysis on the results so that he can best apply resources for further refinement of indicators and hypotheses.   BACH methodology represents a major step forward toward meeting the Director of National Intelligence's goal of improved intelligence analysis. In BACH, the authors have an initially refined prototype capability that has automated this process and provides the user with a straight-forward, collaborative computer-based tool to achieve this goal.