NORTH ATLANTIC TREATY
ORGANIZATION

SCIENCE AND TECHNOLOGY
ORGANIZATION

AC/323()

www.sto.nato.int

STO TECHNICAL REPORT

PUB REF STO-MP-SAS-114-PPG

ANNEX G
Report on SAS-114 Experiment on Analysis of Competing Hypotheses

Mandeep K. Dhami, Ian Belton

A Test of the 'Analysis of Competing Hypotheses' Analytic Technique

Mandeep K. Dhami, Ian Belton
Middlesex University, London, UK
and
David R. Mandel, DRDC-Toronto

Acknowledgments

Dhami and Belton's contribution funded by HM Government

Thanks to JB and SH

Background

- Analysis of Competing Hypotheses (ACH; Heuer, 1999) is a technique used in intelligence community to
  - Identify alternative hypotheses
  - Link evidence to hypotheses
  - Avoid confirmation bias (searching for evidence to support a favoured hyp or giving more weight to such evidence)

- ACH therefore focuses on evidence *inconsistent* with a hypothesis, ignoring evidence *consistent* with it

- ACH consists of eight steps

1. Identify hypotheses
2. List significant evidence
3. Create the ACH matrix
4. Revise the ACH matrix
5. Draw tentative conclusions
6. Perform sensitivity analysis
7. Report conclusions
8. Identify indicators for future observation

- Heuer (2005)
  - incorporates reference to accompanying ACH software
  - reference in step 3 to the ‘diagnosticity’ of evidence is replaced by the concepts of evidence ‘credibility’ and ‘relevance’
  - step 4 includes a requirement to identify gaps in the evidence that may need to be filled
  - step 5 requires the analyst to compare their conclusions with ‘inconsistency’ or ‘weighted inconsistency’ scores generated by the ACH software

- Heuer (2009, p. 4) “I’d love to see our methods tested...”
  - BUT, he believes this is not practical!
  - AND even if the research is done analysts will ignore it!

- Little past research on ACH
  - Wheaton & Chido (2006) – automation, structured ACH
  - Lehner et al. (2008) found some evidence that ACH reduces confirmation bias in non-analysts but no evidence that it reduces bias in experienced analysts
  - Kretz et al. (2012, 2013) found that ACH did not do that much better than other techniques in e.g., generating more hypotheses
  - Trent et al. (2007) were unable to evaluate the impact of ACH on analysis as, despite being trained on the technique, groups resisted using it

Aims

- To examine how well analysts trained to use ACH can actually apply it in practice

- To explore the hypothesis-testing strategies that untrained analysts use

Method

- ***Design***
  - Between subjects
    - ACH training v. no training (control)

- ***Participants***
  - 50 intelligence analysts (25 trained in ACH and 25 given no training)
  - 50% male
  - Average age = 28 (*SD* = 5)
  - Months' experience as analysts ranged from 1 to 144)
  - 78% worked as full-time analysts

- ***ACH Training***

Step 1	**Identify all possible hypotheses. These should be mutually exclusive.**
Step 2	Make a list of significant information/evidence that is relevant for evaluating the hypotheses, including assumptions and the absence of things one might expect if the hypothesis were true.
Step 3	Create a matrix with all the hypotheses across the top and all items of relevant information down the left side. Then, analyse each piece of information by asking if it is Consistent or Inconsistent with the hypothesis or if it is Not Applicable or irrelevant. This can be done by filling each cell of the matrix row-by-row with 'C', 'I' or 'NA'. You can put two 'CCs' or two 'IIs' if the information is particularly compelling. The ratings will likely depend on some assumptions, and if so, then record those assumptions in another column, row-by-row.
Step 4	Think about how the matrix may need revising. To do this, sort the information for diagnosticity (i.e., which items of information are most helpful in comparing hypotheses). Consider how much confidence you have in the assumptions for the highly diagnositic Inconsistent ratings, and re-adjust the ratings accordingly. Delete the rows with non-diagnostic information. Reconsider the hypotheses and decide if any need combining or any if new ones need to be added. Finally, rate the information for the combined or new hypotheses, again making note of any assumptions. You will need to redraw and update the matrix.

Step 5	**Draw tentative conclusions about the relative likelihood of each hypothesis based on the diagnosticity of each item of information. Do this by adding up the number of Inconsistent ratings for each hypothesis to give an 'Inconsistency Score' for each hypothesis. Then, rank the hypotheses so that the highest rank is given to the one with the lowest inconsistency score. The hypothesis with the lowest inconsistency score is tentatively the most likely hypothesis and the hypothesis with the highest inconsistency score is usually the least likely.**
Step 6	Analyse the sensitivity of your tentative conclusion to a change in the interpretation of a few critical items of relevant information. If one or more of these items were wrong, misleading or subject to a different interpretation will your conclusion need to change? If so, then go back and double-check the accuracy of your interpretation.
Step 7	Report your conclusions. Consider the relative likelihood of all of the hypotheses. State which items of information were the most diagnostic, and how compelling a case they make in identifying the most likely hypothesis. Also say why alternative hypotheses were rejected.
Step 8	Identify indicators or milestones for future observation. Create two lists – one focusing on future events or access to additional information that would support your conclusion, and one list focusing on events and information that would suggest your conclusion is less likely to be correct or that the situation has changed.

- ***Analytic test task***
  - 4 specific hypotheses
  - 2 general hypotheses
  - Base-rate information provided
  - 12 pieces of evidence
  - Probability of occurrence of each piece of evidence
  - Target described in terms of 12 pieces of evidence

Abbreviated Instructions

- In this task, you will be asked to assess the tribe membership of a randomly selected person from a region…. After reading the scenario, you will be asked to detail your analysis [using an Analysis of Competing Hypotheses exercise]….

Abbreviated Scenario

- In the Zuma region of Zanda, there are four tribes called Acanda, Bango, Conda, and Dengo.
- They represent 5%, 20%, 30%, and 45% of the Zuma's population, respectively.
- Assume that Acanda and Conda are hostile tribes, whereas Bango and Dengo are friendly.
- Your government would like to improve its understanding of this region and has captured a randomly chosen inhabitant to be interviewed.
- …

- Assume that your government has already determined the following information which is at your disposal:

- Acanda: 10% of the tribe is under 40 years of age, 75% use social media, 50% speak Zebin (one of two languages spoken in Zuma), 25% are employed, 90% practice a religion, 25% come from a large family (i.e., more than 4 children), 50% have been educated up to the age of 16, 75% have a reasonably high socio-economic status relative to the general population, 75% speak Zimban (one of two languages spoken in Zuma), 75% have a political affiliation, 75% wear traditional clothing, and 25% have fair coloured skin.
- Bango: …

- The target person: The target is under 40 years of age, uses social media, speaks Zebin, is employed, does not practice a religion, does not come from a large family, does not have education up to age 16, does not have a reasonably high socio-economic status, speaks Zimban, is not politically affiliated, wears traditional clothing, and does not have fair coloured skin.

- ***Measures - Qualitative***
  - ACH Group:
    - In order to solve the analytic task presented, we would like you to use the technique called 'Analyses of Competing Hypotheses' (ACH). This consists of the steps described below…
  - Control Group:
    - Report your conclusions in the box below. Consider the relative likelihood of all of the hypotheses. State which items of information were the most diagnostic, and how compelling a case they make in identifying the most likely hypothesis. Also say why alternative hypotheses were rejected...

- ***Measures – Quantitative***
  - Rate likelihood of each hypothesis
  - Rate diagnosticity of each piece of evidence

Qualitative Results
(excl. diagnosticity results)

Table 1. Understanding the Task

	ACH Group	Control Group
Drew a matrix	100%	80%
Identified the 4 specific hypotheses	92%	100%
Identified the 2 general hypotheses	4%	0%
Identified the 12 evidence items	68%	68%

Table 2. Reasoning Strategies

	ACH Group	Control Group
Included assumptions on which scoring was based	20%	N/A
Used base-rates	12%	52%
Added up only consistent evidence (CC, C)	4%	22%, *n* = 20
Added up only inconsistent evidence (II, I)	20%	0%, *n* = 20
Added up both consistent and inconsistent evidence	76%	78%, *n* = 20

Table 3. Other Strategies Used by Control Group

Made an evidence matrix/list • Listed evidence percentages for each hypothesis • Scored/ranked each evidence item for each hypothesis • Other approaches	80% • 40% • 28% • 12%
Used some form of scoring system • Added up evidence likelihood percentages • Points-based – matching evidence (>50%) scores 1 point • Points-based – e.g., ≥ 75% = 3, ≥ 50% = 2, ≥ 25% = 1	64% • 22% • 22% • 20%
• *[see last four rows of Table 2 for information about other control group strategies]*	

Table 4. Conclusions Drawn

	ACH Group	Control Group
Provided tentative conclusion	88% but only 64% matched matrix	N/A
Provided final conclusion consistent with their matrix	68%	100%, *n* = 16
Final conclusion matched tentative conclusion	86%	N/A
Made some attempt to assess sensitivity of conclusion	60%	4%
Provided at least one indicator	72%	N/A

Table 5. Within-Analyst Consistency

	ACH Group	Control Group
Consistency of scoring rule used	4%	44%, *n* = 9
Accuracy in transforming negative evidence values	28%	63%, *n* = 16
Consistently applied integration rule across hypotheses	76%	80%, *n* = 15

Table 6. Between-Analyst Consistency

	ACH Group	Control Group
Consistency in scoring of evidence items	*M* Kappa = .37 (*SD* = .23)	Used many different scoring rules
[See last three rows of Table 2 for consistency in how analysts integrated evidence]		

Conclusions

- How well did analysts trained to use ACH actually apply it in practice?
  - Only 1 analyst performed all 8 ACH steps correctly!
  - 32% performed 7 steps correctly
    - 88% only mistake was adding up both positive and negative scores
  - Only 20% focused on inconsistent evidence (a key element of ACH)
  - 68% reached a conclusion that matched the results of their ACH matrix

- What hypothesis-testing strategies did untrained analysts use?
  - 80% used a matrix of some kind to evaluate the evidence
  - 78% took account of both consistent and inconsistent evidence
  - Much *more* likely than ACH participants to take account of base-rates

- THEREFORE:
  - Little evidence that analysts suffer from conformation bias
  - Analysts' intuitive hypothesis testing strategies are more cognitively complex than the strategy applied by ACH
  - ACH may adversely interfere with analysts' intuitive hypothesis testing strategies