Connecting the Dots through the Discovery of Evidence, Hypotheses, and Arguments from Masses of Data

Gheorghe Tecuci, Dorin Marcu, Mihai Boicu, David Schum

Learning Agents Center, Volgenau School of Engineering
George Mason University, Fairfax, VA 22030
tecuci@gmu.edu, http://lac.gmu.edu

George Mason University and IBM ASC Symposium
Fairfax, VA, 4 November 2014

Summary

We present research performed in the Learning Agents Center to develop a computational theory of evidence-based reasoning and to implement it in intelligent analytical tools, such as Disciple-CD (Disciple cognitive assistant for Connecting the Dots) and COGENT (Cognitive Assistant for Cogent Analysis), addressing the complex task of "connecting the dots" to discover knowledge from masses of data of all kinds.

"Connecting the Dots" is performed through a mixed-initiative process of ceaseless discovery of evidence, hypotheses and arguments in a non-stationary world, process integrating analyst's imagination with agent's knowledge and evidence-based reasoning, and involving abductive, deductive, and inductive inference. The analyst and the cognitive assistant marshal thoughts and evidence to generate or discover productive competing hypotheses, use the hypotheses to discover new evidence, and construct defensible and persuasive arguments on the hypotheses believed to be most favored by the evidence that has been gathered and evaluated.

We illustrate our approach to "connecting the dots" in the area of intelligence analysis, and discuss its application to other areas, including cyber insider threat, forensics, medicine, law, and natural sciences.

Overview

Development of Cognitive Assistants

Computational Theory of Evidence-based Reasoning: Application to Intelligence Analysis

From TIACRITIS to Disciple-CD and to COGENT

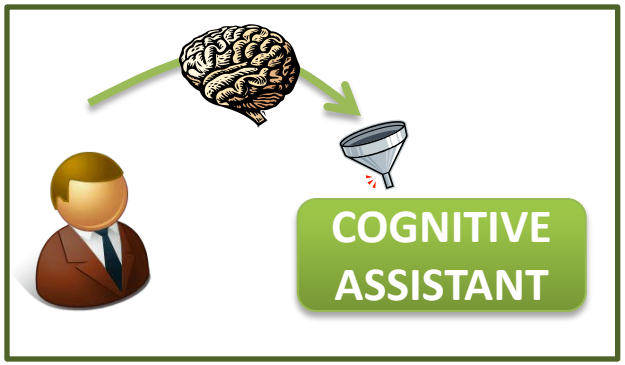
“Knowledge Engineering” Book and Disciple-EBR

“Intelligence Analysis” Book and Disciple-CD

Computational Theory of Evidence-based Reasoning: Application to other Domains

Theory, Methodology, and Tools for the Development of Cognitive Assistants

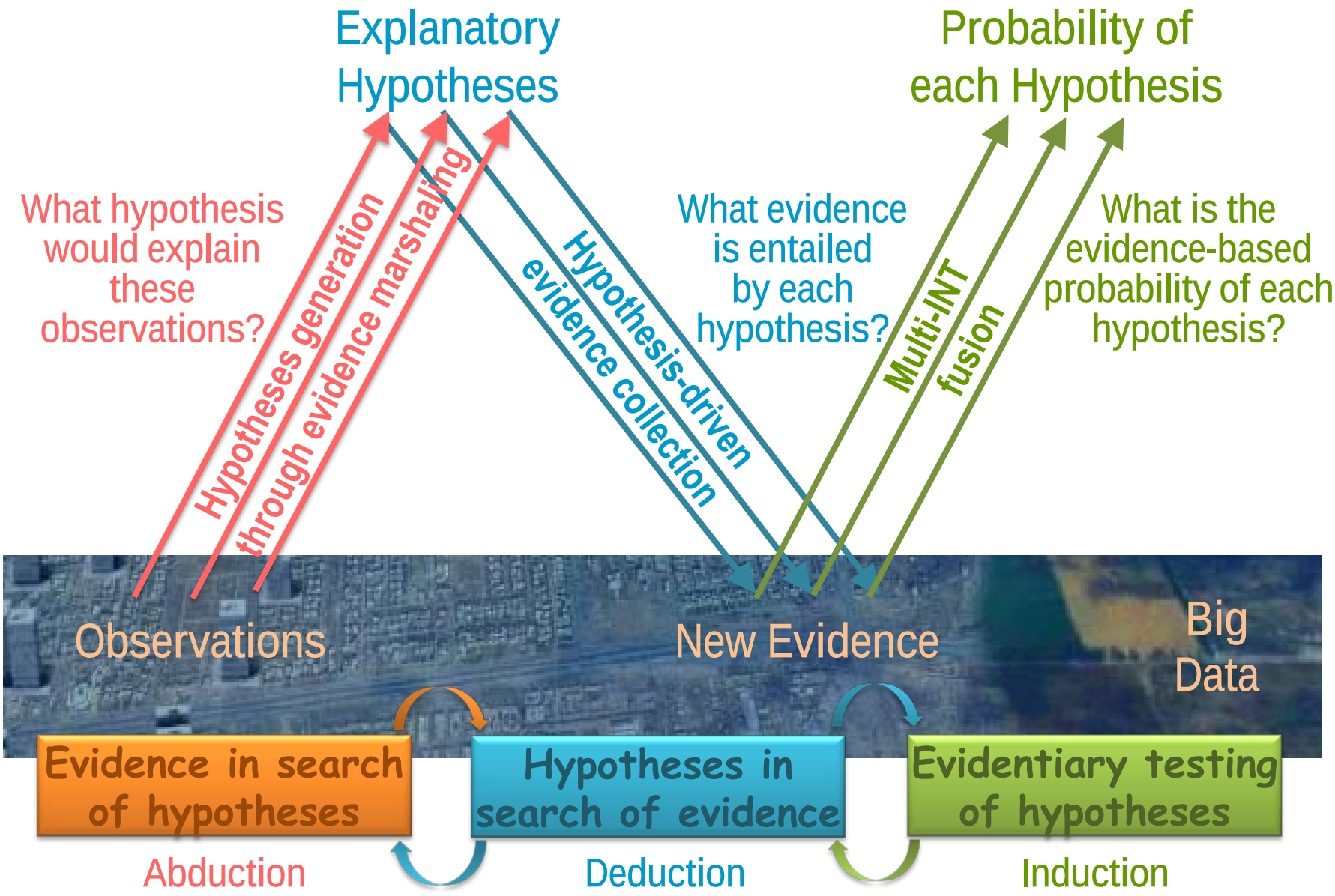
- Learn the (explicit and tacit) knowledge of subject matter experts



- Assist their users in complex problem solving and decision making

- Train junior professionals and students

Computational Theory of Evidence-based Reasoning



Astonishing Complexity of Intelligence Analysis

COMPLEXITY

TIME

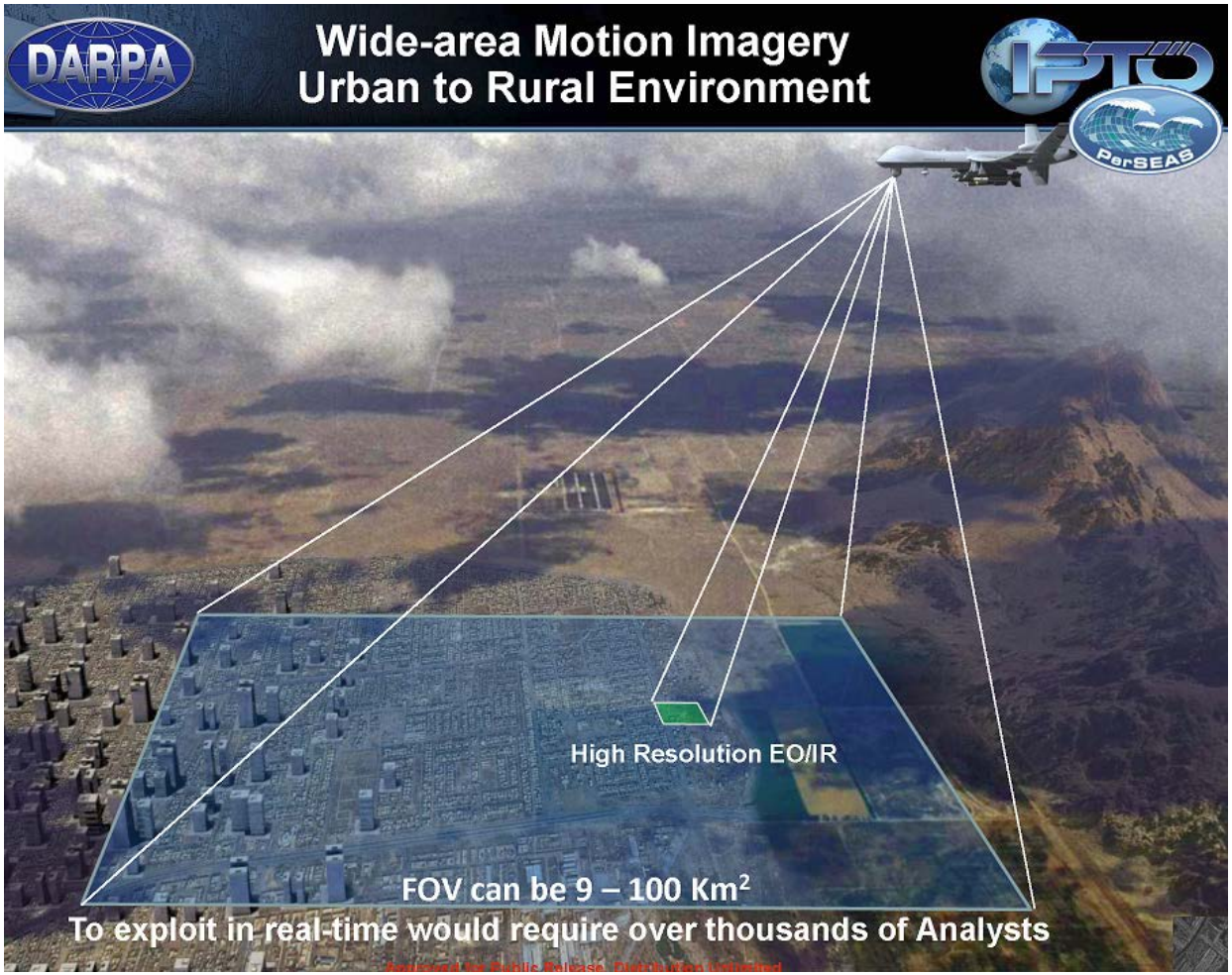
Often Stunningly Complex Arguments

- imaginative reasoning
- critical reasoning

Masses of Evidence

- inconclusive
- incomplete
- dissonant
- ambiguous
- various degrees of credibility

Sample Problem: Analysis of Wide-Area Motion Imagery



From: Mita Desai, Multi-entity activity discovery over large space-time windows, DARPA, http://www.darpa.mil/ipto/solicit/baa/BAA-09-55_ID01.pdf

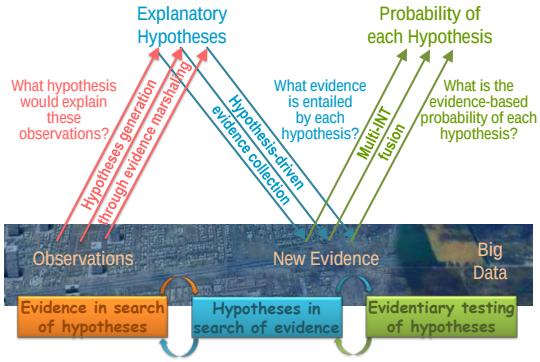
Real-Time Analysis

Discover impending threat events (e.g., ambush, rocket launch, IED, suicide bomber, false check-point, kidnapping, etc.) early enough to be able to interdict them.

Forensic Analysis

Backtrack from a past event (e.g., an ambush) and discover participants, possible related locations and events, and movement patterns.

Discovery of Hypotheses



ambush threat

deceptive threat

ambush preparation

traffic disruption

traffic deviation

road block

road repair

road work

not road work

Evidence in search of hypotheses

E_1: Evidence of road work at Al Batha junction at 1:17am

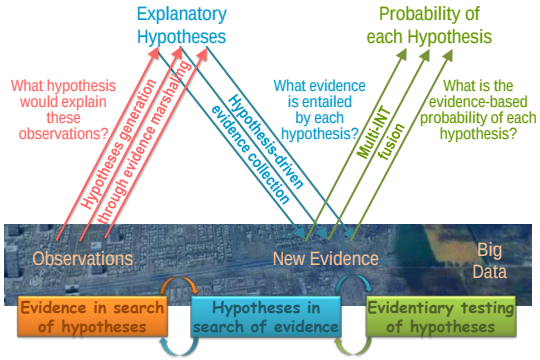
What hypotheses would explain this observation?

Big Data

Discovery of Evidence

Hypotheses in search of evidence

Assuming that this hypothesis is true, what other things should be observable?



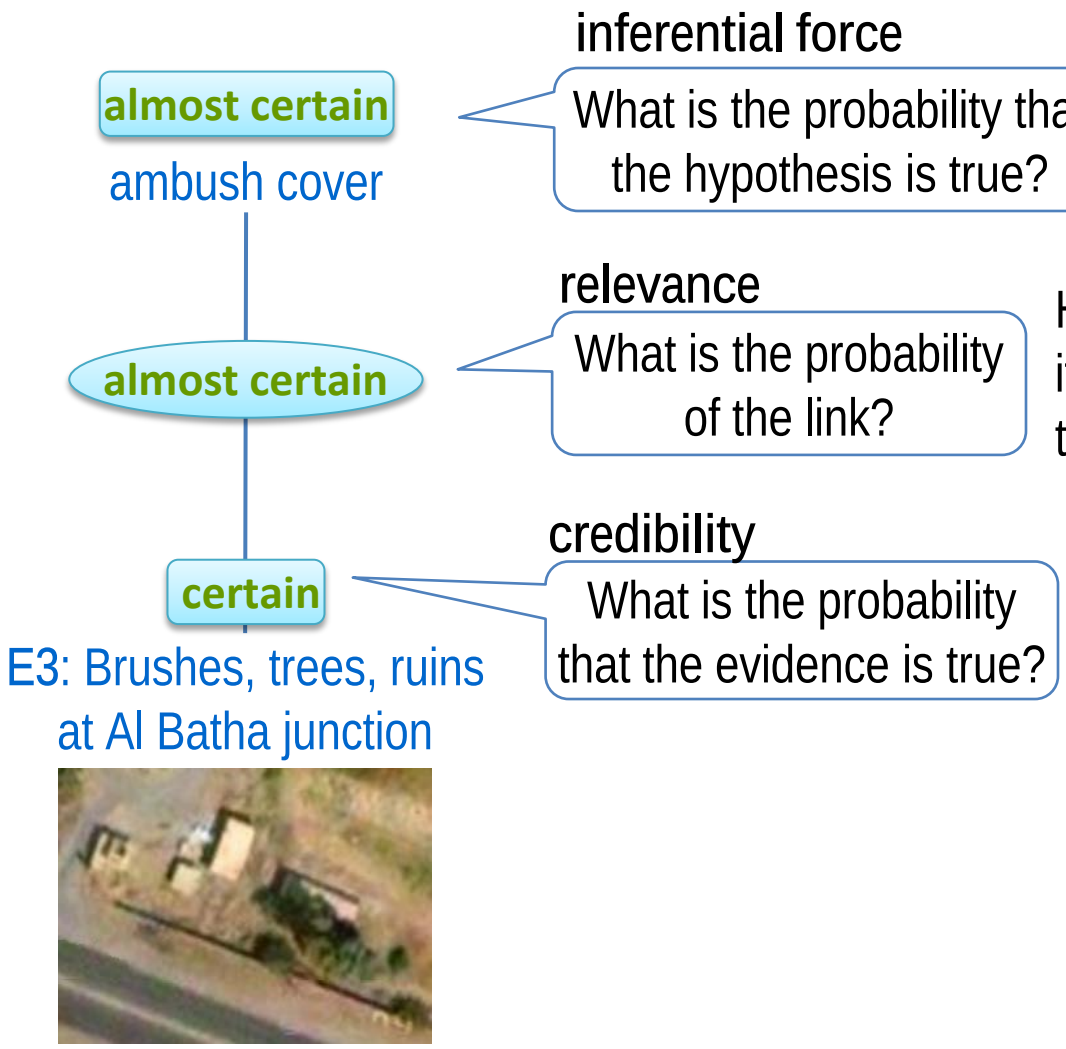
Ambush threat at Al Batha highway junction around 1:17am.

&

- good location
  - &
    - route used by U.S. forces
      - Search for evidence that the Al Batha highway junction is on a route used by the U.S. forces.
    - ambush cover
      - Search for evidence that there is ambush cover near the Al Batha highway junction.
- ambush preparation
  - &
    - deployment of terrorists
      - &
        - people descended from vehicle
          - Search for evidence that people descended at the Al Batha highway junction from a vehicle short before 1:17am.
        - vehicle departed from terrorist facility
          - Search for evidence that the vehicle that drove the people to the Al Batha highway junction short before 1:17am, departed from a terrorist facility.
    - road block
    - people move to cover

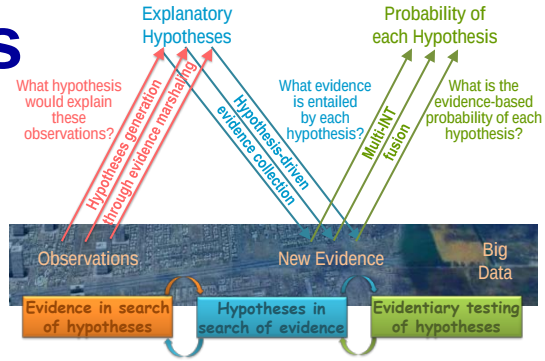
Evidence-based Hypothesis Assessment

Simple, intuitive, rigorous, and less error-prone probability system and evidence-based assessment



Evidentiary testing of hypotheses

Discovery of Arguments



It is very likely that there is an ambush threat to the U.S. forces at the Al Batha junction.

very likely
ambush threat

&

almost certain
good location

&

certain
route used

certain

certain

E_2: AFC confirms route

almost certain
ambush cover

almost certain

certain

E_3: brushes, trees, and ruins

very likely
ambush preparation

&

very likely
deployment of terrorists

almost certain
road block

certain
people move to cover

&

almost certain
people descended from vehicle

certain

almost certain

E_4: People get off pickup truck

very likely
vehicle departed from terrorist facility

very likely

almost certain

E_5: Pickup truck leaves abandoned rice farm

Simple Baconian/Fuzzy composition functions: *Minimum, Maximum, On Balance*

Advanced Tools for Intelligence Analysis: From TIACRITIS to Disciple-CD and to COGENT

Improvements over TIACRITIS

- Probability system
- Argument development
- Evidence-based reasoning
- Knowledge base management
- Usability
- Scalability
- Reliability

Version 1 (Summer 2014)

Cognitive Agent for Cogent Analysis

COGENT

2012-2016

Disciple Assistant for Connecting the Dots

Disciple-CD

2011-2014

Teaching Intelligence Analysts Critical Thinking Skills

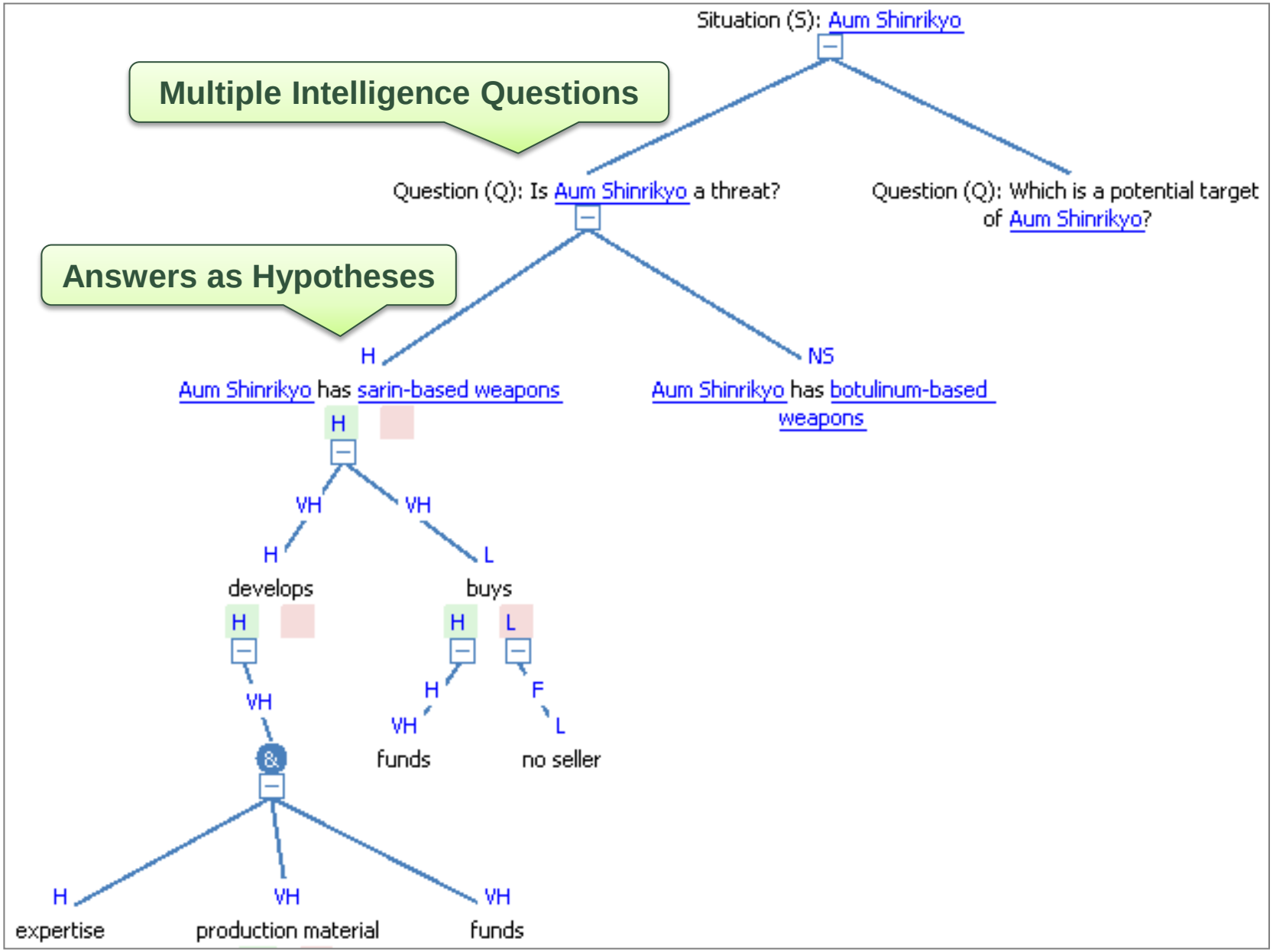
TIACRITIS

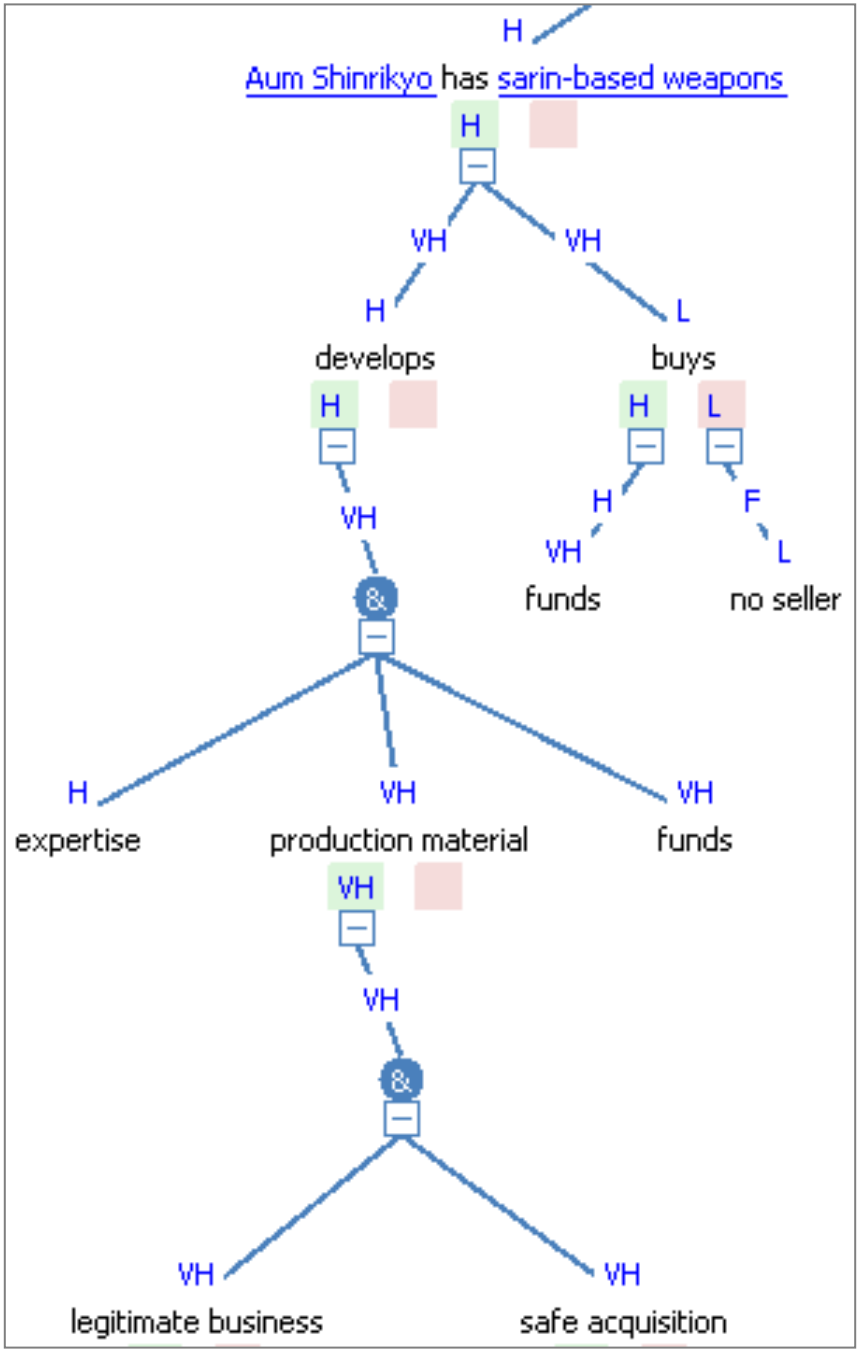
2009-2011

New Generation Tool

- Easy to use
- Enforcing cogent analyses
- Learning and reuse
- Collaborative analysis
- Enabling fast analyses
- Customizable scale

Cogent: Cognitive Agent for Cogent Analysis





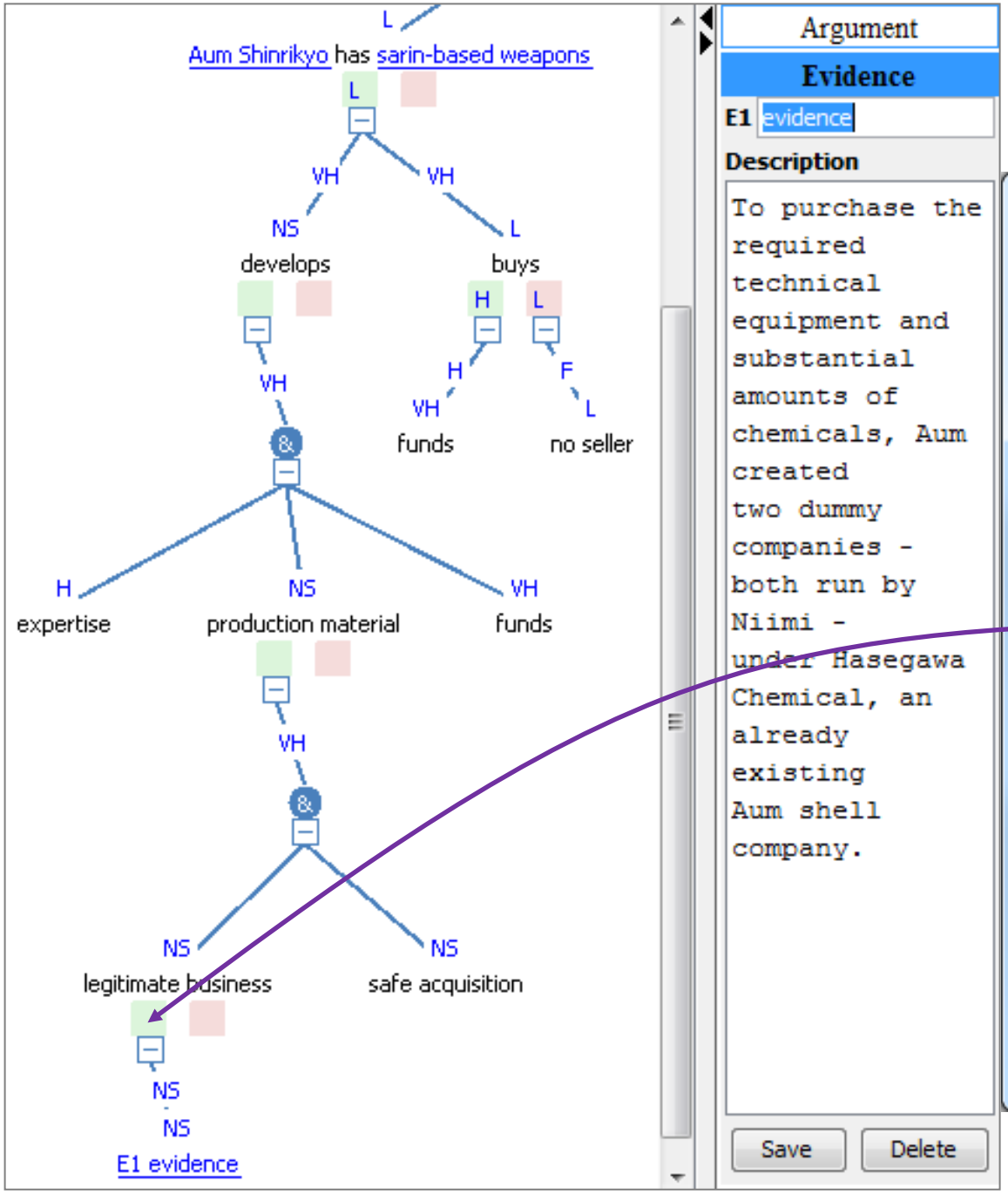
Customizable assessment scale

Strength	Probability	Belief
F (Full strength)	C (Certain)	TB (Total Belief)
VH (Very High)	AC (Almost Certain)	SB (Strong Belief)
H (High)	VL (Very Likely)	MB (Moderate Belief)
M (Medium)	L (Likely)	WB (Weak Belief)
L (Low)	LS (Lack of Support)	LB (Lack of Belief)
VL (Very Low)	NS (Not Set)	NS (Not Set)
N (No strength)		
NS (Not Set)		

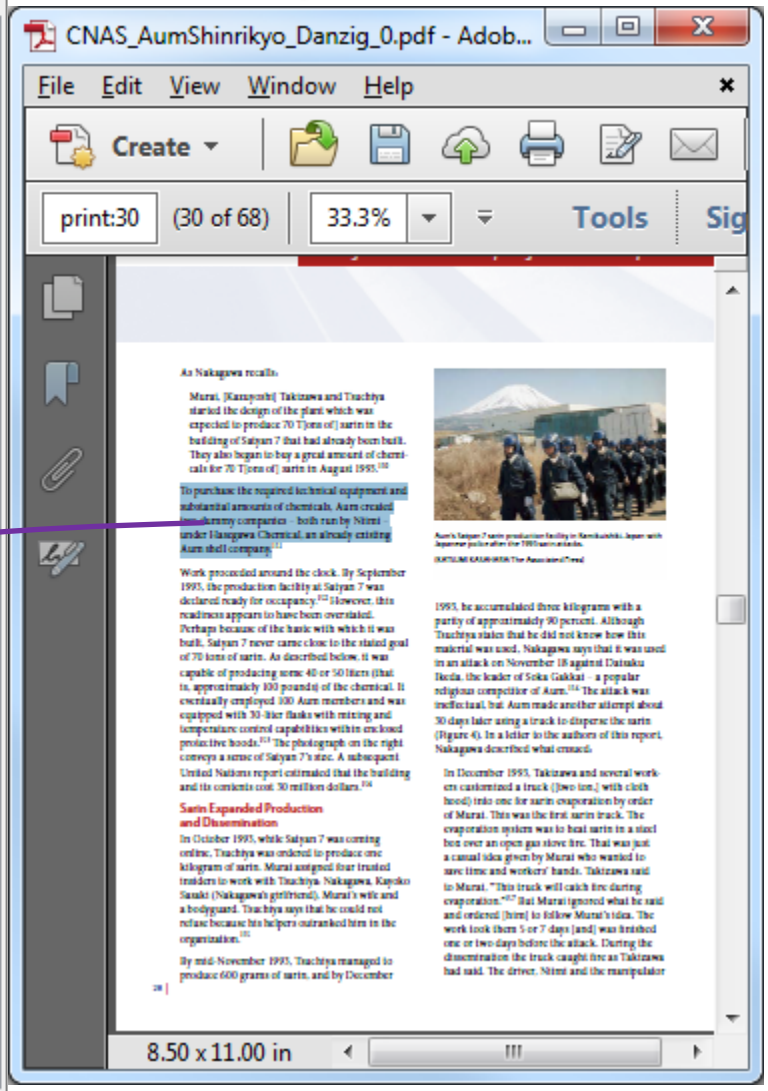
On balance function

Disfavoring arguments
Support for negation of Hypothesis

Favoring arguments Support for Hypothesis / Strength of Hypothesis	N	VL	L	M	H	VH	F
F	F	VH	H	M	L	VL	N
VH	VH	H	M	L	VL	N	N
H	H	M	L	VL	N	N	N
M	M	L	VL	N	N	N	N
L	L	VL	N	N	N	N	N
VL	VL	N	N	N	N	N	N
N	N	N	N	N	N	N	N

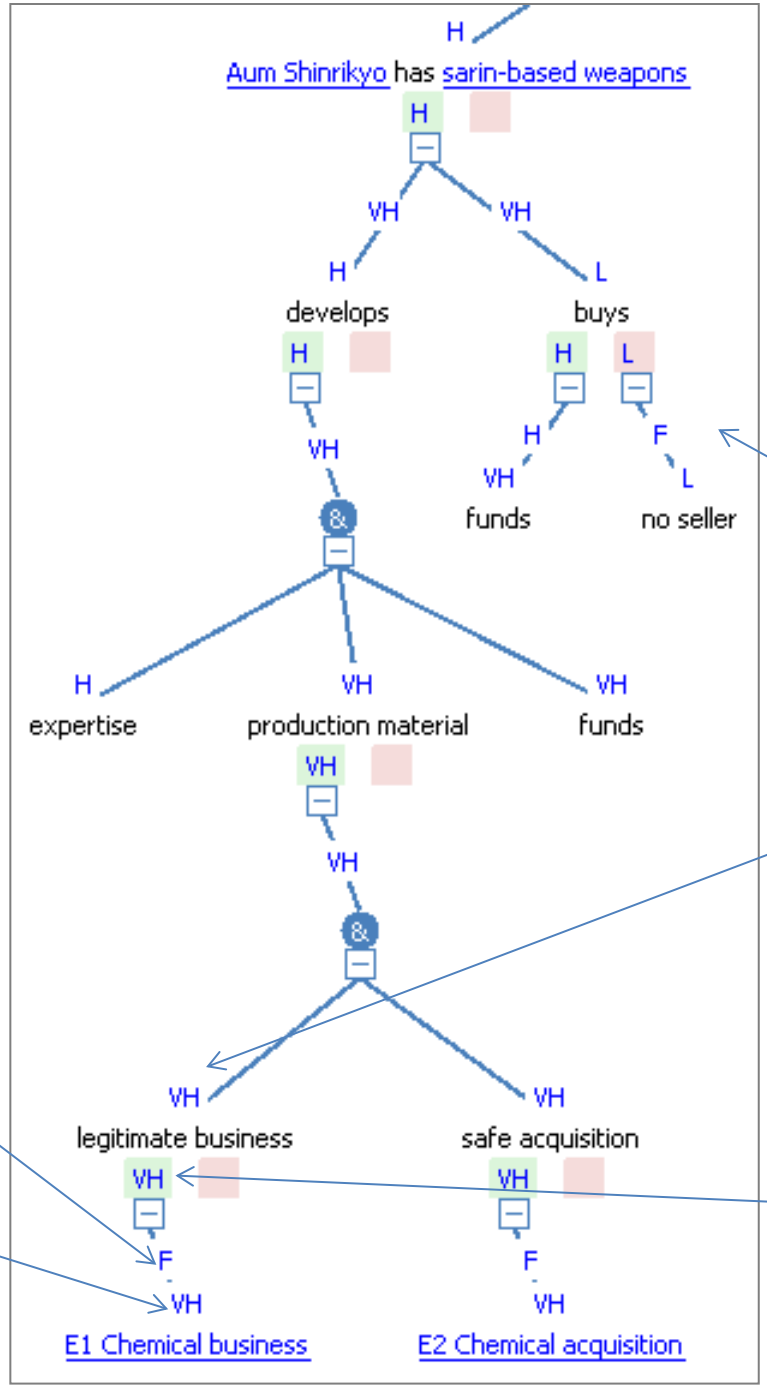


Attaching evidence to hypothesis



Analyst assessments

Cogent assessments



2

Strength of link: How strong is the link between what the evidence states and the hypothesis? That is, assuming that Aum has indeed created the two dummy chemical companies, how strong is the hypothesis that it has a legitimate business which is justified to acquire sarin?

1

Credibility of evidence: How high is the credibility of E1 (i.e., that Aum has indeed created two dummy chemical companies)?

5

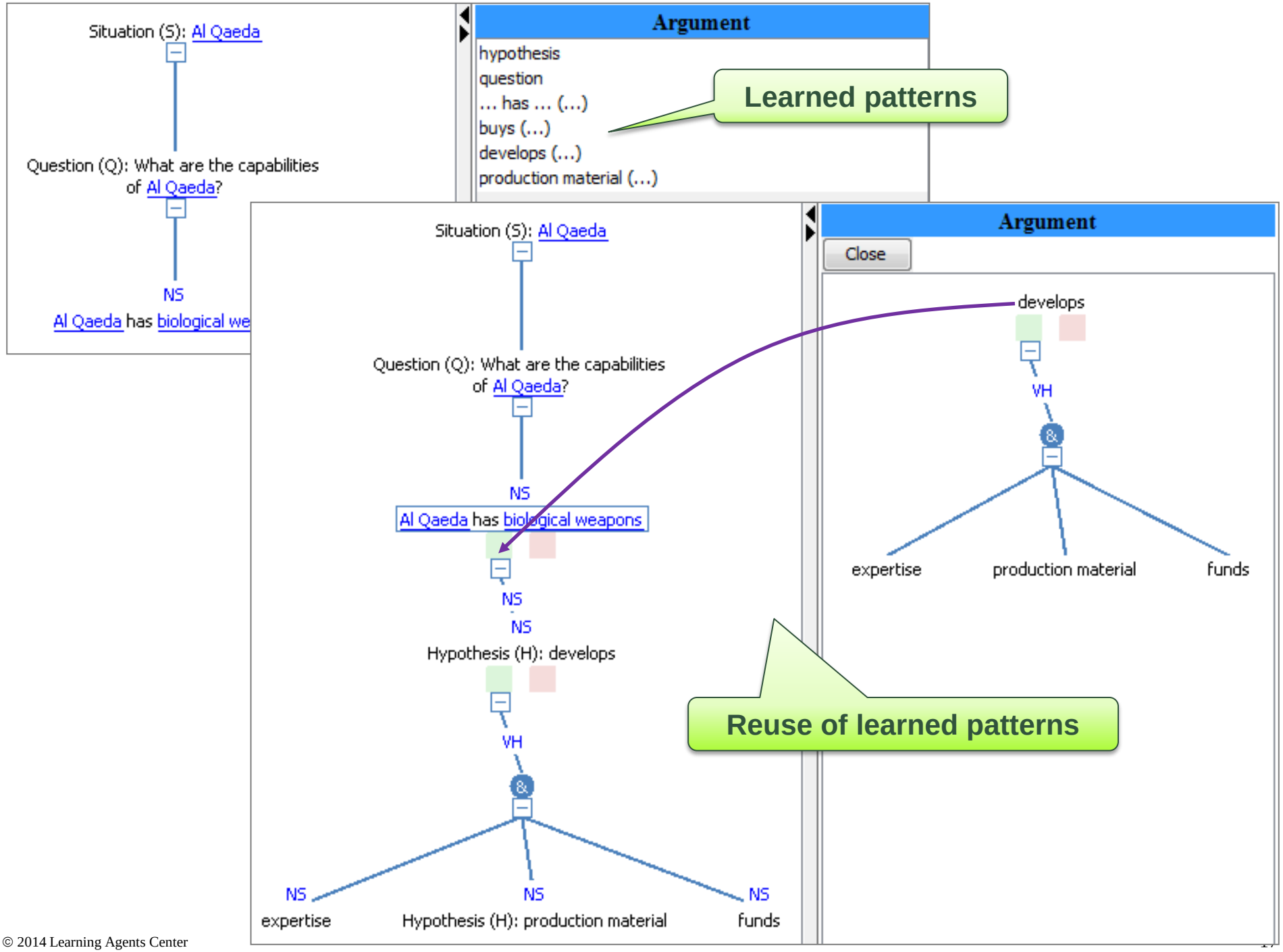
Strength of upper-level hypotheses

4

Strength of hypothesis (based on both favoring and disfavoring arguments)

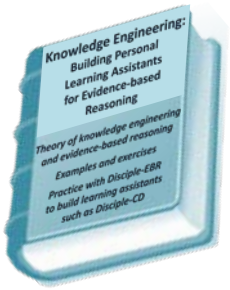
3

Strength of favoring argument: What is the strength of the favoring argument for the "legitimate business" hypothesis, based only on E1?



Knowledge Engineering Book (with Disciple-EBR)

KNOWLEDGE ENGINEERING: *Building Personal Learning Assistants for Evidence-based Reasoning*



- Introduction
- Evidence-based Reasoning: Connecting the Dots
- Methodologies and Tools for System Design and Development
- Modeling the Problem Solving Process
- Ontologies
- Ontology Design and Development
- Reasoning with Ontology and Rules
- Learning for Knowledge-based Systems
- Rule Learning
- Rule Refinement
- Abstraction of Reasoning
- Disciple Agents (Disciple-WA, Disciple-COA, Disciple-COG, and Disciple-VPT)

Theory of knowledge engineering and evidence-based reasoning

Examples and exercises at each chapter

Practice with Disciple-EBR to build learning assistants such as Disciple-CD

Intelligence Analysis Book (with Disciple-CD)

Intelligence Analysis as Discovery of Evidence, Hypotheses, and Arguments:

Finding and Connecting the Dots

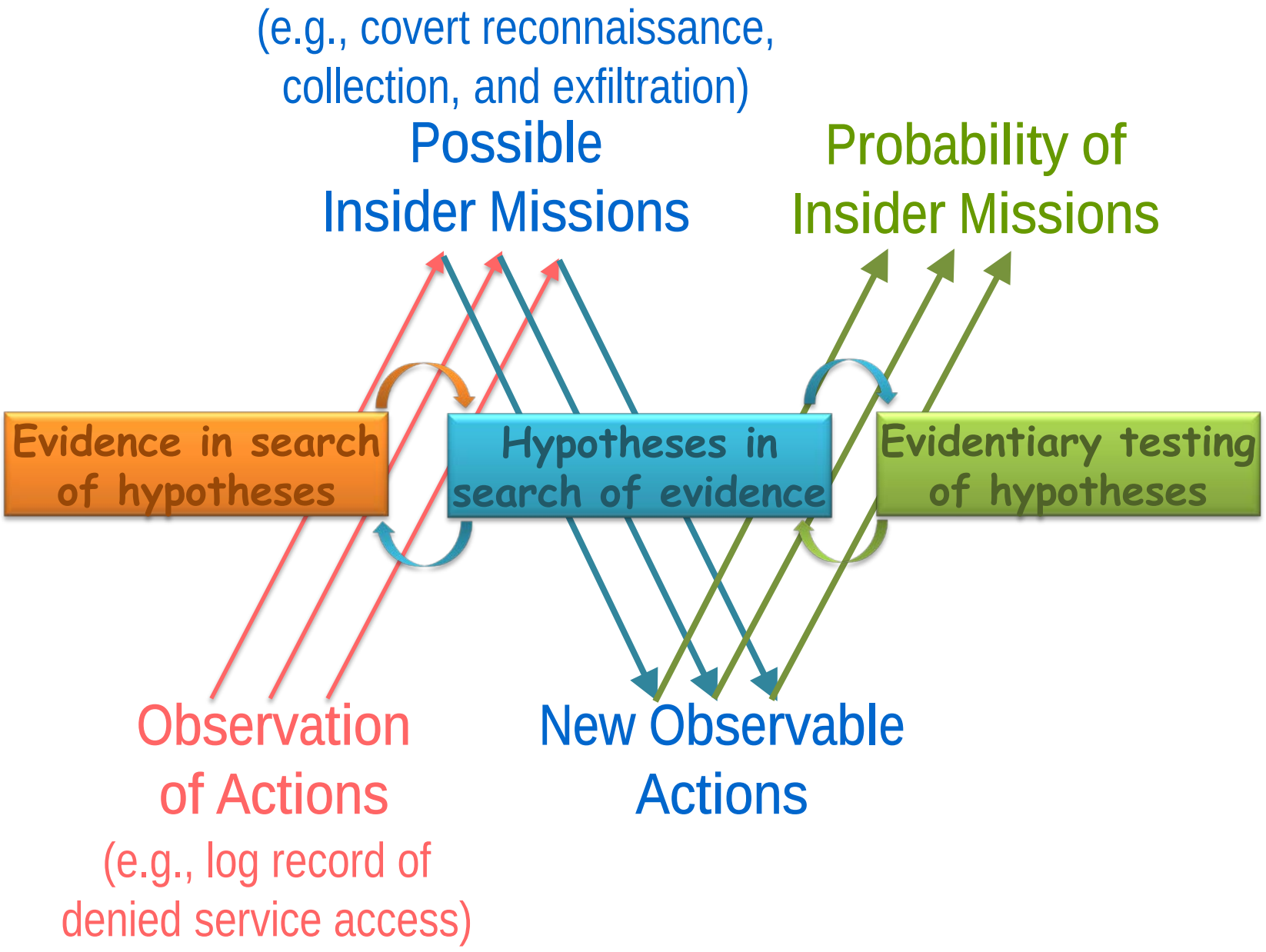
Theory of intelligence analysis and evidence-based reasoning

- Intelligence Analysis: "Connecting the Dots"
- Marshaling Thoughts and Evidence for Imaginative Analysis
- Disciple-CD: A Cognitive Assistant for Intelligence Analysis
- Evidence
- Divide and Conquer: A Necessary Approach to Complex Analyses
- Assessing the Believability of Evidence
- Chains of Custody
- Recurrent Substance-blind Combinations of Evidence
- Major Sources of Uncertainty in Masses of Evidence
- Assessing and Reporting Uncertainty: Some Alternative Methods
- Analytic Bias
- Appendices

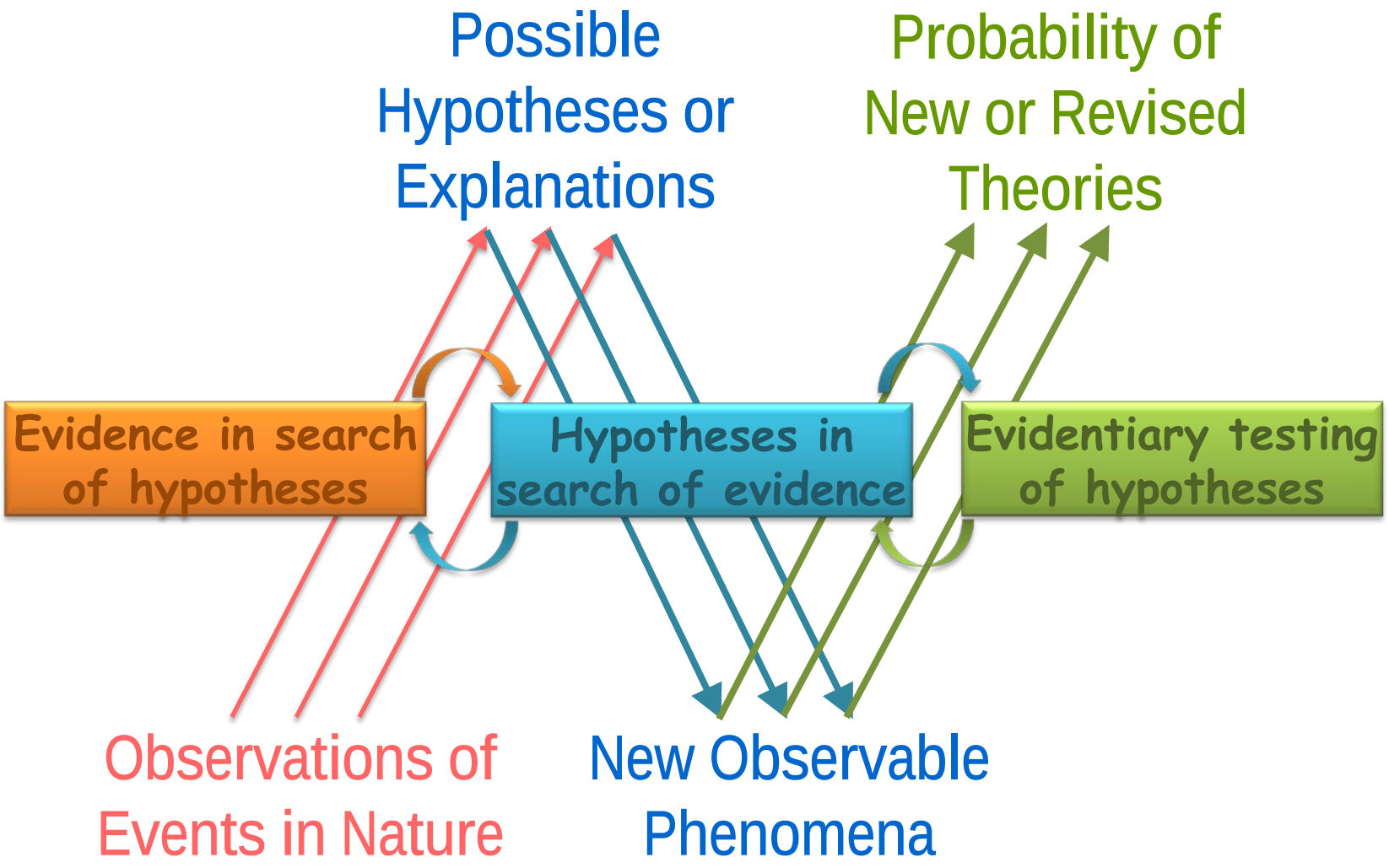
Examples and exercises at each chapter

Basic and advanced practice with Disciple-CD to assess hypotheses based on evidence

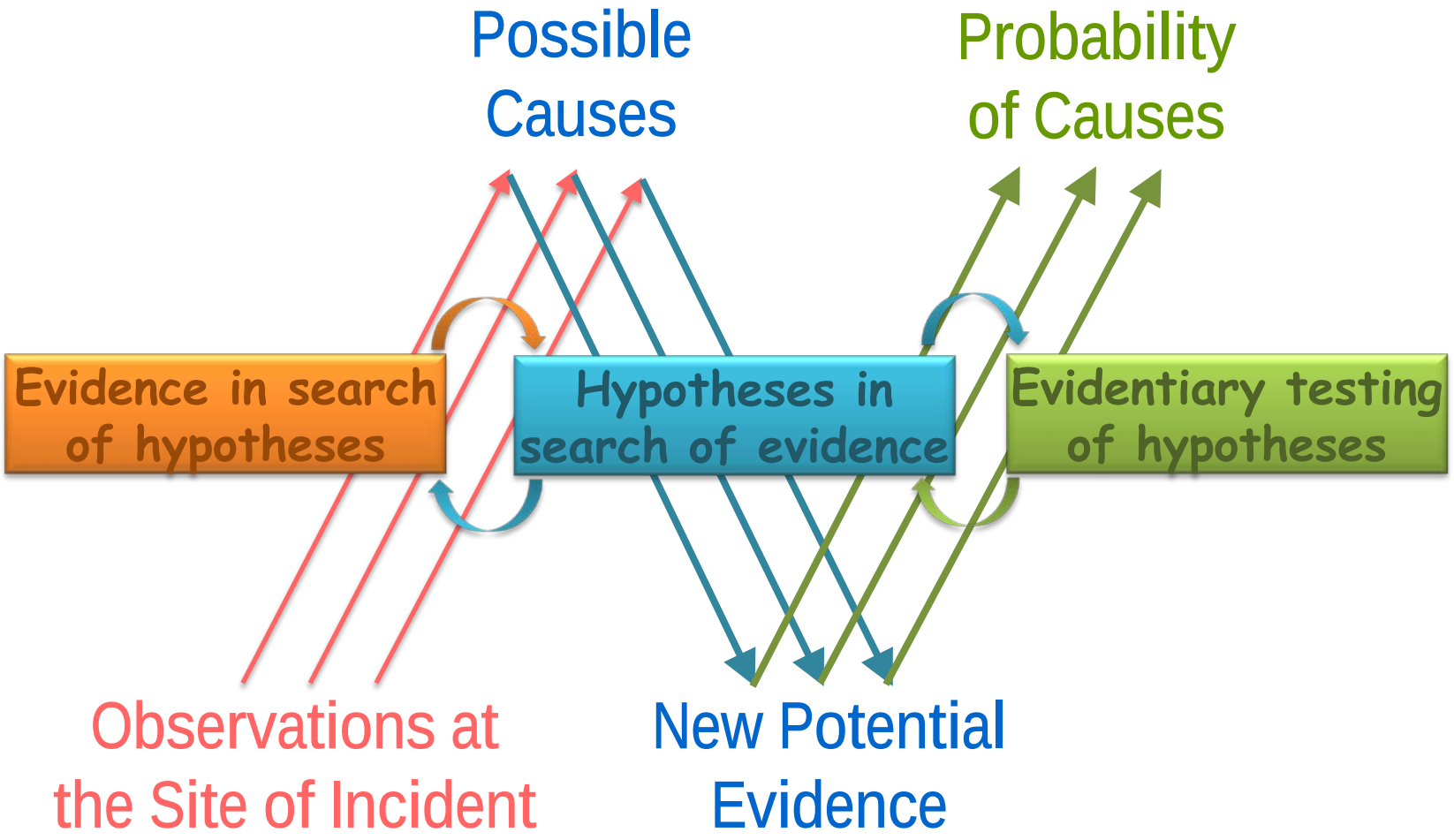
Cyber Insider Threat Discovery and Analysis



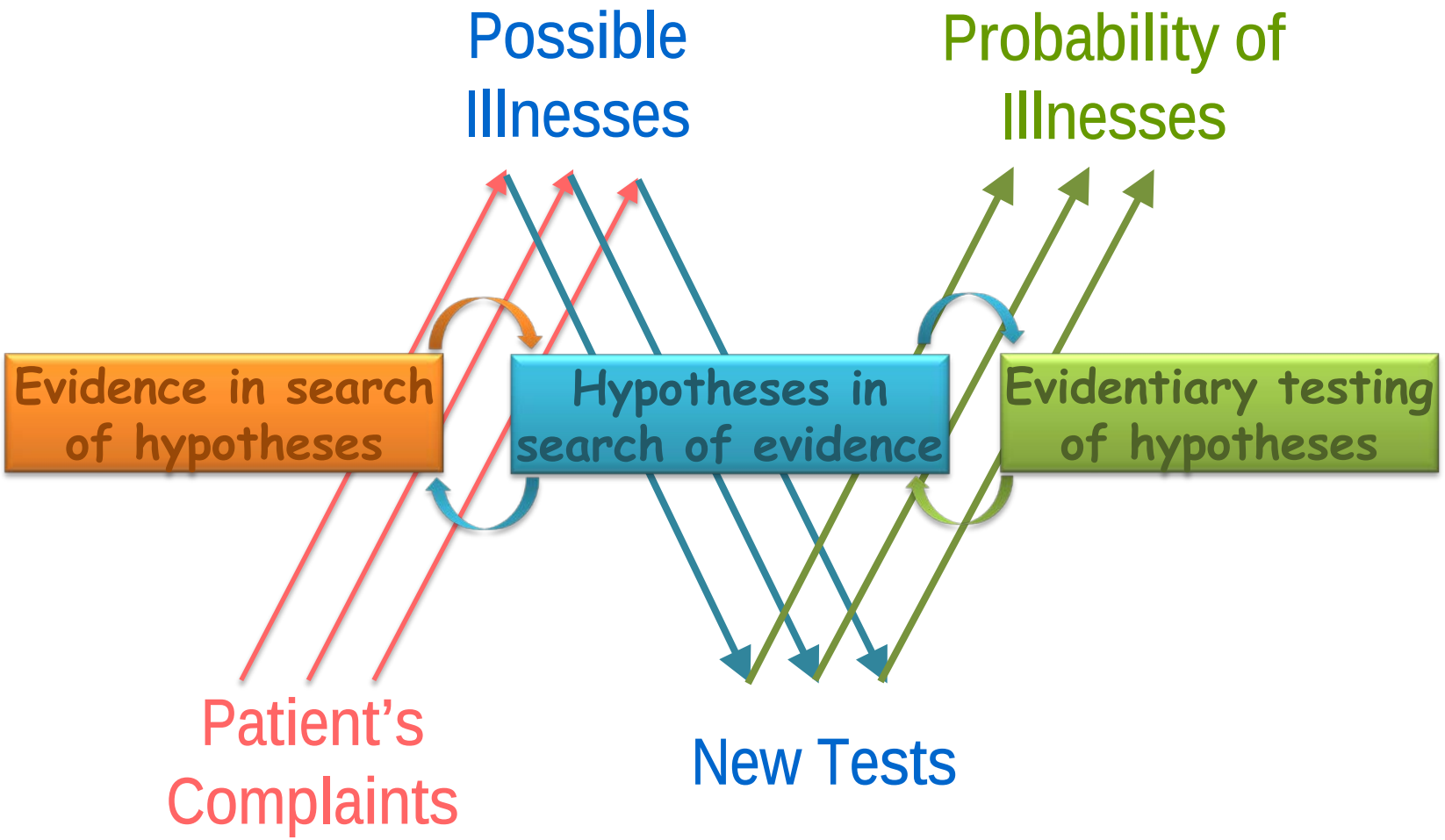
Natural Sciences



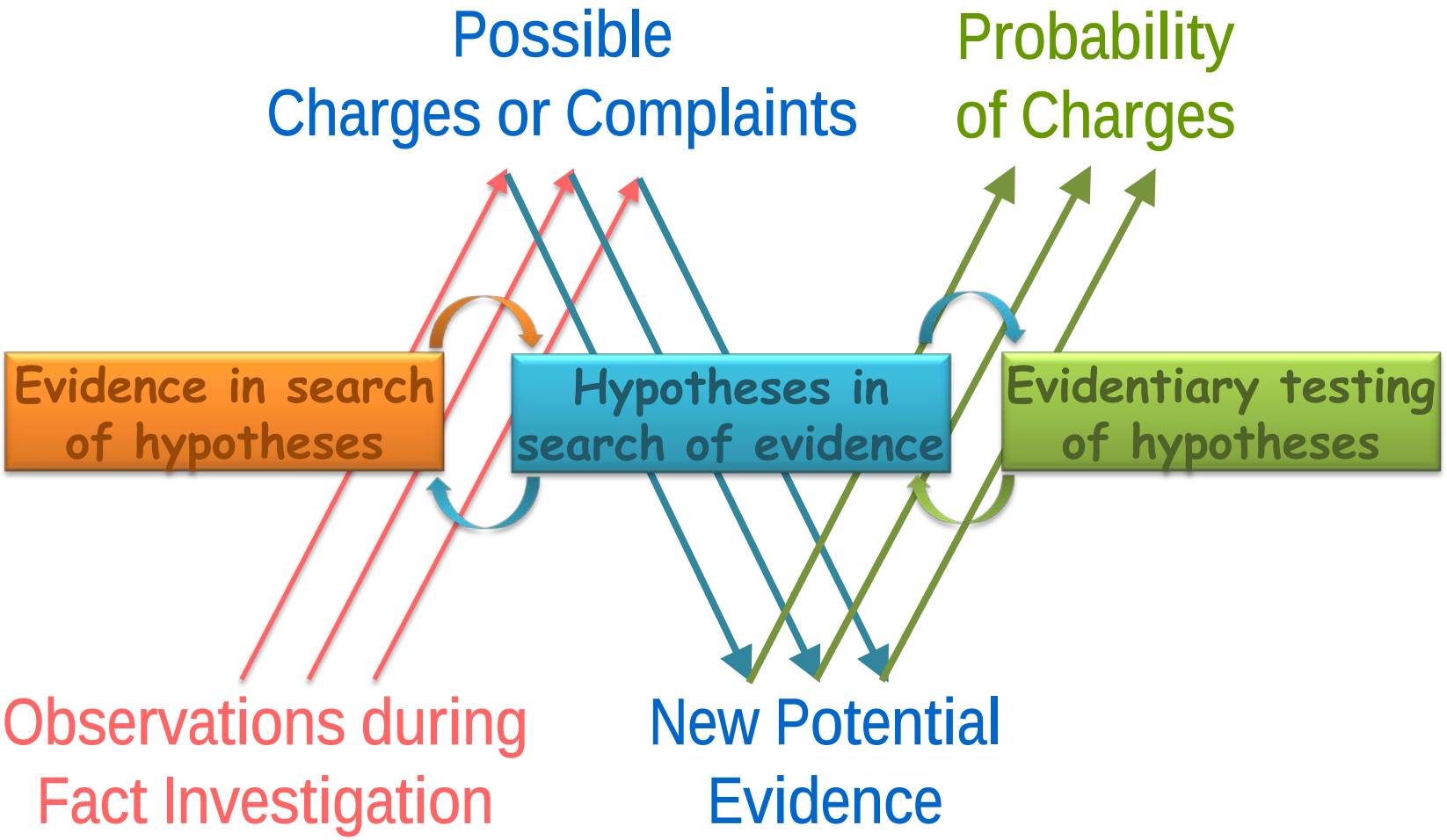
Forensics



Personalized Medicine



Law



Key Elements of the Computational Theory of EBR

- Developed in the framework of the scientific method.
- Systematic approach to evidence-based reasoning through a synergistic integration of abductive, deductive, and inductive reasoning.
- Computational models for essential analytical tasks (e.g., evidence marshaling, hypothesis-driven evidence collection, multi-INT fusion, detection and mitigation of bias).
- General analysis structure with favoring and disfavoring arguments for competing hypotheses.
- Intuitive system of Baconian probabilities with Fuzzy qualifiers, allowing customizable assessment scales.
- Substance-blind ontology of evidence.
- General procedures for credibility assessment.
- Context-based rules learned from expert analysis examples.

Questions

Contact information

Gheorghe Tecuci
Professor of Computer Science and Director of the Learning Agents Center
MSN 6B3, Learning Agents Center, George Mason Univ., Fairfax, VA 22030
tecuci@gmu.edu tel 703 993 1722 http://lac.gmu.edu/