An Overview of Explanation: Concepts, Uses, and Issues

Benjamin Grosof*

Ontology Summit, session on Explanation
45 minutes including discussion. Jan. 23, 2019, via webconference.

These slides will be available soon via link at the author’s website, at:
http://benjamingrosof.com/misc-publications

Thanks also for material courtesy of Coherent Knowledge also co-authored by Michael Kifer**, Paul Fodor**, and Janine Bloomfield***

* Kyndi ; work in part done while previously at Coherent Knowledge
  http://benjamingrosof.com

** Stony Brook University, and Coherent Knowledge
  http://www.cs.stonybrook.edu/~kifer
  http://www.cs.stonybrook.edu/~pfodor

*** Coherent Knowledge
  http://coherentknowledge.com
  https://www.linkedin.com/in/janinebloomfield

Note on Copyright and Permissions:
Slides with Coherent Knowledge logo are, in whole or part, courtesy of Coherent Knowledge; and several other slides as well are, in whole or part, courtesy of Coherent Knowledge.
Bio – Benjamin Grosof

• AI researcher, executive, and entrepreneur
• Chief Scientist at Kyndi – AI startup on NL question answering using ML+KRR
• Co-founder & Board of Coherent Knowledge – AI startup on KRR

Previously:
• Principal Director & Research Fellow in AI at Accenture on BPM
• CTO & CEO of Coherent Knowledge
• Directed advanced AI research at predecessor of Allen Inst. for AI
  • Developed Rulelog KRR theory, algorithms, UI approach
• MIT Sloan professor and DARPA PI
  • Co-Founder of RuleML, key contributor to W3C OWL-RL and RIF standards
• IBM Research, creator IBM Common Rules
  • 1st successful semantic rules product in industry
  • Stanford AI PhD, combining ML with logical and probabilistic reasoning

Themes: flexible clean KRR + NL + ML; many app domains & tasks

• http://www.linkedin.com/in/benjamingrosof
• http://benjamingrosof.com
Outline of Talk

• Observations
  – Concepts of Explanation
  – Applications & Benefits of Explanation
  – Example Industry Leaders in Explanation Technology
  – Issues Today in the Field of Explanation

• Case Study
  – Proof style explanation in
    Financial Regulatory/Policy Compliance
“Gold Standard” concepts of explanation:

1. **Proof** – deductive, in a formal knowledge representation (KR)
   - (There are multiple formalizations of deductive models / knowledge / reasoning)
   - E.g., natural deduction – structured kinda like a proof in high school geometry
     - humble(Socrates). Derived from the knowledge base (KB):
       - humble(Socrates) :- mortal(?person) \and wise(?person).
       - mortal(?person) :- human(?person) \and nonfictional(?person).
       - nonfictional(?person) :- historical(?person).
       - historical(Socrates). historical(Barbara).
       - wise(Socrates). wise(Barbara).
   - E.g., probabilistic (uncertain)
     - p(slippery) = 0.31. Derived from the KB:
       - p(slippery \| rain) = 0.8.
       - p(slippery \| \neg rain) = 0.1.
       - p(rain) = 0.3.
       - p(\neg rain) = 0.7.

2. **Causal**
   - Causal models / knowledge / reasoning
   - (There are multiple formalizations of causal relationships)
Other, Frequent concepts of explanation:

- **Provenance** (i.e., about source)
  - E.g., fact sentence F41 was drawn from document D73, at URL U84, in section 22, line 18.
  - E.g., rule sentence R56 was encoded from text sentence in document D73, at URL U84, in section 15, lines 11-12.

- **Transparency** – of an algorithm/process for inference
  - (“Inference” here is superclass of deduction, induction, and abduction.)
  - E.g., facts FGH -- or predictions PQR -- or decisions MNO – were extracted as results from machine learning (ML) algorithm XYZ, selected by person B in organization O, run on input data CDE, in implementation I by person A at time T.

- **Drill down** – interactivity for exploration, e.g., within graph/tree

- **Focus** – i.e., locus of attention
  - E.g., heat, or weight of activity

- **Specialized presentations of information**
  - E.g., comparative bar/pie charts, geospatial maps, timelines
Concepts of Explanation (III)

Other, Trending-Up concepts of explanation:

- **Influentiality**
  - E.g., heavily weighted hidden nodes and edges affecting discrimination of output, in a deep learning neural network (NN)

- **Reconstruction – simpler / easier-to-comprehend model**
  - Typically this is approximative and/or partial
  - “Model the model”, e.g., learn a NN model then learn a second simpler decision-tree model

- **Lateral relevance – interactivity for exploration**
  - E.g., within graphs: neighborhoods, centrality, top-k within categories

- **Conversational human-computer interaction (HCI) -- dialogue**
  - With previous knowledge of user (audience) knowledge and goals
  - And/or adaptively learn user knowledge and goals
Additional Aspects/Modifiers of explanation:

• Summarization, grain (coarse vs. fine), drill-down, elaboration
• Partial vs. complete
  – E.g., top-k in size among contributors to an aggregate sum
  – E.g., top-k in surprise/unexpectedness relative to some context
• Approximate vs. precise
  – E.g., present the top-3 weighted sub-models within an overall ML ensemble model
• Structuring of inference in presentation
  – E.g., present a clausal resolution graph or a Bayes net graph
• Assumptions and presumptions
• Targeting to user knowledge and goals (i.e., user model)
• Natural language (NL) generation
• Graphical presentation
• Terminology – e.g., ontologies, thesauri, vocabularies
Applications and Benefits of Explanation

- Semi-automatic decision support
- Fully-automatic decision making
  - E.g., in deep deduction about policies and legal
- Education and training, i.e., e-learning
  - E.g., Digital Socrates concept by Janine Bloomfield of Coherent Knowledge
- Accountability
- Knowledge debugging in KB development
- Trust in systems
  - Competence and correctness
  - Ethicality, fairness, and legality
- Human-machine interaction
  - User engagement
- Reuse / transfer of knowledge
Example Industry Leaders in Explanation Technology

• Coherent Knowledge – semantic logical deductive reasoning
  – Proof – natural deduction – in (declarative) extended logic programs
  – ... with NL generation + drill-down + interactive navigation + provenance

• Tableau Software
  – Specialized presentation of information via bar etc. charts

• Kyndi – cognitive search in NL knowledge graphs (KG)
  – Provenance + focus + lateral relevance
  – ... within extended KG derived from combining NLP + ML + KRR
Issues in the Field of Explanation Today

• Confusion about concepts
  – Esp. among non-research industry and media
  – But needs to be addressed first in the research community

• Mission creep, i.e., expansivity of task/aspect
  – Esp. among researchers. E.g., IJCAI-18 workshop on explainable AI.

• Ignorance of what’s already practical
  – E.g., in deep policy/legal deduction for decisions: full explanation of extended logic programs, with NL generation and interactive drill-down navigation
  – E.g., in cognitive search: provenance and focus and lateral relevance, in extended knowledge graphs

• Disconnect between users and investors
  – Users often perceive critical benefits/requirements for explanation
  – Investors (both venture and enterprise-internal) often fail to perceive value of explanation
Case Study: Automated Decision Support for Financial Regulatory/Policy Compliance

Problem: Current methods are expensive and unwieldy, often inaccurate

Solution Approach – using Textual Rulelog software technology:
• Encode regulations and related info as semantic rules and ontologies
• Fully, robustly automate run-time decisions and related querying
• Provide understandable full explanations in English
  • Proof: Electronic audit trail, with provenance
• Handles increasing complexity of real-world challenges
  • Data integration, system integration
  • Conflicting policies, special cases, exceptions
  • What-if scenarios to analyze impact of new regulations and policies

Business Benefits – compared to currently deployed methods:
• More Accurate
• Better explanations
• More Cost Effective – less labor; subject matter experts in closer loop
• More Agile – faster to update
• More Overall Effectiveness: less exposure to risk of non-compliance
Demo of Rulelog for Compliance Automation: US Federal Reserve Regulation W

- EDM Council Financial Industry Consortium Proof of Concept – successful and touted pilot
  - Enterprise Data Management Council (Trade Assoc.)
  - Coherent Knowledge (USA, Technology)
  - SRI International (USA, Technology)
  - Wells Fargo (Financial Services)
  - Governance, Risk and Compliance Technology Centre (Ireland, Technology)

- Reg W regulates and limits $ amount of transactions that can occur between banks and their affiliates. Designed to limit risks to each bank and to financial system.

- Must answer 3 key aspects:
  1. Is the transaction’s counterparty an affiliate of the bank?
  2. Is the transaction contemplated a covered transaction?
  3. Is the amount of the transaction permitted?

The Starting Point - Text of Regulation W

Determining Whether Regulation W Applies

Two initial questions need to be answered in determining whether a transaction is subject to Regulation W. The first is whether the transaction is between a bank and an “affiliate” of the bank. The second is whether the transaction is a “covered transaction.”

Affiliate Definition. Regulation W applies to covered transactions between a bank and an affiliate of the bank.

The definition of an affiliate for purposes of Regulation W is set forth in section 223.2. The definition is broad, and includes:

- Any company that controls the bank;
- Any company that is controlled by a company that controls the bank;
- Any company that is controlled, directly or indirectly, by trust or otherwise, by or for the benefit of shareholders who beneficially or otherwise control, directly or indirectly, by trust or otherwise, the bank or any company that controls the bank;
- Any company in which a majority of its directors, trustees, or general partners (or individuals exercising similar functions) constitute a majority of the persons holding any such office with the bank or any company that controls the bank;
- Any company, including a real estate investment trust, that is sponsored and advised on a contractual basis by the bank or an affiliate of the bank;
- Any registered investment company for which the bank or any affiliate of the bank serves as an investment adviser;
- Any unregistered investment fund for which the bank or any affiliate of the bank serves as an investment adviser, if the bank and its affiliates own or control in the aggregate more than 5 percent of any class of voting securities or more than 5 percent of the equity capital of the fund;
Reg W Demo Drill-down – outline

• Uses ErgoAI implementation of Rulelog

• Video available at
  http://coherentknowledge.com

• We show here: screenshots, example KB rules
Series of Advances → Rulelog’s Core Expressive Features

- Well-founded semantics; basic tabling algorithms
  - *Undefined* for paradox; smart cacheing; intuitionistic disjunction
- Higher-order syntax (Hilog); frame syntax
  - Associated optimizations of LP tabling etc. algorithms
- Statement id’s for meta; argumentation meta-rules for defeasibility; provenance
- General formulas with all usual classical connectives and quantifiers (omniformity)
- Restraint bounded rationality
  - Use 3rd truth value *undefined* for “don’t-care”
  - Radial, skipping; naf unsafety; external-query unsafety, unreturn


**Textual extension of Rulelog**

- High-level concept of approach: *Textual* Rulelog (TR)
- Extends Rulelog with natural language processing (NLP)
  - Start with English as the NL
- Rulelog logic (extended logic programs) itself is utilized to map:
  - Rulelog logic syntax $\leftarrow \rightarrow$ NL syntax
  - I.e., use logic to help do: *text interpretation* and *text generation*
- Mapping is much simpler and closer than with other KR’s
  - Rulelog’s high expressiveness is much closer to NL’s conceptual abstraction level
  - More often doable and useful:
    - 1 English sentence $\leftrightarrow$ 1 Rulelog sentence (rule)
- In principle, almost any NL sentence can be represented with deep semantics as a logical sentence in Rulelog
  - Leverage the general quantified formulas expressive feature of Rulelog
ErgoAI Architecture

Optional Custom Solutions

ErgoAI suite
- ErgoAI Studio (Integrated Development Env.)
  - GUI for explanation, editing, querying, debugging, tuning
- ErgoAI Reasoner
  - querying, compilation, caching, explanation generation, execution monitoring

Knowledge Base

queries, assertions

answers, explanations

Users

External Info (multi-source)
- Data
- Views, Rules
- Schemas & Ontologies
- Results of ML

External Services & Frameworks
- Relational DB
- RDF/Graph DB
- Other Sem. Tech
- Machine Learning
- Apps, Docker, ...

Complex Information
- English/NL Doc.'s etc.
- Policies, Regulations
- Financial, Legal, Medical, Science

Query is asked in English

![Query example](Coherent Query interface displaying a query and its result.

```
What proposed transactions are prohibited by RegW? Show (Bank, Company, Amount).
```

<table>
<thead>
<tr>
<th>Bank</th>
<th>Company</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>'Pacific Bank'</td>
<td>'Maui Sunset'</td>
<td>23.0</td>
</tr>
</tbody>
</table>

Why?
- Explanation Game
- See answer term as free
Example of Explanation in Rulelog

Why is the proposed transaction prohibited by Regulation W?

With drill down on: Why is the aggregate-affiliates limit $10 million?
User Clicks the handles to expand the Explanations
Why is the proposed transaction prohibited by Regulation W?

1. Is the transaction’s counterparty an “affiliate” of the bank?

   YES.

   And here’s why ...

   - RegW prohibits the proposed transaction by Pacific Bank with Maui Sunset of $23.0 million
   - The proposed transaction by Pacific Bank with Maui Sunset of $23.0 million is a RegW covered transaction
   - Maui Sunset is a RegW affiliate of Pacific Bank
   - Hawaii Bank is a RegW affiliate of Pacific Bank
   - There is common control of Hawaii Bank and Pacific Bank
   - Hawaii Bank is controlled by Americas Bank
   - Pacific Bank is controlled by Americas Bank
   - Pacific Bank is a subsidiary of Americas Bank
   - Maui Sunset is advised by Hawaii Bank
   - There is a proposed loan from Pacific Bank to Maui Sunset of $23.0 million
   - There is a limit of $10.0 million for any proposed RegW covered transaction by Pacific Bank with Maui Sunset
   - The proposed transaction of $23.0 million is greater than the RegW limit of $10.0 million
Concept of **Humagic Knowledge**

- Humagic = human-machine logic
- A humagic KB consists of a set of linked sentences
  - Assertions, queries, conclusions (answers & explanations)
- NL-syntax sentence may have 1 or more logic-syntax sentences associated with it
  - E.g., that encode it, or give its provenance, or represent its text interpretation
- Logic-syntax sentence may have 1 or more NL-syntax sentences associated with it
  - E.g., results of text generation on it
  - E.g., source sentence in text interpretation, that produced it
- Other sentences can be in a mix of NL-syntax and logic-syntax
  - Using textual templates, for text interpretation and text generation
Reg W Example Sentence using Templates

• Example of hybrid-syntax sentence – executable in ErgoAI:

\((The \text{ individual affiliate threshold for transaction under Regulation W by } \text{ ?Bank} \text{ with } \text{ ?Counterparty} \text{ is } \text{ ?Amount}) \ :
l
\((\text{ ?Counterparty is deemed an affiliate of } \text{ ?Bank} \text{ under Regulation W}) \ \text{ and } \n\((\text{ ?Bank has capital stock and surplus } \text{ ?Capital}) \ \text{ and } \n\text{ (the threshold percentage for an individual affiliate is } \text{ ?Percentage} \) \ \text{ and } \n\text{ ?Amount is } \text{ ?Capital} \times \text{ ?Percentage/100.})
Reg W example Template definition in ErgoAI

template(headbody,
   \(\text{The proposed transaction } ?\text{Id by } ?\text{Bank with } ?\text{Affiliate of } $?\text{Amount is a RegW covered transaction}\),

   covered(proposed(transaction))(by(?Bank))(with(?Affiliate))
   (of(amount(?Amount)))(having(id(?Id)))
).

• The templates are self-documenting
Examples of the Underlying Textual Rulelog

Executable **Fact** Assertions

- subsidiary(of)('Pacific Bank','Americas Bank').
- advised(by)('Maui Sunset','Hawaii Bank').
- bank('Hawaii Bank').
- company('Maui Sunset').
- capital(stock(and(surplus)))('Pacific Bank',2500.0).
- proposed(loan) (from('Pacific Bank'))(to('Maui Sunset')) (of(amount(23.0))) (having(id(1101))).
- previous(loan)(from('Pacific Bank'))(to('Hawaii Bank')) (of(amount(145.0))) (having(id(1001))).
- proposed(asset(purchase))(by('Pacific Bank')) (of(asset(common(stock)(of('Flixado'))))) (from('Maui Sunset')) (of(amount(90.0))) (having(id(1202))).
Executable Assertions: non-fact Rules

/* A company is controlled by another company when the first company
   is a subsidiary of a subsidiary of the second company. */
@!{rule103b} /* declares rule id */
@@{defeasible} /* indicates the rule can have exceptions */
controlled(by)(?x1,?x2)
:- /* if */
   subsidiary(of)(?x1,?x3) \and
   subsidiary(of)(?x3,?x2).

/*A case of an affiliate is: Any company that is advised on a contractual basis by
 the bank or an affiliate of the bank. */
@!{rule102b} @@{defeasible}
affiliate(of)(?x1,?x2) :-
( advised(by)(?x1,?x2)
 \or
 (affiliate(of)(?x3,?x2) \and advised(by)(?x1,?x3))).
Executable Assertions: **Exception Rule**

@!{rule104e}
@{‘ready market exemption case for covered transaction'} /* tag for prioritizing */
\neg covered(transaction)(by(?x1))(with(?x2))
  (of(amount(?x3)))(having(id(?Id))) :-
  affiliate(of)(?x2,?x1) \and
  asset(purchase)(by(?x1))(of(asset(?x6)))(from(?x2))(of(amount(?x3)))
  (having(id(?Id))) \and
  asset(?x6)(has(ready(market))).

/* prioritization info, specified as one tag being higher than another */
\overrides(‘ready market exemption case for covered transaction',
  'general case of covered transaction').

/* If a company is listed on the New York Stock Exchange (NYSE), then the common stock of that company has a ready market. */
@!{rule201} @@{defeasible}
asset(common(stock)(of(?Company)))(has(ready(market))) :-
  exchange(listed(company))(?Company)(on('NYSE')).
Executable Assertions: Import of OWL

:- iriprefix fibof = /* declares an abbreviation */
   "http://www.omg.org/spec/FIBO/FIBO-Foundation/20120501/ontology/".

/* Imported OWL knowledge: from Financial Business Industry Ontology (FIBO) */
rdfs#subClassOf(fibob#BankingAffiliate, fibob#BodyCorporate).
rdfs#range(fibob#whollyOwnedAndControlledBy, fibob#FormalOrganization).
owl#disjointWith(edmc#Broad_Based_Index_Credit_Default_Swap_Contract,
    edmc#Narrow_Based_Index_Credit_Default_Swap_Contract).

/* Ontology Mappings between textual terminology and FIBO OWL vocabulary */
company(?co) :- fibob#BodyCorporate(?co).
fibob#whollyOwnedAndControlledBy(?sub,?parent) :- subsidiary(of)(?sub,?parent).

/* Semantics of OWL - specified as general Rulelog axioms */
?r(?y) :- rdfs#range(?p,?r), ?p(?x,?y).
?p(?x,?y) :- owl#subPropertyOf(?q,?p), ?q(?x,?y).
Outline of Talk

• Observations
  – Concepts of Explanation
  – Applications & Benefits of Explanation
  – Example Industry Leaders in Explanation Technology
  – Issues Today in the Field of Explanation

• Case Study
  – Proof style explanation in
    Financial Regulatory/Policy Compliance
OPTIONAL SLIDES FOLLOW
Rulelog: Software Tools

- Lots of Rulelog expressiveness:
  - Ergo Lite: Large subset of Rulelog. Open source.
    - A.k.a. as Flora-2, originally
  - ErgoAI (from Coherent): Most of Rulelog. Has IDE.
    - The most complete & highly optimized implementation available. Free for academic research use. Free trials available for everyone. (Support time may cost, tho’.)

- Much smaller subsets of Rulelog expressiveness:
  - XSB Prolog: Most of LP -- with functions and well-founded negation. Plus a bit more. Open source.
  - Jena: Function-free negation-free LP, focused on RDF. Plus a bit more. Open source.
    - Similar: misc. other, e.g., that implement SWRL or SPIN
ErgoAI: Reasoner, Studio, Connectors

- ErgoAI Reasoner has sophisticated algorithms & data structures
  - Smart cacheing with dependency-aware updating. Leverages LP & DBMS techniques.
  - Transformation, compilation, subgoal reordering, indexing, modularization, dependency/loop analysis, performance monitoring/analysis, pausing, virtual machine, programming kernel, external import/querying
  - Java API. Other interfaces: command line, web, C. Additional APIs for Python, REST, more.
  - Scales well: Millions of sentences on 1 processor; Trillions on distributed nodes

- Explanations: for every answer; interactively drill down tree; in NL
  - Rulelog enables natural deduction style proofs, automated NL generation via rules

- ErgoAI Studio is a graphical Integrated Development Environment
  - Interactive editing, querying, visualization of knowledge
  - Fast edit-test loop with award-winning advanced knowledge debugging/monitoring

- ErgoAI Connectors federate knowledge & reasoning
  - Import/query dynamically via: SPARQL, OWL, RDF; SQL; CSV; JSON; more
  - Federation distributes reasoning (i.e., its processing) across multiple nodes

- Open, standards-based approach; a portion is open source
  - Rulelog is draft industry standard from RuleML (submission to W3C & Oasis)
Concept: Virtual Data Stores

- Rulelog orchestrates overall federated reasoning by sub-goaling dynamically.

- A variety of other (“external”) structured information systems are treated as virtual data/knowledge stores, via Rulelog federation connectors, which import/query and translate.

- Each virtual data/knowledge store is treated as having expressiveness that is a subset of Rulelog. E.g.,
  - An external database fact is treated as a logical atom in Rulelog.
  - An OWL axiom is treated as a fact but also supplemented by semantic axioms about OWL’s constructs in general.
Kinds of Virtual Data in ErgoAI

- Graph databases: via SPARQL/RDF connector
  - Description logic ontologies: via OWL connector
- Relational databases: via SQL connector
- Spreadsheets and web logs: via DSV connector
- JSON connector; XML connector; Web services via those
- Extensible to almost any kind of (semi-)structured info
  - E.g., Machine Learning (ML) and NLP systems
    - Represent \text{prob(content\_sentence, lower\_bound, upper\_bound, confidence\_level, statistical\_procedure)} as an ErgoAI sentence
  - E.g., legacy applications in Java
    - Get\_foo method is treated like a query
- Importing RDF & OWL knowledge into ErgoAI

- Screenshot of ErgoAI OWL connector part of ErgoAI Studio

Translates RDF & OWL to ErgoAI

Define IRIs in ErgoAI Studio

N-triples and N-quads

RDF/OWL XML, JSON-LD, or Turtle as input. Predicate or Frame syntax output.
Some Key Tasks in Textual Rulelog

- **TR text interpretation**: Rulelog rules map from NL to logic
- **TR text generation**: Rulelog rules map from logic to NL
- **TR terminology mapping**: Rulelog rules map between phrasings and ontologies – in NL or logic
  - “moving a bomb” implies “transporting weaponized material”
  - isBomb(?x) implies rdftriple(?x,rdftype,bomb)
Some Techniques for Textual Rulelog

1. Word as functor (WAF): treat a NL word as a logical functor

2. Phrasal terms (phrasts): treat a NL phrase as a logical term

3. Phrasal mapping from paraphrase knowledge (PMK): e.g., synonyms, hypernyms, hyponyms, equivalent named entities; other implications b/ phrases

4. Textual templates (TET): hybrid of text syntax and logic syntax

5. Quantification of NL determiners (QUD): e.g., treat “every” and “some” as relativized universal and existential logical quantifiers

6. Deep extended NL parsing (DEP): logically represent dependency parse tree extended with coreference analysis and named entity recognition

• These can be combined
  – Many interesting directions & open areas for research! E.g., DEP and QUD.
END OF OPTIONAL SLIDES