Scaling Challenges in Explanatory Reasoning

(#paper33-langley)

Pat Langley

Institute for the Study of Learning and Expertise
Palo Alto, California

Mohan Sridharan

School of Computer Science
University of Birmingham, UK

Thanks to B. Meadows, P. Bello, and W. Bridewell for contributions to this research, partly funded by ONR Grant No. N00014-17-1-2434 and N00014-10-1-0487.

A Motivating Example

Humans understand many social interactions with little effort. Consider a simple example:

- Suppose we hear that *Abe has some cash* and *Bob has a car*.
- We also hear that, later, *Abe possesses the same car*.

We do not observe any transaction, but we can assume one took place. Two reasonable explanations come to mind:

- *Abe bought the car from Bob using his money*.
- *Abe stole the car from Bob by threatening him*.

We also know these two explanations are mutually exclusive.

Later, we may hear Abe gave money to Bob, eliminating theft as an alternative. We want a theory of such reasoning ability.

Target Abilities

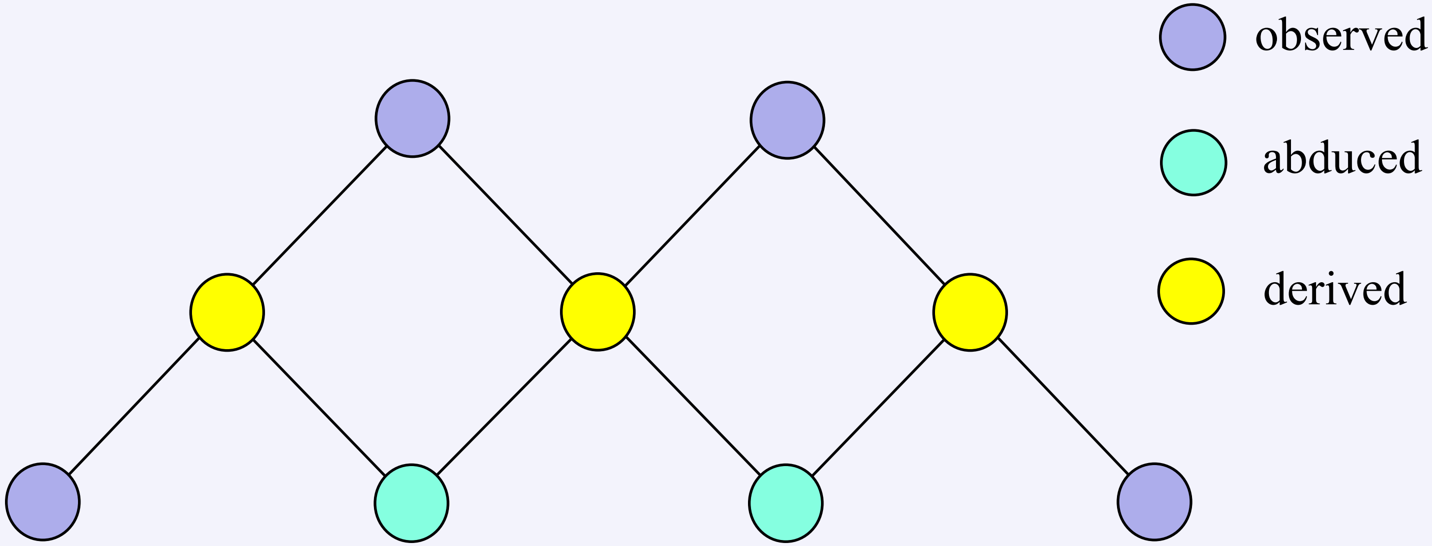
We can identify five abilities humans exhibit when they make observations:

- Explain these events by connecting them through knowledge.
- Introduce plausible assumptions about unobserved events.
- Incorporate observations into explanations incrementally.
- Detect inconsistent beliefs and address these conflicts.
- Generative alternative explanations of these observations.

These are distinctive features of human intelligence and thus natural targets for cognitive systems research.

Traditional Formulations of Abduction

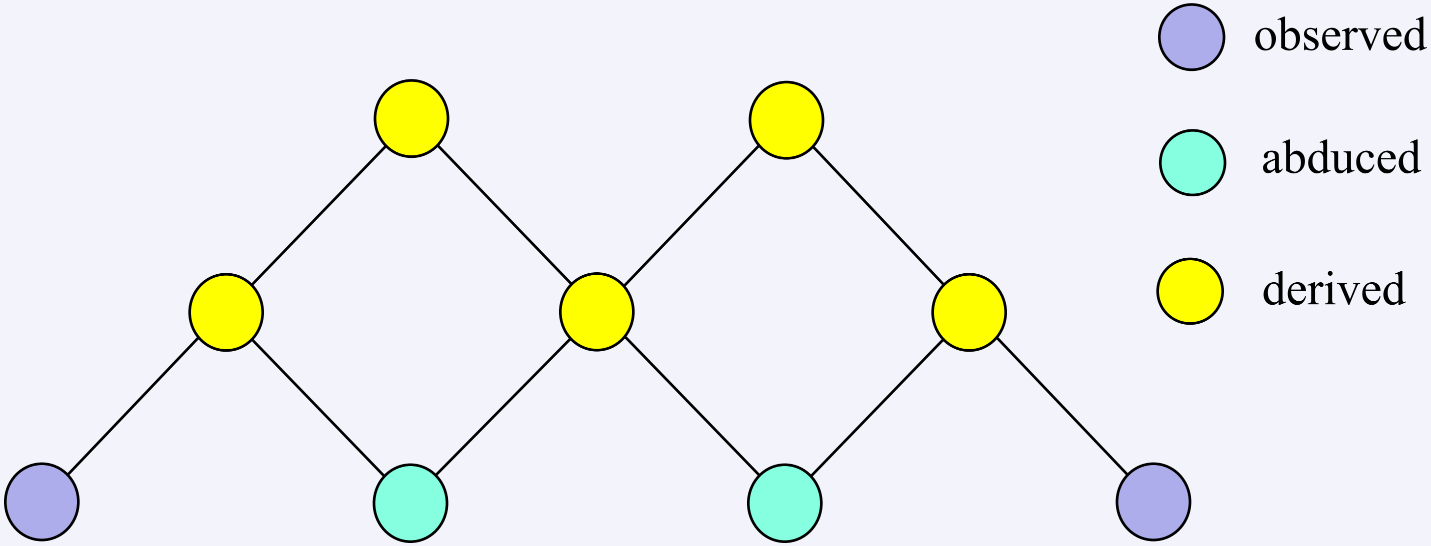
Classic treatments of abduction construct proof graphs with observations as *roots* and assumptions as *terminal nodes*.



We refer to this as *derivational abduction* because observations must be derived from other beliefs.

A Different Formulation of Abduction

Another framework for abduction constructs proof graphs with both observations and assumptions *only* as terminal nodes.



We refer to this as *associative abduction* because observations are explained if they hang together, as in ‘guilt by association’.

A Theory of Associative Abduction

Our theory of associative abduction (Langley & Meadows, 2019) incorporates:

- *Structural* postulates (5): representation and organization of explanations.
- *Processing* postulates (3): mechanisms that generate and revise explanations.

This theory comprises postulates about cognitive structures and their interpretation.

We also have a system that instantiates this theory, but they are conceptually distinct.

R1: Two Types of Knowledge

The theory posits two complementary types of knowledge:

1. *Definitions* specify high-level predicates as conjunctions of simpler ones.
   a. High-level definition for “purchasing” or “robbery”; low-level rules for transferring property.
   b. Similar to organization in logic program, context-free grammar.

2. *Constraints* specify relations that are mutually exclusive.
   a. Cannot buy and steal an item!
   b. Indicate inconsistency when satisfied jointly.

Definitions are *generative*, while constraints are *restrictive*.

R2: Three Types of Beliefs (Dynamic Memory)

There are three different kinds of short-term mental elements:

1. *Observed beliefs*, which come from external perceptions.
   a. Observed Abe with car, so Abe has possession of car.

2. *Abduced* beliefs, which are introduced as assumptions (from unmatched antecedents of definitions).
   a. Abe bought or stole the car!

3. *Derived beliefs*, which are deduced from other beliefs using knowledge (from the consequents of definitions).
   a. Abe gave money to Bob, so Abe bought car.

Beliefs take the form of ground literals, predicates with zero or more arguments; possibly skolems (invented symbols).

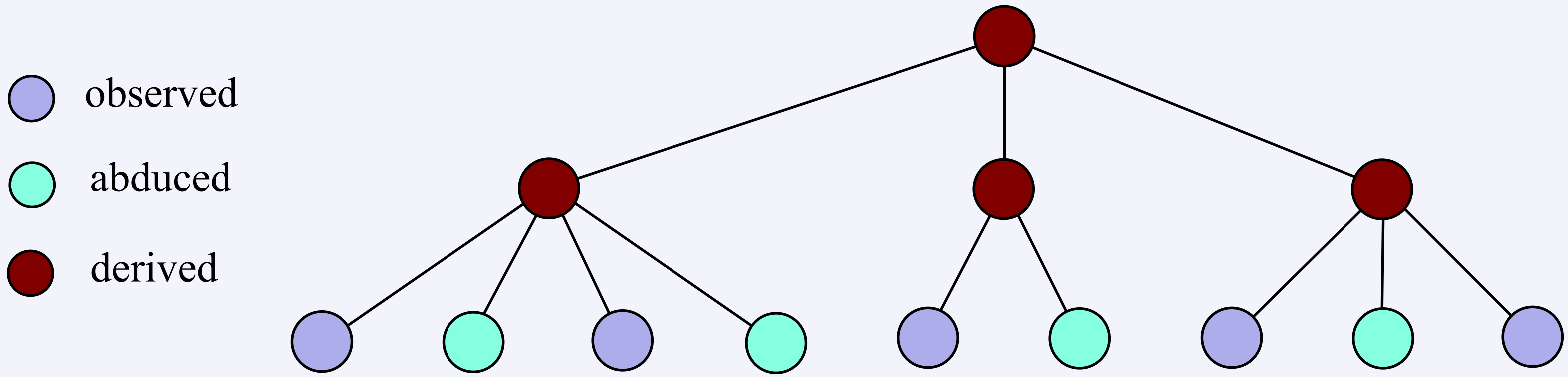
R3: Structure of Explanations

Justifications (instances of applied definitions) are organized into higher-level explanations. An *explanation* is a connected proof graph with four elements:

1. A set of *observed* beliefs *O* to be explained (terminal nodes)
2. A set of *abduced* (assumed) beliefs *A* (terminal nodes)
3. A set of *derived* beliefs *D* that follow from *O* and *A*
4. A set of *justifications* that show how D follows from *O* and *A*

E.g., parse trees; observed words are terminal nodes, non-terminal nodes derived, different parses have different justifications.

An explanation may have more than one derived root node, but it must be *connected*.

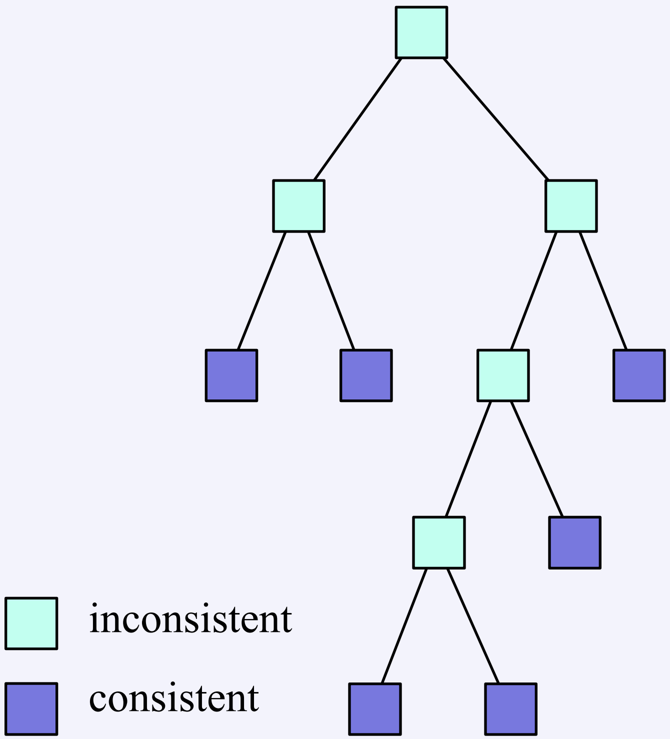


Observations are *terminal* nodes, not *root* nodes, as in most abduction work.

R4: A Tree of Possible Worlds

- Explanations are stored as sets of justifications and beliefs called *worlds*.
- Justification can contribute to competing accounts, e.g., two parses of a sentence share subtrees, each associated with multiple worlds.
- Worlds organized in a phylogenetic tree that traces their evolution.
- Root node: initial set of beliefs. Each child omits some elements from its parent world to sidestep an inconsistency.
- Terminal nodes denote worlds (potentially) consistent with observations and knowledge.
- Closed worlds: known constraint violations; Active worlds: (frontier) internally consistent.

Siblings in world tree offer competing explanations of observations.



R5: Distributed Representation

- Beliefs are stored in *one working memory*, with each element specifying worlds in which it does ***not*** hold.
- Alternative worlds are encoded in a distributed manner: takes advantage of shared observations, abductions, derivations.
- Avoids repeating the same inferences during reasoning, which supports an implicit form of parallelism.
- Storing worlds where beliefs do not hold reduces memory load, provided elements held in common are in the majority.
- Serves as a heuristic measure that has no guarantees but is often effective.

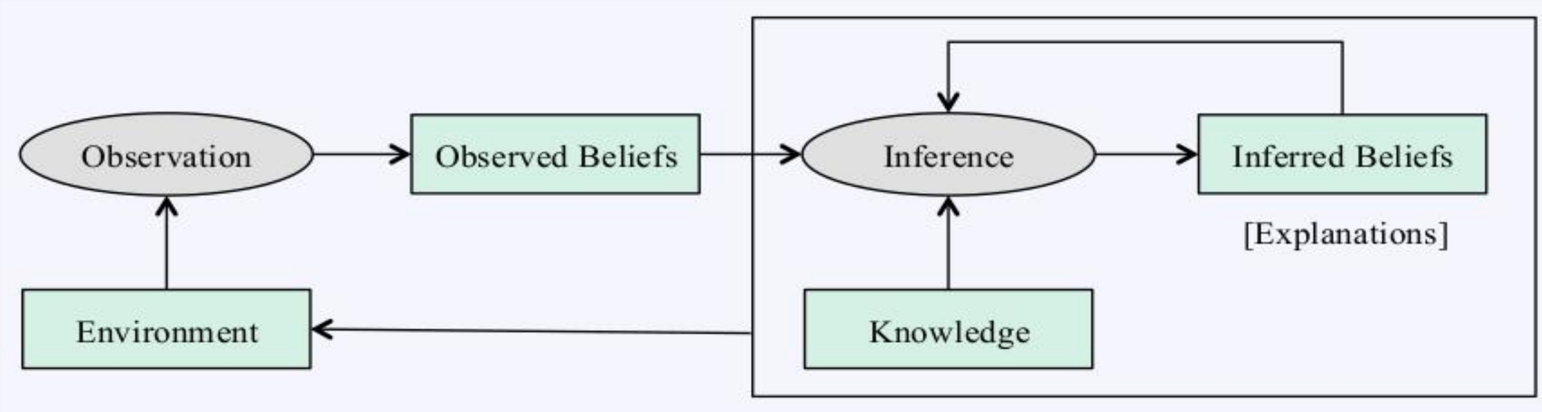
P1: Incremental Processing

Explanation process alternates between two cognitive cycles:

1. *Observation* (outer) loop accepts inputs from the environment.
   a. E.g., vision, language, produces new *observed* beliefs.

2. *Inference* (inner) loop extends and revises explanations.
   a. Repeatedly select focus belief, invoke definitions to elaborate explanations, use constraints to detect+repair inconsistencies.
   b. Focus belief determines relevant knowledge; antecedent unifies with it.

Produces *derived* beliefs and *abduced* beliefs; constructs explanations *incrementally* and *bottom-up*.



P2: Two Varieties of Inference

Explanation relies on two forms of inferential processing:

1. *Elaboration* involves applying a conceptual definition.
   a. Produces new belief based on the rule's head (*deduction*).
   b. Adds assumptions if some antecedents are absent (*abduction*).

2. *Repair* detects a violated constraint (B1 / B2) and eliminates it:
   a. Deactivates each world W with the conflict, generates one child of W with B1 and another with B2.
   b. New worlds retain beliefs from ancestors not responsible for, or implied by, removed beliefs

Inference alternates between elaborating worlds (*monotonic*) and spawning worlds to fix inconsistencies (*non-monotonic*).

P3: Focus of Attention

Explanation construction is aided by knowledge but driven by observations obtained incrementally.

Multiple accounts of observed fact possible; search through explanations consistent with data.

Explanatory inference relies on focus of attention to provide *heuristic guidance*:

- In each cycle, select belief F (observed, derived, or abduced) to focus on.
- During elaboration and repair, only consider definitions and constraints with antecedents that unify with F.

Worlds encoded in distributed manner:

- Each inference step can elaborate/repair worlds that share belief.
- ‘Spreading activation’ in which one idea leads to others, ‘stream of consciousness’.

This mechanism makes retrieval / matching tractable but can overlook useful inferences and inconsistencies.

The PENUMBRA System

Embedded ideas in PENUMBRA, an architecture for explanatory inference that operates incrementally.

Like most cognitive architectures, this one comes with:

- A *syntax* for knowledge elements and working memory.
- An *interpreter* that operates over these structures.

PENUMBRA offers a programming language that incorporates theoretical assumptions about the mind.

The system shares many features with UMBRA (Meadows et al., 2014), an earlier system for abductive explanation.

Scalability Analysis: Analytical, Empirical

Parameterize performance using variables:

- Processing times of inference cycle stages: select focus, check constraints, select definition.
- Relevant factors and independent counts.

Analytical computation of costs of inference (in paper) provides hypotheses.

- Focus belief selection time $T_F = j \,.\, N_B$
- Constraint checking time $T_C = i \,.\, C_P \,.\, (A_C - 1) \,.\, B_P$
- Definition selection time $T_D = k \,.\, D_P \,.\, (B_P + 1)^{(A\_D - 1)}$

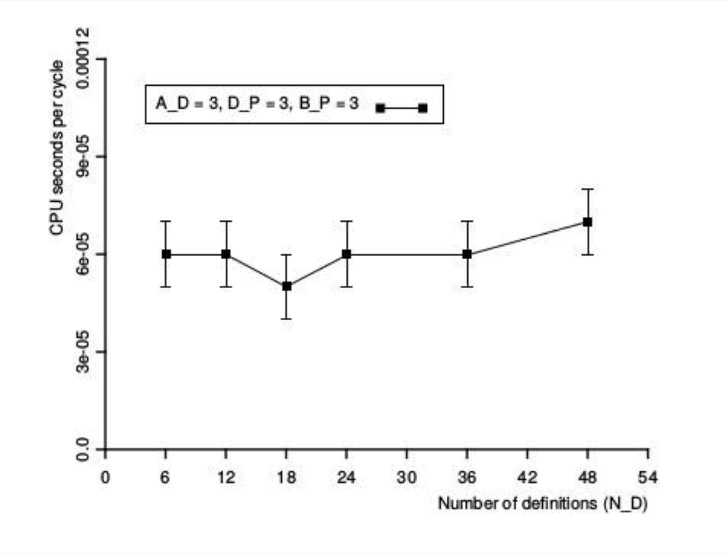
Begin by empirically evaluating cost of selecting definition for elaboration:

- More expensive than other steps.
- Use synthetic datasets (see paper).

Scalability Analysis: Analytical, Empirical

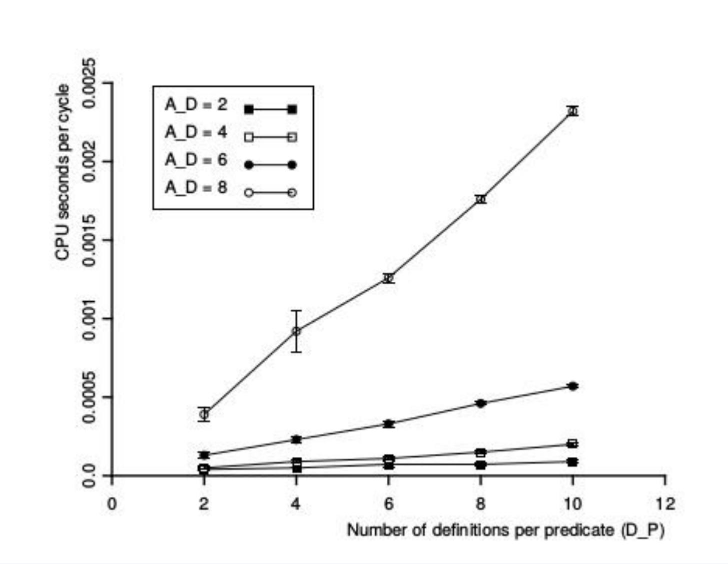
Processing time per definition selection is independent of number of definitions.

A_D = no. of antecedents/definition;
D_P = no. of definitions/predicate;
B_P = no.of beliefs/predicate;
Varied N_D = no. of definitions.



Processing time per definition selection is linear function of average number of definitions per predicate.

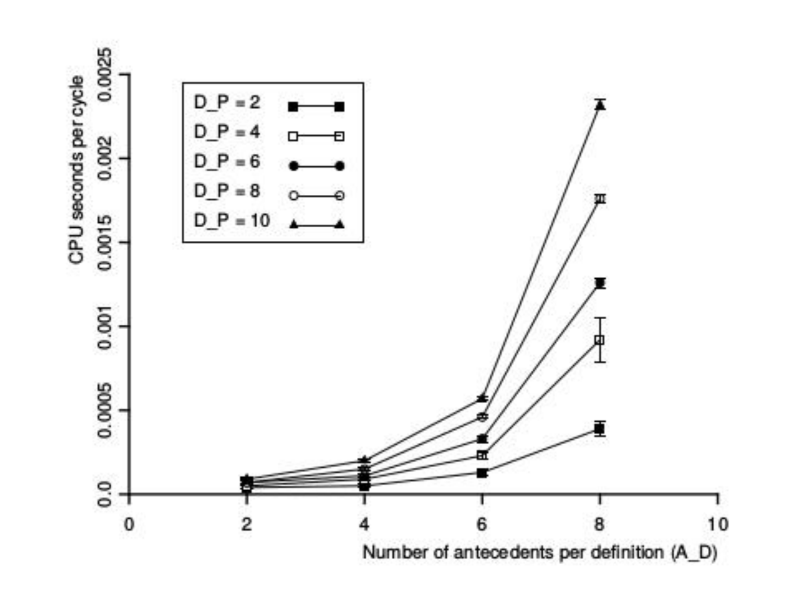
A_D = no. of antecedents/definition;
D_P = no. of definitions/predicate;
$B_P = 3$;



Scalability of Rule Selection: Continued...

Processing time per definition selection is exponential function of average number of antecedents per definitions.

A_D = no. of antecedents/definition;
D_P = no. of definitions/predicate;
B_P = 3;



Experimental studies of definition selection consistent with analytical calculations:

- Processing time grows slowly with N_D, D_P, and B_P.
- Exponential in A_D due to need to consider partial matches; bound by limiting antecedents per rule=>hierarchical organization of such knowledge.

Scalability Analysis: Explanation Construction

Full explanation needs to be scalable.

Explore scalability to number of alternative explanations; human language processing indicates use of effective heuristics to guide choices.

PENUMBRA heuristics: focus belief selection, definition selection.

Hypothesis: *Given effective heuristics, time to find best explanation independent of no. of consistent worlds*.

"Best" explanation?

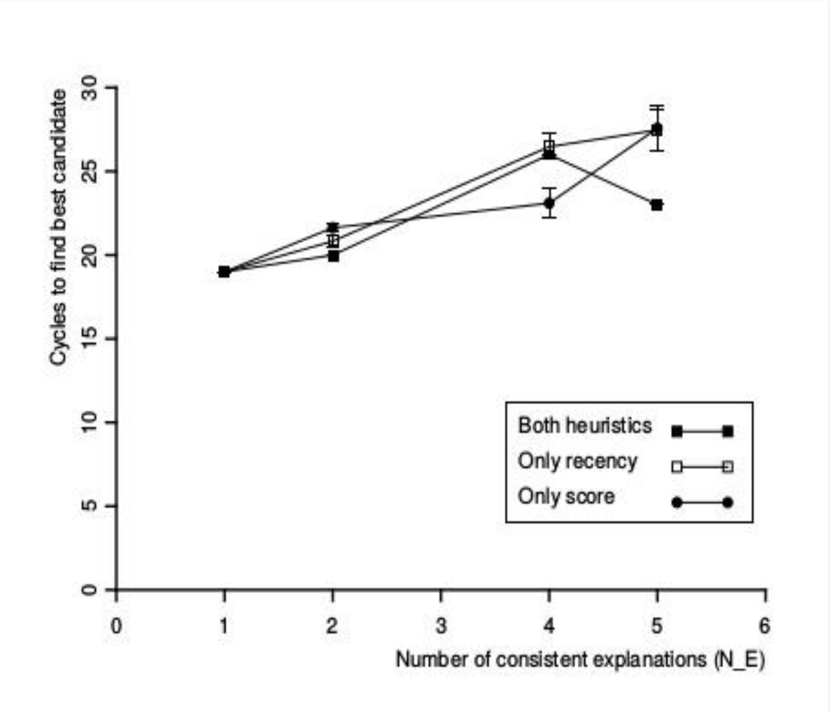
- Simplicity, coherence, *summed weights of assumptions*, probability of parse trees.
- Use variant of Hobbs et al. (1993); select recent beliefs, rules with higher scores.
- *Depth-first search through space of explanations, apply definitions that elaborate on most promising world before others*. May occasionally take you down wrong path!

Scalability Analysis: Best Explanation First?

- Consider no. of observations to be explained, complexity of explanations.

- Sentence parsing task.
- Parses map to explanations; terminal nodes (words), root node (root of explanation); different parses set up to have different scores.
- English syntax (subset) as CFG.

- Cycles to find best parse, as a function of number of consistent explanations (i.e., parses).



- Compare with random selection of beliefs and/or rules; should not work well.
  Results not quite as expected: need better heuristics?

Related Research

Our explanatory inference approach borrows ideas from prior work:

- *Explanation relies on abduction that posits plausible assumptions*
  - Gordon (2018), Molineaux et al. (2012), Friedman et al. (2018)
- *Incremental associative abduction guided by focus of attention*
  - Bridewell and Langley (2011), Meadows et al. (2014)
- *Encoding alternative situations by associating beliefs with worlds*
  - Fahlman (2011), Bello (2012)
- *Nonmonotonic repair of inconsistencies via truth maintenance*
  - de Kleer (1986), Doyle (1979)

Our approach builds on these traditions, but combines them in novel ways to explain the explanation process.

Concluding Remarks

Computational account of explanatory inference:

- Two forms of knowledge, three types of dynamic beliefs organized as linked justifications associated with one or more worlds
- Three mechanisms: focus attention, apply definitions, repair constraint violations.

An implemented version of the theory in PENUMBRA.

Scalability analysis:

- Analytical computation of computational costs; empirical evaluation with synthetic data.
- Processing time scales well except antecedents per definition; *we can bound this*.
- Qualitative hypotheses about ability to find best account before alternatives.
- Heuristics for selecting focus beliefs and definitions (to be applied) need to be improved.

In future research, we plan to explore:

- Scalability to more complex problems and large databases.
- Other criteria for explanation quality and heuristics, e.g., probabilities for alternative accounts, explanatory coherence, other heuristics for selecting focus beliefs and rules.

These will provide a fuller account of everyday explanation.

That’s all folks!