Understanding Social Interactions by Incremental Abductive Inference

Pat Langley

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

Department of Computer Science
University of Auckland, Auckland, NZ

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

Social Understanding

Humans understand many social interactions with little effort; we easily generate hypotheses about:

- Other agents' beliefs and goals
- Their beliefs and goals about others' mental states
- Their awareness / ignorance of the true situation
- Even their intentions to deceive third parties

Such abilities are a distinctive feature of human intelligence and thus a natural target for AI research.

Social Understanding in Fables

Aesop-like fables present an interesting variation on the task of social understanding:

> *The Lion and the Sheep.* A lion is too old to hunt animals for prey. The lion announces he is ill. The sheep, believing he is harmless, follows social convention and visits the lion's cave to pay respects to the sick. The lion kills and devours him.

Such stories are usually brief, focus on goal-directed behavior, and center on high-level social interaction.

Explanations of fables often revolve around agents' beliefs and goals about others' beliefs and goals.

Some Related Paradigms

The task of social understanding is related to a number of other research paradigms, including:

- Activity recognition (e.g., Aggarwal & Ryoo, 2011)
- Plan recognition (e.g., Goldman, Geib, & Miller, 1999)
- Behavior explanation (e.g., Malle, 1999)
- Collaborative planning (Rao, Georgeoff, & Sonenberg, 1992)
- Story understanding (e.g., Wilensky, 1978; Mueller, 2002)

These differ in important ways, but we have incorporated some of their key ideas into our work.

The Task of Social Understanding

We can formulate the problem of social plan understanding as:

- *Given:* A sequence *S* of observed actions by agent *A*
- *Given:* Knowledge of situations and activities available to *A,* including *social* actions and their effects
- *Infer:* An explanation, *E*, organized as a proof lattice, that accounts for *S* in terms of *A*'s goals, beliefs, and intentions

We assume that content has already been translated from language or vision into some internal representation.

Theoretical Tenets

Our computational theory of social understanding postulates that this process:

- Involves the *abductive* generation of *explanations* that introduce default assumptions
- Involves inference about the participating agents' *mental states* (beliefs / goals about situations and activities)
- Operates in an *incremental* manner to process observations that arrive sequentially
- Proceeds in a *data-driven* fashion because it draws on observations about agents' activities

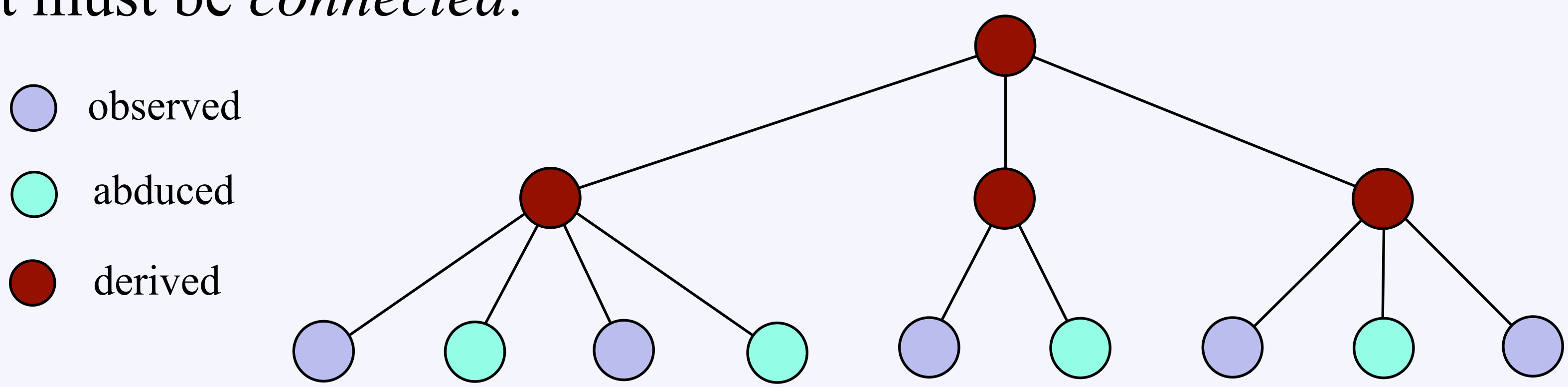
These assumptions place strong constraints on our account of this important capability.

The Structure of Explanations

An explanation is a connected proof graph with four elements:

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

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



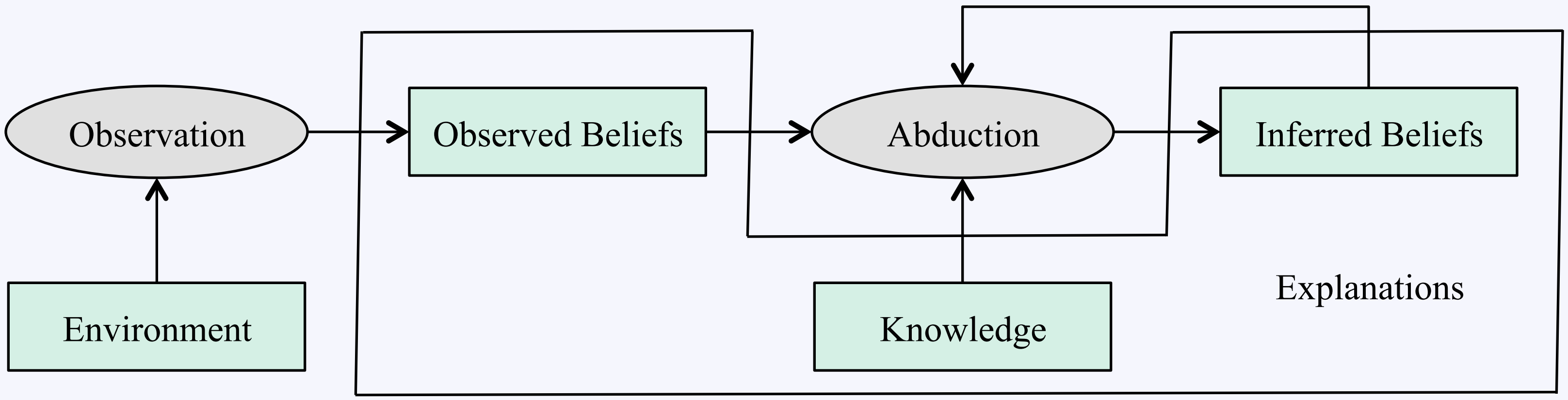
Observations are *terminal* nodes, not *root* nodes, as in most abduction work.

Incremental Construction of Explanations

The explanation process alternates between two core activities:

- Accepting inputs from the environment (vision, language)
  - This process produces new *observed* beliefs
- Elaborating these observations through abductive inference
  - This process produces *derived* and *abduced* beliefs

This two-level cycle constructs explanations incrementally.



The UMBRA System

In previous work (Meadows et al., 2014), we developed UMBRA, an abduction system that:

- Accepts observations and adds them to working memory
- Extends an explanation incrementally:
  - Finds rules with antecedents that unify with memory elements
  - Tentatively completes each rule instance's missing antecedents
  - Selects the rule instance R with best evaluation score
  - Adds R's inferred elements to memory as default assumptions
  - Repeats this process to make a number of inferences
- Continues until no further observations arrive

UMBRA's aim is to produce a *coherent explanation* in terms of its available knowledge.

Early Results with UMBRA

In prior work, we tested UMBRA on disaster-response scenarios from the Monroe corpus (Blaylock & Allen, 2005), such as:

A work crew and a lines crew are both working at the location Park Ridge. There is a fallen pine tree at Park Ridge. The work crew is a tree crew. The work crew cut the pine tree.

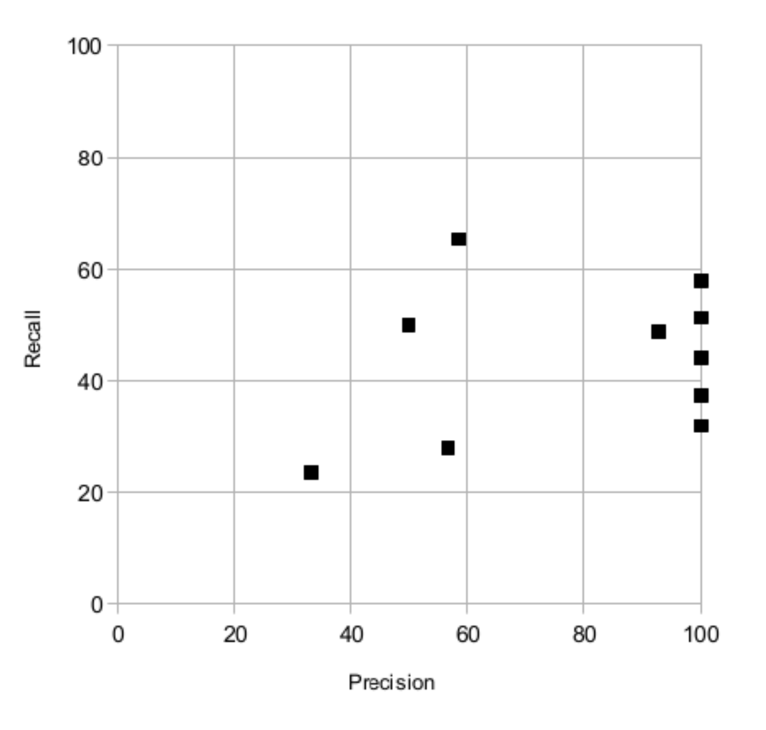
Each scenario was encoded as a set of logical literals including the target hierarchical plan.

- However, we presented the system only with observed actions and with partial state descriptions.

We also provided UMBRA with an HTN with 51 methods for 38 tasks, 41 actions, and 55 conceptual definitions.

Early Results with UMBRA

UMBRA reconstructs much of the plan structure from the input traces, even when only a fraction of agents' actions are observed



Precision and recall declines gracefully as fewer observations are available to the system.

Input	Precision	Recall
Batch 100%	79.8 ± *8.3*	44.1 ± *4.3*
Batch 75%	79.4 ± *9.8*	41.4 ± *4.9*
Batch 50%	70.4 ± *11.5*	29.0 ± *4.7*
Batch 25%	55.2 ± *12.6*	16.4 ± *2.3*
Incremental	70.6 ± *10.2*	53.6 ± *4.3*

Limitations of UMBRA

Despite these promising results, UMBRA had limited abilities for social understanding, in that it could not:

- Represent agents' models of others' beliefs and goals
- Specify the times at which these elements actually held
- Encode knowledge about social activities and their effects
- Reason about different levels of embedded beliefs / goals

In response, we have extended the system's representations and processes to address these drawbacks.

Extension 1: Timing and Constraints

To support social understanding, we augmented the formalism to incorporate:

- Start and end times for each belief and goal:
  - belief(lion, prey(sheep), 6:00, s1)
  - goal(lion, healthy(lion), 12:00, 12:30)
- Constraints on timing and equality:
  - constraint(fox, between(s2, s4, 8:00), 5:35, 6:00)
  - constraint(lion, nequal(sheep, s3), 5:00, s2)

Constraints appear as first-class structures at the same level as beliefs and goals.

Extension 2: Embedded Structures

The extended UMBRA represents agents' mental states, some with embedded structures:

- belief(fox, has(crow, grapes, 09:30, s1), 09:31, s2)
- goal(crow, acquire_edible_food(crow, s3, s4))
- belief(snake,
  belief(lion, at_location(lion, river, 09:00, s5), 09:02, s6), 09:02, s7)
- belief(snake,
  goal(fox, trade(crow, fox, grapes, grain, 09:40, s8), 09:30, s9),
  09:30, s10)
- goal(lion, belief(sheep, sick(lion, 09:00, 24:00), 09:45, s12), 09:00, s13)

Embedded structures appear in working memory and social rules, but *not typically in domain-level knowledge*.

Extension 3: Inference Processes

These representational changes required some extensions to our inference mechanisms:

- Introducing start times for inferences based on the current cycle
- Inferring timing and equality constraints when a rule fires
- Using constraints to eliminate rules that would create inconsistent default assumptions
- Reasoning over embedded beliefs and goals using rules that have non-embedded structures

We did not alter the basic abduction mechanism to operate over social knowledge, despite its more abstract character.

Empirical Claims About UMBRA

We make three claims about the extensions to UMBRA that let it support social understanding:

- The system generates reasonable explanations for fables from partial information
- Applying knowledge at different levels of embedding is critical to this ability
- Abstract knowledge about social interactions is also essential to this functionality

We designed and carried out experiments designed to test these three claims.

Tests of Social Understanding

We devised eight fables that require social understanding at different levels of complexity:

- *Nested understanding:* UMBRA interprets an agent's mental states and/or plan based on observed behavior.
- *Deeply nested understanding:* The system infers an agent's inferences about another agent's mental states.
- *Inferring mistakes:* The program infers that an agent has mistaken beliefs, its reasons, and how they differ from the true account.
- *Reasoning about opportunism:* UMBRA understands how an agent has capitalized upon another's false beliefs.
- *Reasoning about deception:* The system infers that an agent engenders false beliefs in another agent to achieve its goals.

We then used these scenarios to test UMBRA's ability to construct social explanations.

Tests of Social Understanding

We created a knowledge base for these scenarios that includes:

- About 60 distinct operators and methods
  - alternative decompositions
  - many with overlapping conditions
  - only ten percent used in any ‘correct’ fable explanation
  - about 500 domain-level conditions, excluding constraints
- About 100 distinct domain-level predicates

Domain knowledge describes physical situations and activities at a single level of embedding.

Social knowledge uses embeddings to support reasoning about others’ mental states.

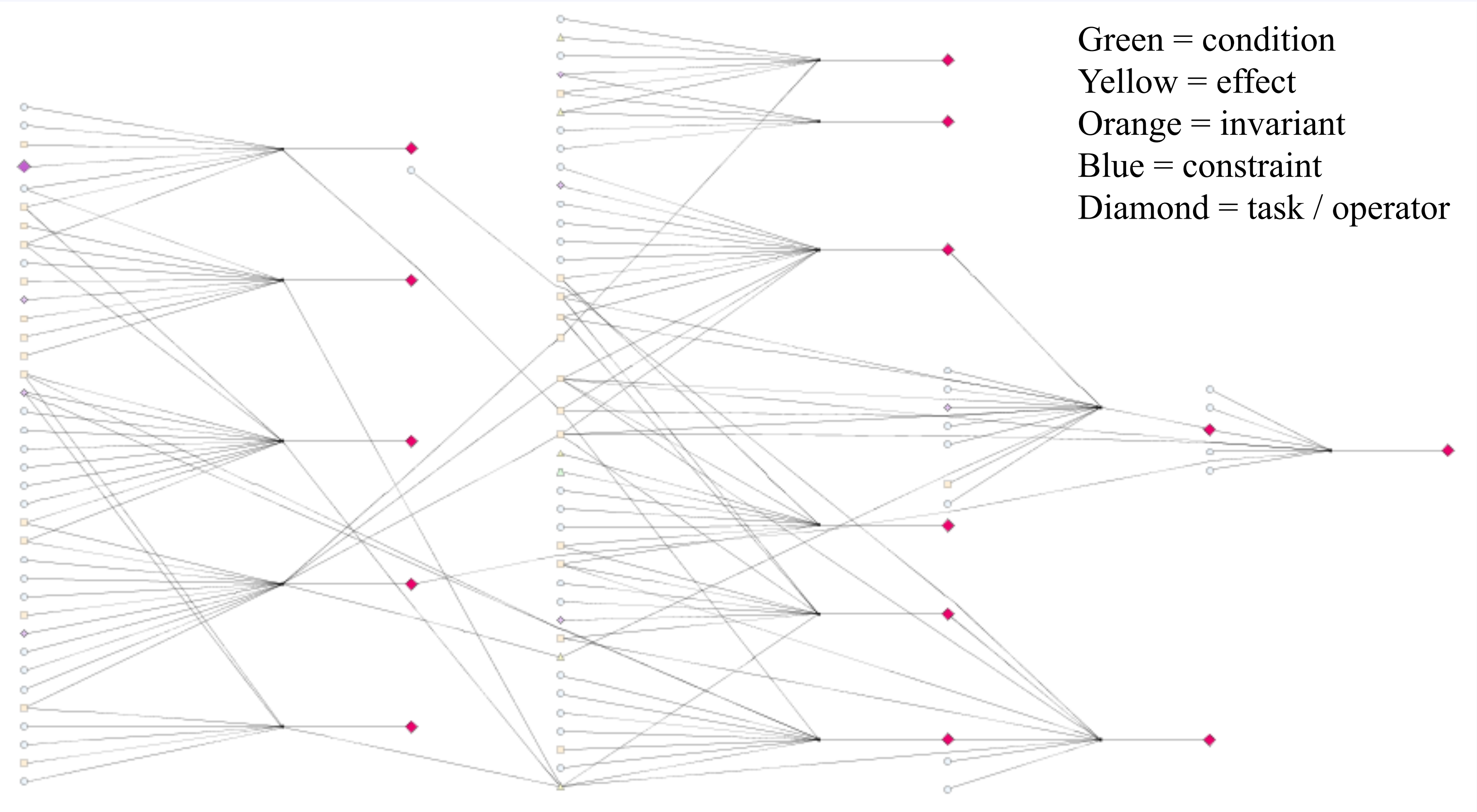
Social Operators

UMBRA's social knowledge includes 13 operators that describe personal interactions:

- announce_genuine, announce_wrong, announce_false
- interpret_as_real, interpret_as_real_agent, interpret_as_real_attributed
- interpret_as_image, interpret_as_image_attributed
- become_jealous
- judge_not_a_threat
- pretend_attribute
- suggest_trade_good_faith, suggest_trade_bad_faith

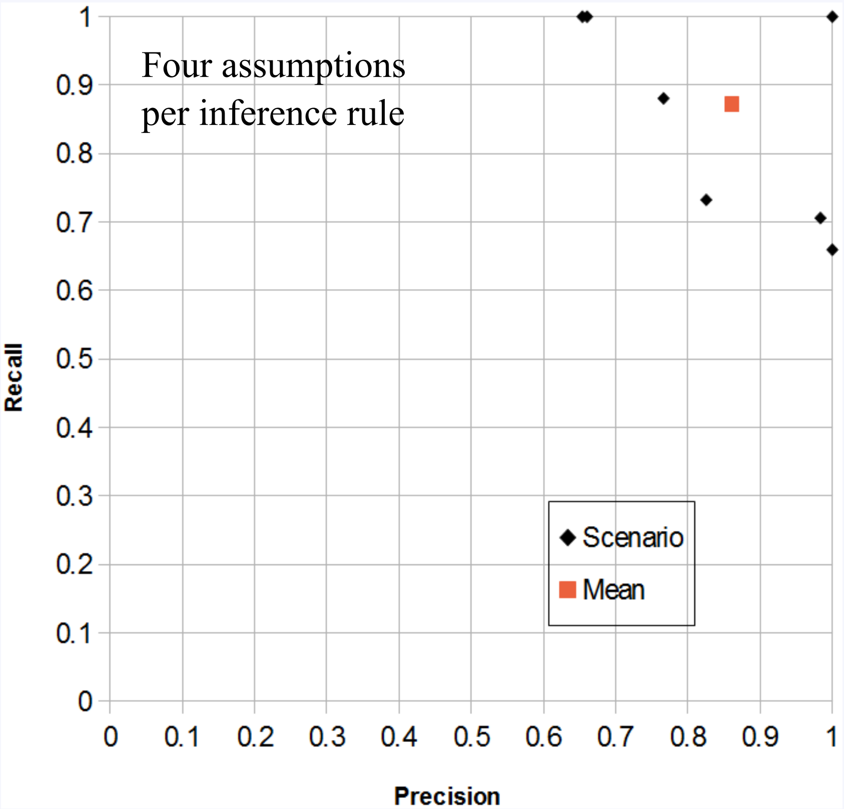
Each operator describes an activity that alters the mental states of the participating agents.

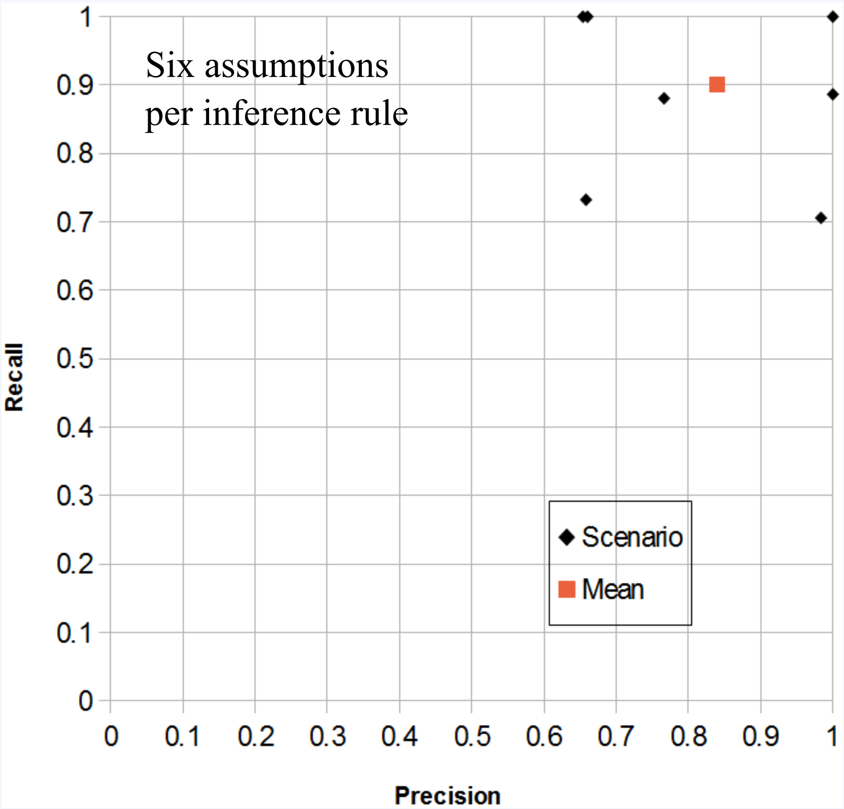
Structure of a Fable Explanation



Basic Results on Fable Understanding

The extended UMBRA draws correct inferences with high precision and recall from story facts (less than 40 percent of target explanations).

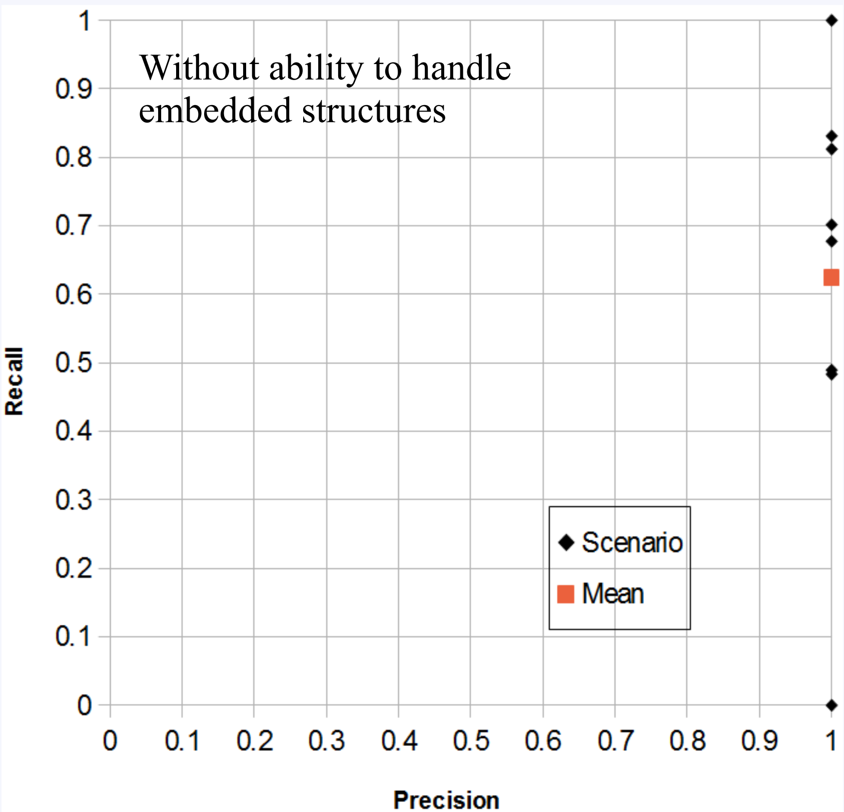


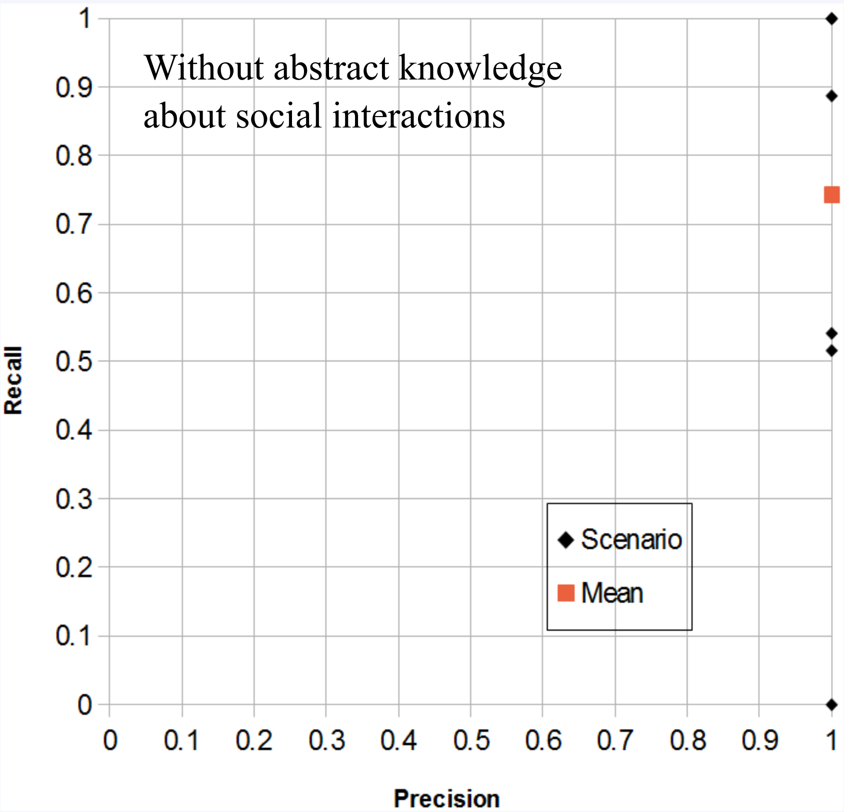


Changes to the system’s parameters have little effect on these scores.

Results from Lesion Studies

We also ran UMBRA with its ability to handle embedded structures and its social knowledge removed.





Even when given all terminal literals, recall was still reduced greatly.

Related Research

Our approach relies on three assumptions that have been explored in prior research:

- *Social cognition relies on representing and reasoning about other agents' mental states.*
  - Fahlman (2011), Bello (2012), Bridewell and Isaac (2011)
- *Plan understanding involves incremental abduction that constructs an explanation of observed inputs.*
  - Ng and Mooney (1990), Bridewell and Langley (2011)
- *Social understanding depends on general knowledge about social interactions and their effects on mental states.*
  - Wilensky (1978), Winston (2012)

Our work borrows ideas from these traditions, but combines them in novel ways to support social understanding.

Ongoing Research

One of UMBRA's key drawbacks is its reliance on greedy search through the space of explanations.

In response, we developing a successor system that instead:

- Considers and elaborates on multiple explanations
- Detects when an explanation has inconsistent beliefs
- Generates revised accounts that are internally consistent
- Organizes alternative accounts in a tree of possible worlds

The system keeps UMBRA's incremental, data-driven, abductive approach to social understanding, but it should be more robust.

Concluding Remarks

We have extended UMBRA, which constructs explanations with an incremental form of abductive inference, to:

- Represent other agents' mental states as embedded structures
- Encode information about timing and constraints
- Store domain-independent knowledge about social interactions
- Reason over this content to understand Aesop-like fables

Experiments suggest that our approach can create plausible and coherent social explanations from partial information.

End of Presentation