A POSSIBLE WORLDS MODEL OF BELIEF FOR STATE-SPACE NARRATIVE PLANNING

Alireza Shirvani, Stephen G. Ware, Rachelyn Farrell

Presented by Shane Racey

INTRODUCTION

Traditional State Space Planners

- Model the world using a single truth-based state
- Characters act with perfect knowledge or a shared "public" belief

How agents act based on their (possibly wrong) beliefs and how their beliefs change is an essential part of many stories

- Treat the search space of the problem as a map of temporally possible worlds
- Add epistemically possible worlds to represent what agents believe

TREASURE ISLAND EXAMPLE

Jim Hawkins acquires a treasure map but must enlist the help of the pirate Long John Silver to get it.

Silver believes the treasure does not exist, but Hawkins spreads a rumor causing him to believe it is on the island.

Silver takes Hawkins to the island in the hopes of claiming the treasure for himself, but Hawkins digs it up and makes off with it.

NARRATIVE PLANNING PROBLEM

We define a narrative planning problem as ⟨s0,g,A,C⟩ ,

- s0 is the initial state
- g is a goal expression
- A is a set of actions
- C is a set of characters

For every action a∈A, we define CON(a) to be the set of characters ∈C who must consent to take that action. We define OBS(a) to be the set of characters ∈C who observe the action when it occurs

NARRATIVE SPACE

State Space: A graph whose nodes are states and whose edges are actions

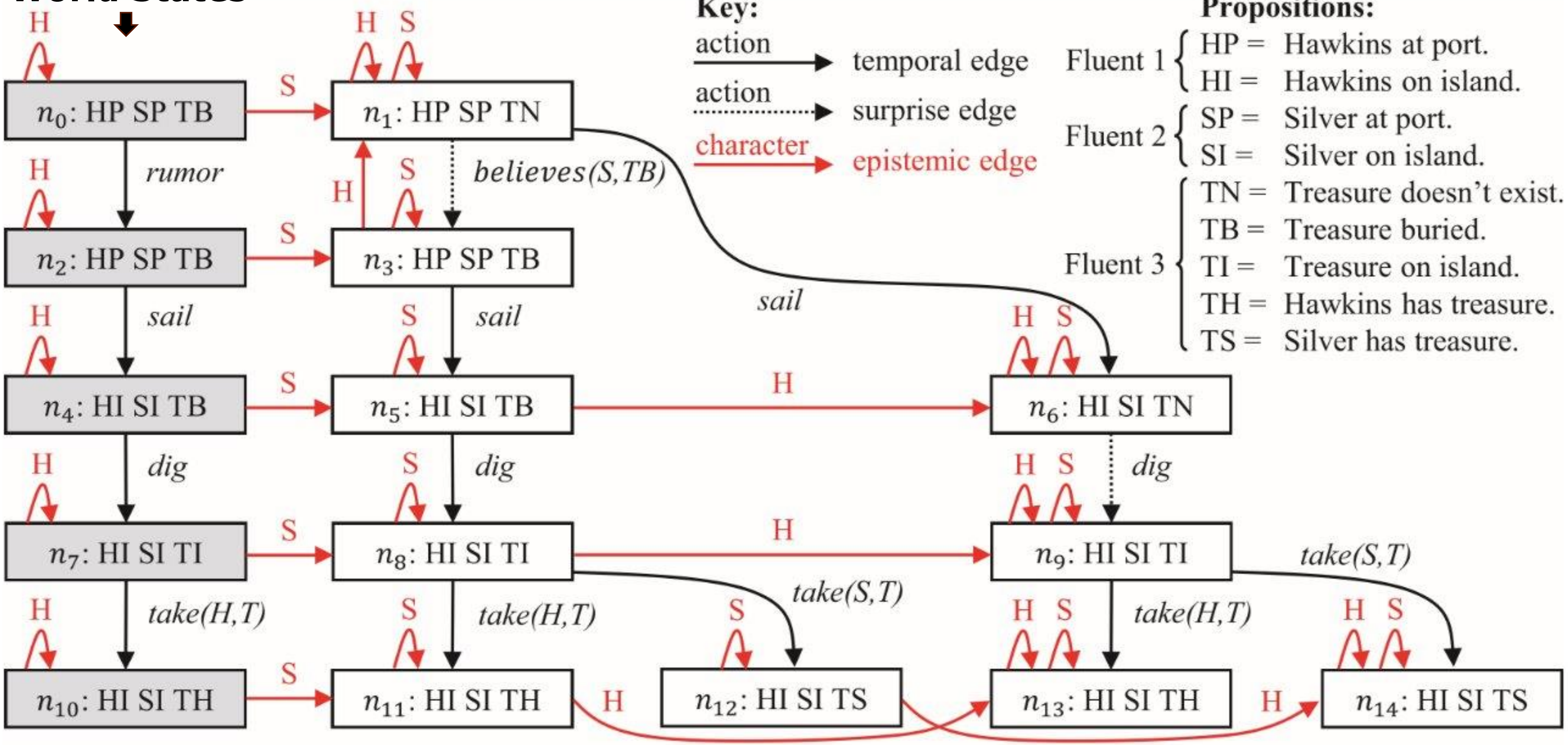
Temporal Edges: The set of possible future worlds (drawn in black)

- Has precondition PRE(a) and effect EFF(a) regarding propositions

Epistemic Edges: Exists for every character (drawn in red)

- When a world is in state n1, character c believes that world to be in state n2

World States



Initial State:
HP SP TB *b*(S,TN)
Author's Goal: TH
Hawkins' Goal: TH
Silver's Goal: TS

Actions:

	rumor	*believes(S,TB)*	*sail*	*dig*	*take(H,T)*	*take(S,T)*
PRE:	TB	*false*	HP SP	HI TB	HI TI	SI TI
EFF:	*b*(S,TB)	TB	HI SI	TI	TH	TS
CON:	H	*none*	S	H	H	S
OBS:	H	S	H S	H S	H S	H S

NARRATIVE SPACE

b(c, p) represents whether character c believes proposition p

- The expression b(H, b(S, TB)) means "Hawkins believes that Silver believes the treasure is buried." (n2 –H> n2)(n2 –S> n3)

This belief-aware framework introduces two key changes:

- States are no longer defined solely by the world's physical facts but also by characters' beliefs
- Actions can occur even when a character believes the preconditions are false, leading to surprise edges that represent unexpected events from a character's perspective

EXPANDING THE NARRATIVE SPACE

Initialization

- Graph composed of only the node for s0
- Closed world assumption + characters believe other characters believe the same things they do, unless otherwise stated

$$\forall c \in C\,(b(c,p) \rightarrow \forall d \in C\,(b(c,\beta(d,p))))$$

procedure EXPAND(a, n) ▷ To compute $\alpha(a, n)$.
  if $\alpha(a, n)$ does not exist **then**
    Let n_* be a new node.
    $\forall c \in C$ let $\beta(c, n_*) = \beta(c, n)$.
    $\forall o \in \text{OBS}(a)$ let $\beta(o, n_*) = \alpha(a, \beta(o, n))$.
    Let $\alpha(a, n) = n_*$.
    Impose EFF(a) on n_*.
  end if
end procedure

Algorithm 1: Adding a temporal edge to the search space.

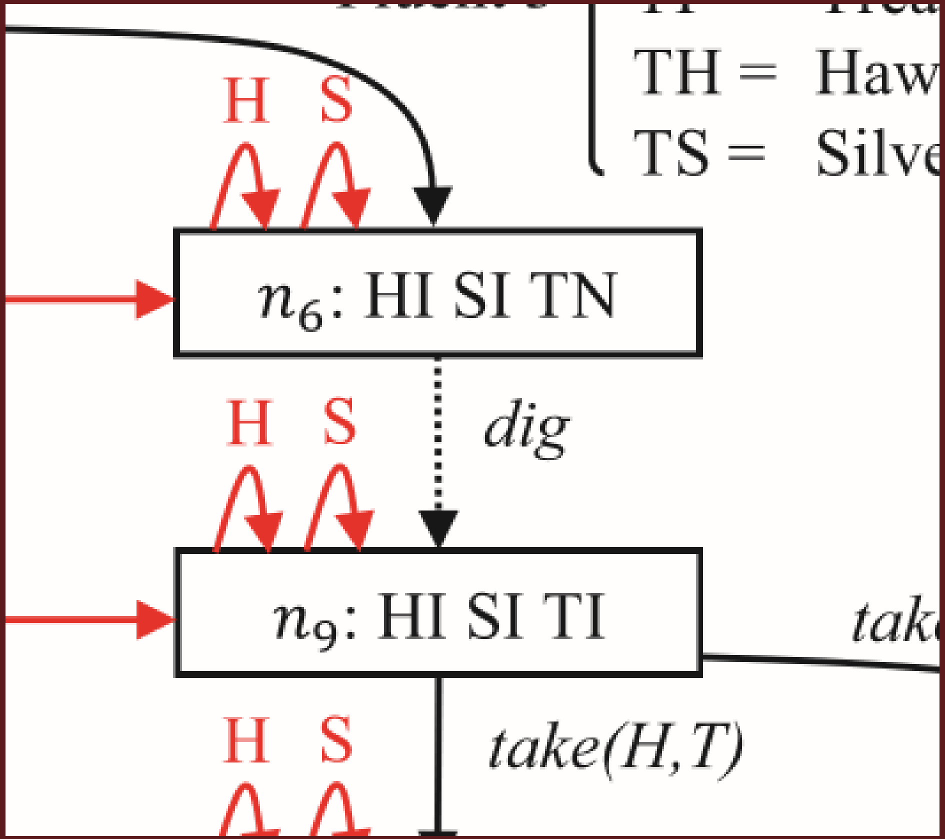
Expanding the Nodes

- Expands the space by adding temporal edges until reaching a state where g holds
- Example: add temporal edge (n4 -dig> n7)
- Compute a(dig, n5) and a(dig, n6) recursively to update the epistemic edges

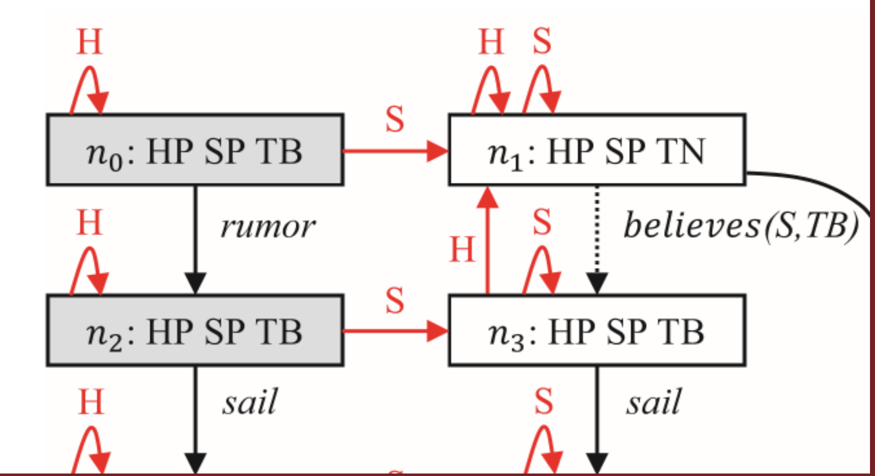
IMPOSING EFFECTS

Imposing a surprise action

- Update beliefs to allow the action to be possible
- Update their beliefs based on the action's effects
- Example: (n6 -dig> n9)



IMPOSING EFFECTS



Imposing a belief effect

- To impose the belief b(c, p), create a dummy action d such that PRE(d)=false, EFF(d)=p, CON(d)=∅, and OBS(d)=c
- Example: adding edge (n0 –rumor> n2)
- To impose the effect b(S, TB) on n2, we create the dummy action believes (S, TB) and impose it on β(S, n2)

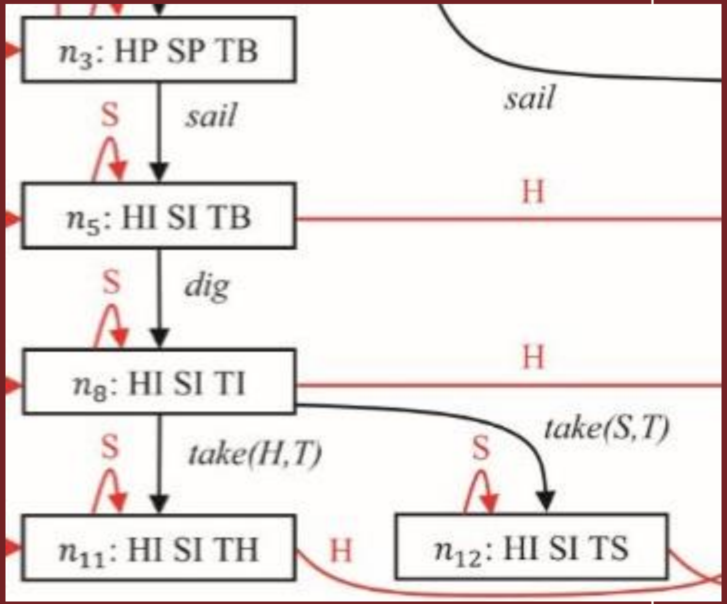
VALID NARRATIVE PLANS

- Every character must believe that every action they take will contribute to one of their goals

Chains

- Causal Chain: Sequence of causally linked actions and propositions
- Intentional Chain: A causal chain where…
  - Character intends final goal
  - No surprise actions or negations
  - Not all actions require character consent (relaxation)

○ n3→n5→n8→n12



S achieves TS via $\langle sail, HI, dig, TI, take(T,S), TS \rangle$

EVALUATION

Model has two important advantages

- Accurately simulates how people process beliefs
- Does so more accurately than models lacking nested beliefs

Other models to compare to (layers of belief):

- 0 Layers - All agents are omniscient, and characters never make mistakes
- 1+1 Layers – Agents share a belief state that updates on publicly observable actions
- 2+ Layers - Each character has their own belief state and uses epistemic edges to track beliefs per agent, including false beliefs and nested beliefs

EXAMPLE STORIES

To evaluate the 2+ model, we used it to produce three stories representing different narrative elements including deception, cooperation, anticipation, and surprise

The Most Wanted (18 questions)

1. Sheriff William wants to trick a local gunman named Jack. The Sheriff loads some money from the town bank into a wagon.
2. Sheriff William goes to the saloon to inform Jack that the money is in the wagon.
3. Sheriff William returns to the bank and takes the money out of the wagon.
4. Sheriff William climbs into the wagon himself.
5. The wagon sets off to the desert. Jack follows it and holds it up at gunpoint. He opens the wagon to find Sheriff William waiting inside to arrest him.

What does Jack believe is in the wagon?
What does Jack believe William believes?
What does William believe Jack believes?

Homecoming (24 questions)

1. Ed finds out about a party at school and invites his friend Mike.
2. Mike invites Jenny to the party because he has a crush on her.
3. Ed discovers the party is actually at a fraternity house.
4. Ed calls Mike to inform him the party is at the fraternity house.
5. Mike does not have Jenny's phone number, so he goes to school, waits for Jenny to arrive, and then informs her the party is at the fraternity house.

Where does Mike believe the party is?
Where does Jenny believe the party is?
Where does Mike believe Jenny believes the party is?
Where does Ed believe Mike believes the party is?

The Forty Thieves (24 questions)

1. Alibaba overhears forty thieves tell their boss, Jafar, of a treasure hidden in a cave.
2. Alibaba goes to the cave, gets the treasure, and takes it to his house.
3. Jafar goes to the cave to find the treasure missing. The thieves search the town and find it in Alababa's house. Alibaba notices the thieves placing a mark on his house.
4. The thieves report to Jafar that the treasure is in Alibaba's house, which is marked.
5. Alibaba removes the mark from his house and marks the town`s guardhouse. Jafar and the thieves arrive in town and see the mark.
6. Jafar and the thieves break into the guardhouse and are arrested.

EXPERIMENTAL DESIGN

Goal:

Evaluate whether the 2+ belief model (with nested beliefs) aligns better with human understanding than simpler models

Method:

1. Read one of three stories, presented step-by-step (via temporal edges)
2. After each step, answer multiple-choice questions about what characters believe (1–2 layers deep)
3. After the story, answer 5 comprehension questions
4. Only responses with all comprehension questions correct were included (to filter noise)

RESULTS

		0		1+1	
		Correct	Incorrect	Correct	Incorrect
2+	Correct	6	18	15	9
	Incorrect	0	0	0	0

(a) Homecoming

		0		1+1	
		Correct	Incorrect	Correct	Incorrect
2+	Correct	11	9	11	9
	Incorrect	0	0	0	0

(b) The Forty Thieves

		0		1+1	
		Correct	Incorrect	Correct	Incorrect
2+	Correct	6	9	8	7
	Incorrect	2	1	0	3

(c) The Most Wanted

Table 1: Contingency tables for the 0, 1+1, and 2+ models

Reliability of Human Judgments (Krippendorff's α):

- Homecoming: 0.79
- The Forty Thieves: 0.70
- The Most Wanted: 0.20 (lower reliability)

Model and Accuracy

- 2+ (nested beliefs): 95%
- 1+1 (shared belief): 54%
- 0 (omniscient agents): 40%

The 2+ belief model aligned far more closely with how human participants understood character beliefs—especially for nested beliefs, deception, and surprise

FURTHER NOISE REDUCTION

Disagreements often stemmed from the initial state of the story especially in The Most Wanted. Misunderstanding the start affects belief tracking throughout

Additional Filtering Step:

- Removed participants who misunderstood the initial state
- Excluded initial state questions from the accuracy analysis
- Discarded 5 more belief questions (9%) due to lack of consensus
- No questions had multiple correct answers

Updated Model and Accuracy

- 2+ (nested beliefs): 100%
- 1+1 (shared belief): 49%
- 0 (omniscient agents): 44%

FORMAL EVALUATION

The 2+ belief model generates a superset of stories produced by the 0 and 1+1

- Planning without belief is a special case of planning with belief

Expressive power of 2+

- 2+ explains actions driven by false beliefs
- 0 and 1+1 models cannot generate such narratives
- Builds on validated models of intentionality
- If A believes B could and would act, it's rational for A to anticipate that action

THANK YOU