# Sabre: A Narrative Planner Supporting Intention and Deep Theory of Mind

Stephen G. Ware, Cory Siler 2021

# Why narrative planners struggle

- Emergent simulations → believable agents, weak plot control
- Classical planners → strong plot control, wooden agents
- Calls for a system that hits author goals and keeps every character rational

# Sabre in a nutshell

- Forward-chaining state-space planner
- Supports full ADL + numeric fluents
- Unlimited theory of mind (nested beliefs)
- Utility-driven intention model for author + all characters
- Every action must be explained in the actor’s belief world
- Solves benchmark stories up to 836 actions / 896 triggers

# Problem representation

- Characters: Entities that appear to reason and act
- Fluents: State variables
- Literals: Basic facts, belief literals
- Events: Character taking an action
- Pre-conditions & effect conditions:  When an event can fire, what its branches look like
- Utility functions: Numeric preferences for the author and each character

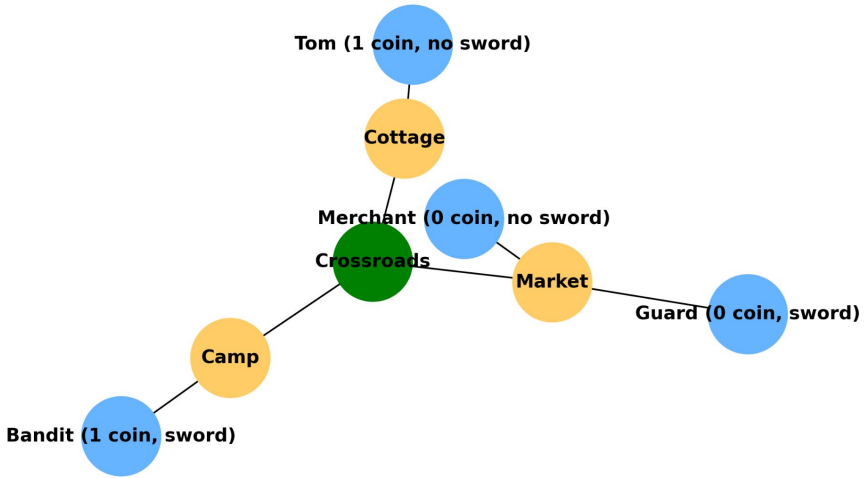
# Problem representation

- Characters: Tom, Merchant, Guard, Bandit
- Fluents: #coins, has sword
- Belief: b(Tom, loc(Merchant)=Market)
- Utility functions:

loc(Tom)=Cottage ^ loc(medicine)=Tom → 2

alive(Tom)=⊥ → 1

⊤ → 0



- Events
  - Action: Chose by the planner
  - Trigger: Fires instantly
  - Surprise: Observe while not believing

- Explained Actions
  - When the character believes the action can lead to increasing their utility

# Believed State

- Maintaining a believed state for each character
  - β(c,s): The state that **c** believes to be the case when real state is **s**
  - Triggers must be checked not only in the real world but in every β(c,s) as well

- Dealing with triggers and beliefs

**Original trigger (fires in real world)**

$loc(guard)=market \land loc(bandit)=market \land b(guard,loc(bandit)\neq market)$

**Effect**: $\top \rightarrow b(guard,loc(bandit)=market)$

**Copied trigger (fires only in β(Tom, s))**

$b(Tom,loc(guard)=market) \land b(Tom,loc(bandit)=market) \land b(Tom,b(guard,loc(bandit)\neq market))$

**Effect**: $\top \rightarrow b(Tom,\ b(guard,\ loc(bandit)=market))$

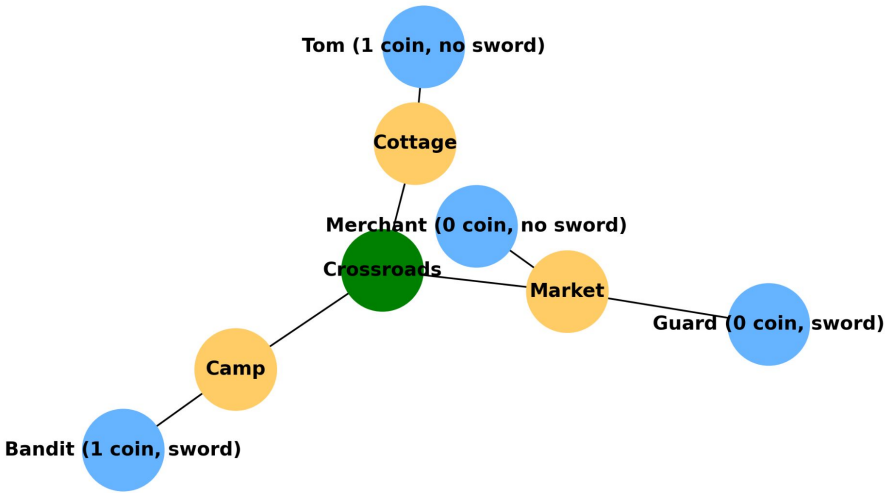
# State Space

- Compact history-based representation
  - Sabre stores the sequence of events that led to the state
- σ(e,s): Single-event transition
  - is defined only if the pre-condition PRE(e) holds in s
- α: Apply event and all mandatory triggers
  - α(e,s) equals:
    - Undefined, if σ(e,s) undefined
    - α(t,σ(e,s)), if some trigger t now enabled
    - σ(e,s), otherwise

# Solution

- A plan is a solution iff
  - Executable from the initial state
  - Raises author utility
  - Every action is explained in the state where it occurs
  - No redundant actions included

# Example: Get a potion for grandmother




1. Author starts in the initial state s0
2. Chooses the first action (Tom walks to the crossroads)
3. Runs a SABRE sub-call
4. Action passes → it is appended; world becomes s1
5. Repeat for the second walk to the market
6. Third action is buy: two consent sub-calls
7. After both consent, buy is appended → s3
8. Final walk home passes Tom's consent → s4, author utility = 2

# Evaluation

- No peer planner supports ADL + numeric + deep Theory of Mind
- Use benchmark narrative domains → problem scale
- Problem sizes & performance
  - 32 actions / 0 triggers → solved in 6ms
  - 836 actions / 896 triggers → solved in 6.2h
- Ablation study
  - Why both intention and belief reasoning

# Performance

| Domain | Char. | Fluents | Actions | Triggers | Time | Visited | Generated |
|---|---|---|---|---|---|---|---|
| Raiders (1) | 3 | 21 | 39 | 66 | 1.4 s | 905 | 17,815 |
| Space (1) | 2 | 16 | 32 | 0 | 6 ms | 18 | 192 |
| Treasure (1) | 3 | 4 | 34 | 0 | 1 ms | 22 | 288 |
| Hubris (1) | 2 | 29 | 14 | 0 | 47 ms | 58 | 831 |
| BearBirdJr (1) | 2 | 13 | 20 | 0 | 14.0 m | 290,711 | 34,084,608 |
| Lovers (9) | 3 | 111.3 | 312.0 | 370.0 | 40.3 m | 126,983 | 5,198,414 |
| Grandma (2) | 4 | 61.0 | 836.0 | 896.0 | 6.2 h | 598,577 | 105,178,466 |

# Ablation study

| Domain | Full Sabre solutions | Intention-only ✓ | Intention-only ✗ | Belief-only ✓ | Belief-only ✗ |
|---|---|---|---|---|---|
| Raiders | **3** | 0 | **110** | 0 | **110** |
| BearBirdJr | **6** | 0 | **110** | 0 | **110** |
| Lovers (avg) | **10** | 0 | **50.7** | 0 | **32.0** |