

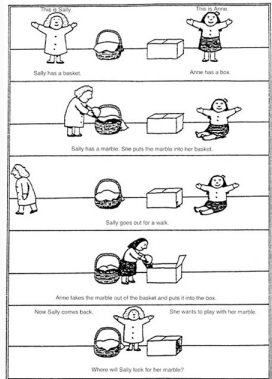
Planning with a Theory of Mind

ICAPS Tutorial: Narrative Planning

ICAPS 2024

Planning with a Theory of Mind

- *Theory of Mind*: the capacity to understand others have a mind of their own, and reason about what goes on in it.
- To be believable, characters must behave consistently with what an audience believes is in their minds.
- Characters may need to reason about what other characters want, believe, and intend.



(Baron-Cohen, Leslie and Frith 1985)
https://en.wikipedia.org/wiki/Sally%E2%80%93Anne_test

- Planning from the author's perspective:
 - Must reason about what characters want, believe and plan, to preserve the illusion that characters are planning and acting of their own will.
 - May reason about what characters believe or assume about what other characters want, believe and plan (because that is what we do)
 - May reason about what the audience believes, to create dramatic effect.
- Planning from a character's perspective:
 - Must work from the beliefs and goals the character has (or is intended or perceived to have).
 - May reason about what other characters want, believe and plan, to predict or manipulate their actions.

- Riedl & Young (2010) distinguish *intentional* actions, performed by one or more characters (its *actors*), from non-volitional events.
- A plan is *intentional* iff every intentional action contributes, directly or indirectly, to achieve an intention (goal) of the character(s) who perform it.
- Characters can acquire intentions as an effect of events or other characters' actions, depending on their character traits.

- Formalised by Riedl & Young in the context of POCL plans.
- A *frame of commitment* is a subset S' of plan steps such that:
 - character A is an actor of every step in S' ;
 - there is a *final step* $s_{fin} \in S'$ that adds g ;
 - there is a *motivating step* s_m that adds (intends A g) and that precedes all steps in S' ; and
 - for each step in S' other than s_{fin} there is a path of causal and/or motivational links to s_{fin} .
- A plan is *intentional* iff every intentional action with actor A belongs to a frame of commitment of A .

Conflict

- Riedl & Young's definition does not allow plans in which characters *fail*, either because
 - they planned from mistaken beliefs; or
 - another character's actions thwarted their plan.
- Ware & Young (2011) amend the definition of intentional plans to allow for plans with inter-character *conflict*, by allowing characters' plan, while complete, to be only partially executed.
- Formally, they distinguish between *executed* and *non-executed* steps in the plan:
 - both executed and non-executed steps may be used to complete a frame of commitment; but
 - only executed steps are subject to normal plan validity.

- Representing character *beliefs* about the world state means representing a potentially different state for each character.
- Character's beliefs can change as a consequence of
 - their actions (failure uncovers false belief);
 - other character's actions (telling or revealing); and
 - non-volitional events (observation).
- Formalised (e.g., Teutenberg & Porteous 2015; Christensen et al. 2020):
 - Character's plans (frames of intention) must be complete with respect to the facts that the character believes.
 - Requires modelling the effects of a character attempting an action whose preconditions are in fact not true.

- Although properties like intentionality, conflict, belief are often formalised in the context of (extended) POCL plans, planners achieving them can be *implemented* in many ways, *e.g.*:
 - By modification/extension of classical POCL algorithms (Riedl & Young 2010; Ware & Young 2011).
 - By modification/extension of (heuristic) state-space search planners (Ware 2012; Ware & Young 2014; Sanghrajka et al. 2022).
 - Through multi-agent planning (Brenner 2010; Teutenberg & Porteous 2013, 2015; Porteous & Lindsay 2019).
 - By reformulation/compilation (Haslum 2012; Christensen et al. 2020).

Planner as an Audience Model

- A planner, using the facts of the story narrated so far, can be used as a model of the audience's expectations:
 - If there are few or unlikely plans in which the story's protagonist is successful from the current state, then the state is *suspenseful* (O'Neill & Riedl 2014; Cheong & Young, 2015).
 - If the story diverges from the plan generated by the audience model, then the story is *surprising* (Bae & Young, 2014).
- Can also incorporate reasoning about the audience's likely inferences from narrated facts (*e.g.*, Arinbjarnar 2008; Chieppe et al. 2022).