Capabilities in Agent Systems

Lin Padgham

Associate Professor in Artificial Intelligence

RMIT University

Melbourne Australia

Overview of talk

- Beliefs, goals and intentions
- Capability concept in a single agent
  - definitions of capability
  - role of capabilities in theory and systems
  - representation of capabilities
- Use of capabilities in multi-agent systems

Mental attitudes

- Need to represent “mental attitudes” in theory and systems to define appropriate, desired, rational reasoning.
- Some researchers argue that beliefs and goals are the only necessary attitudes. Others argue intentions also primary.
- Most popular theories and systems use beliefs, goals (or desires) and intentions.

Beliefs, Goals, Intentions

- **Beliefs** - what the agent knows/believes about the world.
- **Goals** - states/situations the agent wants to achieve. Required to be consistent.
- **Intention** - an action plan the agent has committed to in order to achieve one of its goals.

Semantics

- Much work on desired semantics of beliefs, goals and intentions

e.g.

BEL(Clever(John)) AND John = Partner(Sally)

doesn't imply BEL(Clever(Partner(Sally)))

GOAL(Dentist) AND BEL(Dentist implies Pain)

doesn't imply GOAL(Pain)

Semantics

- But many implications you **do** want:

e.g.

INTEND(X) implies GOAL(X)

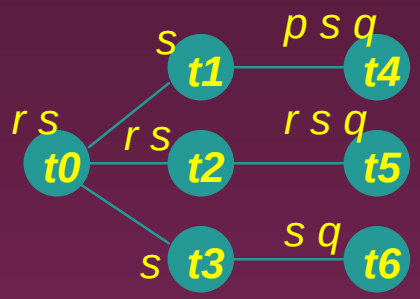
you don’t want arbitrary intentions

GOAL(P AND Q) implies

GOAL(P) AND GOAL(Q)

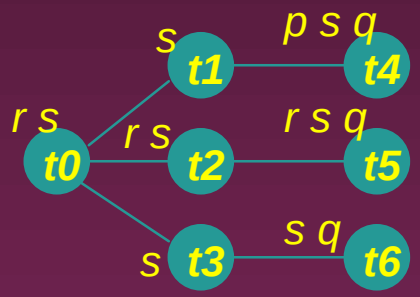
Modal branching-time logic of Rao and Georgeff

Logic is somewhat complex, but captures desired semantics well



Modal branching-time logic of Rao and Georgeff

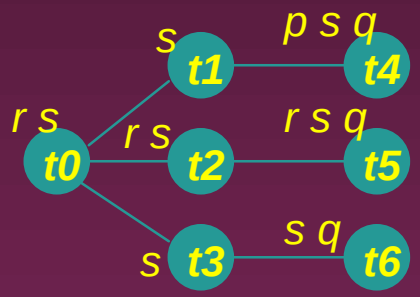
Logic is somewhat complex, but captures desired semantics well



BEL(inevitably always s)

Modal branching-time logic of Rao and Georgeff

Logic is somewhat complex, but captures desired semantics well

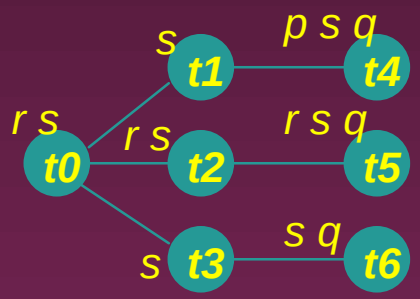


BEL(inevitably always s)

BEL(optionally eventually p)

Modal branching-time logic of Rao and Georgeff

Logic is somewhat complex, but captures desired semantics well



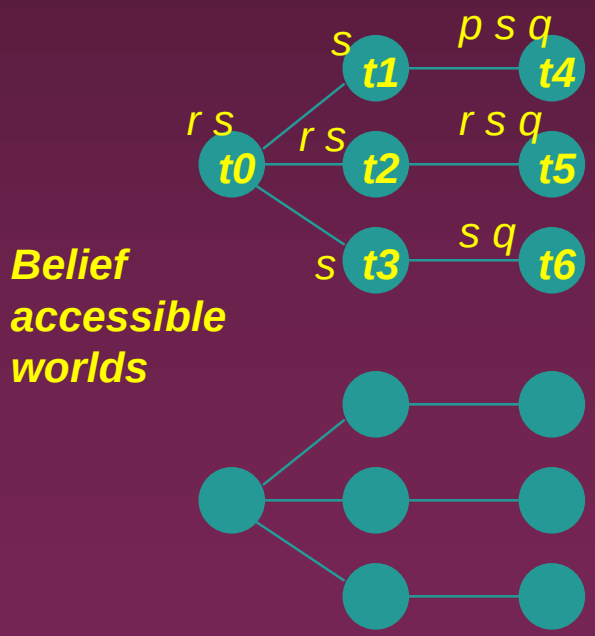
BEL(inevitably always s)

BEL(optionally eventually p)

BEL(inevitably eventually q)

Modal branching-time logic of Rao and Georgeff

Logic is somewhat complex, but captures desired semantics well

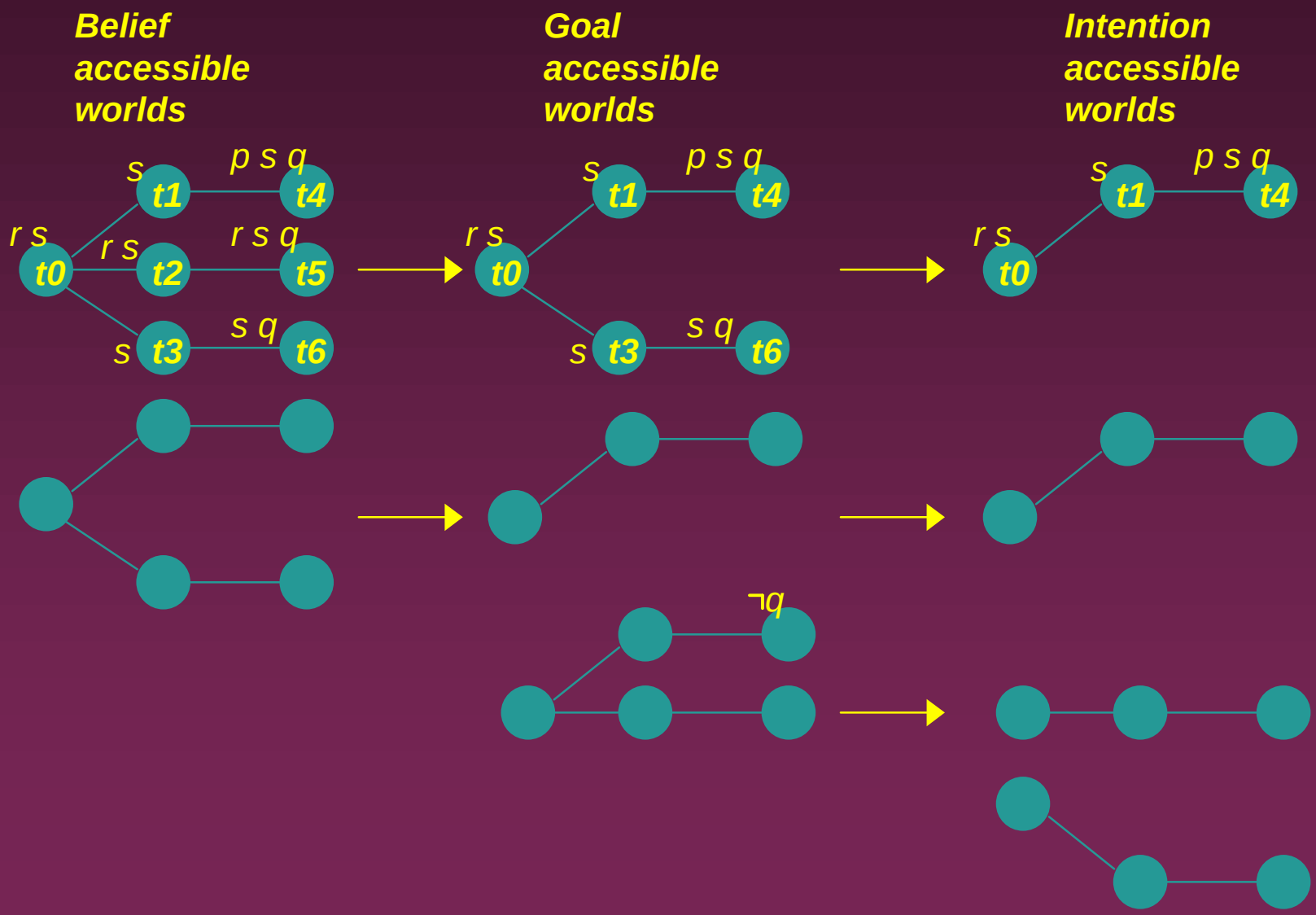


BEL(inevitably always s)

BEL(optionally eventually p)

BEL(inevitably eventually q)

Modal branching-time logic of Rao and Georgeff



Why capabilities?

- Intuitive mismatches and anomalies between BDI theories and systems
- Recent implementation of capabilities in BDI system (JACK)
- “Natural” concept for agent reasoning

What is a capability?

In our work a capability X is:

- ability to take action towards a goal X
- this ability may be limited to specific situations
- equivalent to a plan (or possibly set of plans)
- the plan(s) representing the capability will have trigger ***"achieve goal X"***
- the plans may have restricted applicability

Other versions of capability

Levesque and Lin

- Achievable goals given start state and actions

Lesperance, Levesque, Lin and Scherl

- Agent has ***ability*** X in situation S iff it has an action selection function that allows it to know in S that it can get to S' where X holds.

Dung

- C(p,s) - a function representing all paths that may be taken by the plan p from the state s.

Other versions of capability - Systems

KAoS

- Services or functions

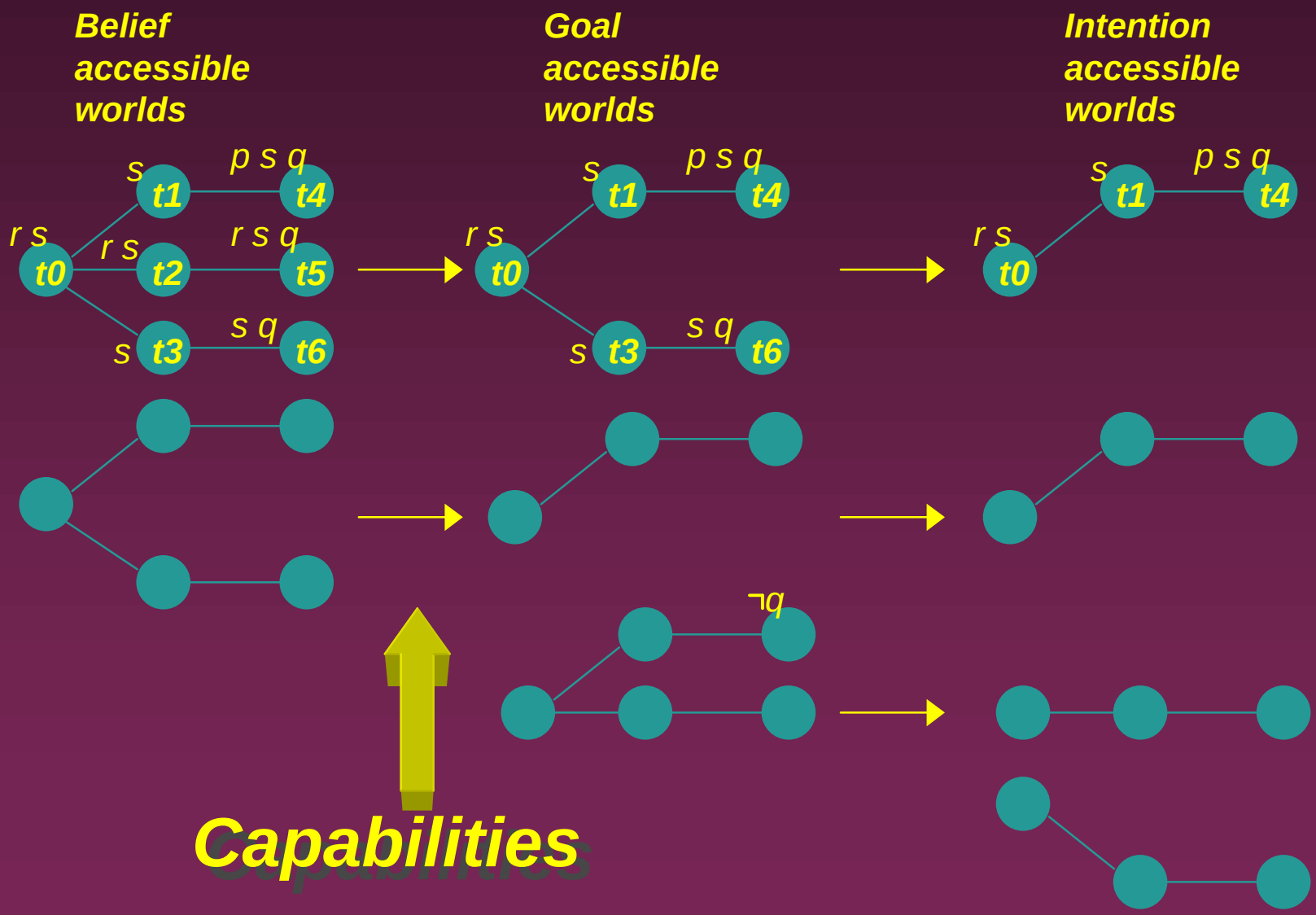
LARKS

- frame with slots for context, types,input, output, constraints,descriptions

JACK

- plan set, KB fragment, interface specification

Modal branching-time logic of Rao and Georgeff



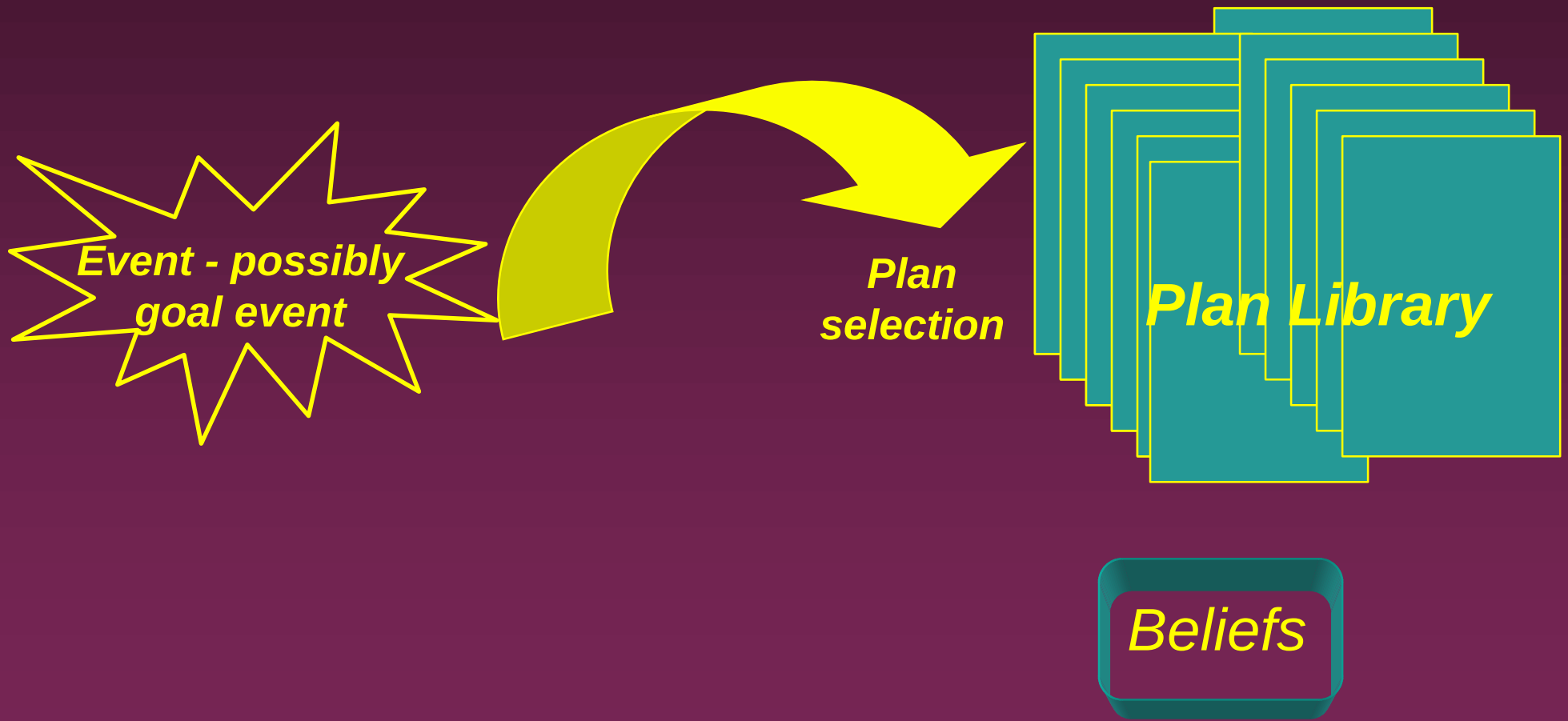
Role of capabilities theoretically

- capabilities constrain goals
- having a goal implies you have a capability to achieve that goal
- e.g. may have a desire for rain, but nonsensical to adopt goal for rain (unless you can control the weather)
- as goals constrain intentions, capabilities in effect also constrain intentions
- our current work suggests goals are constrained by desire and capability

Current systems

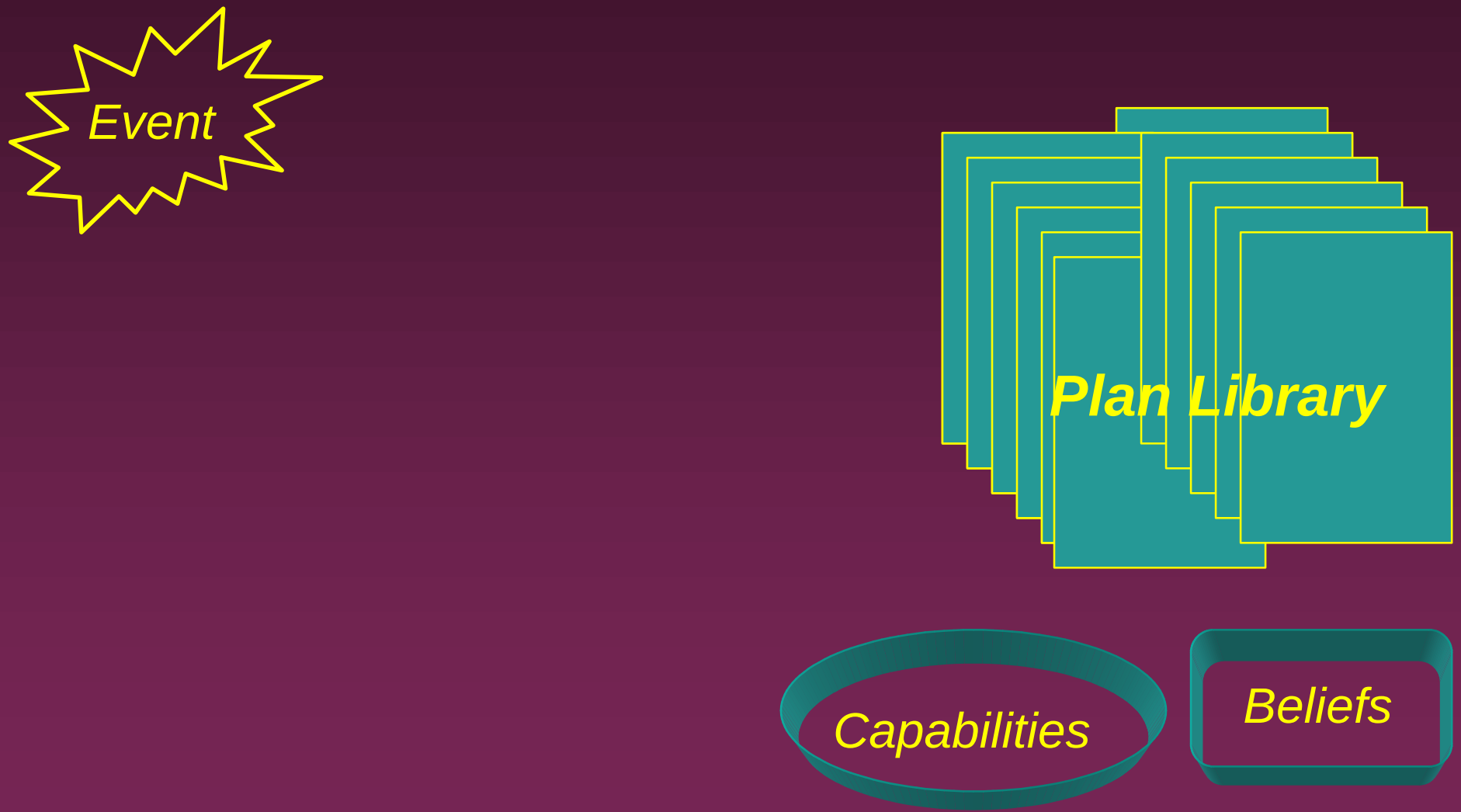
Current systems

Current systems

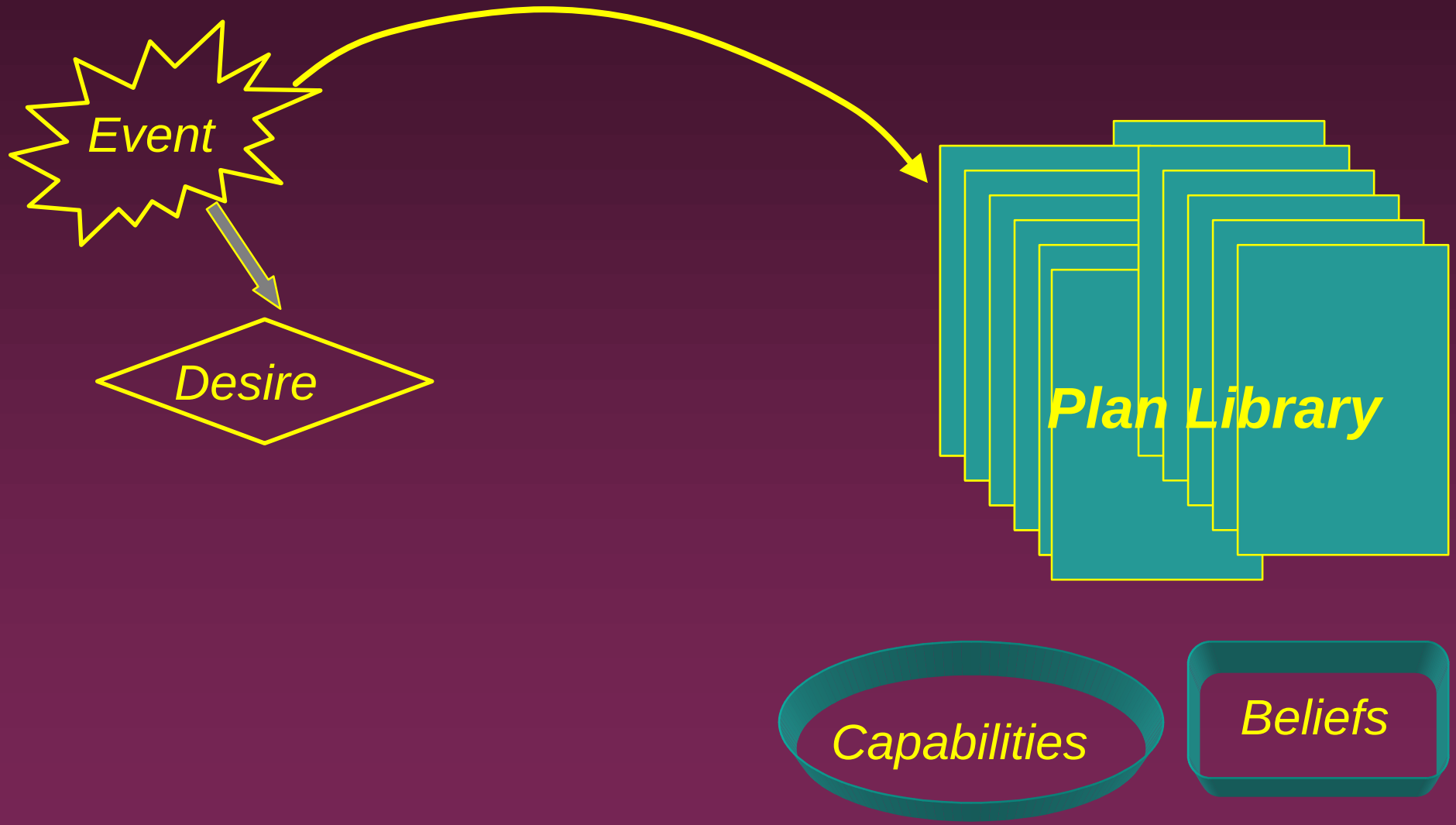


Our model

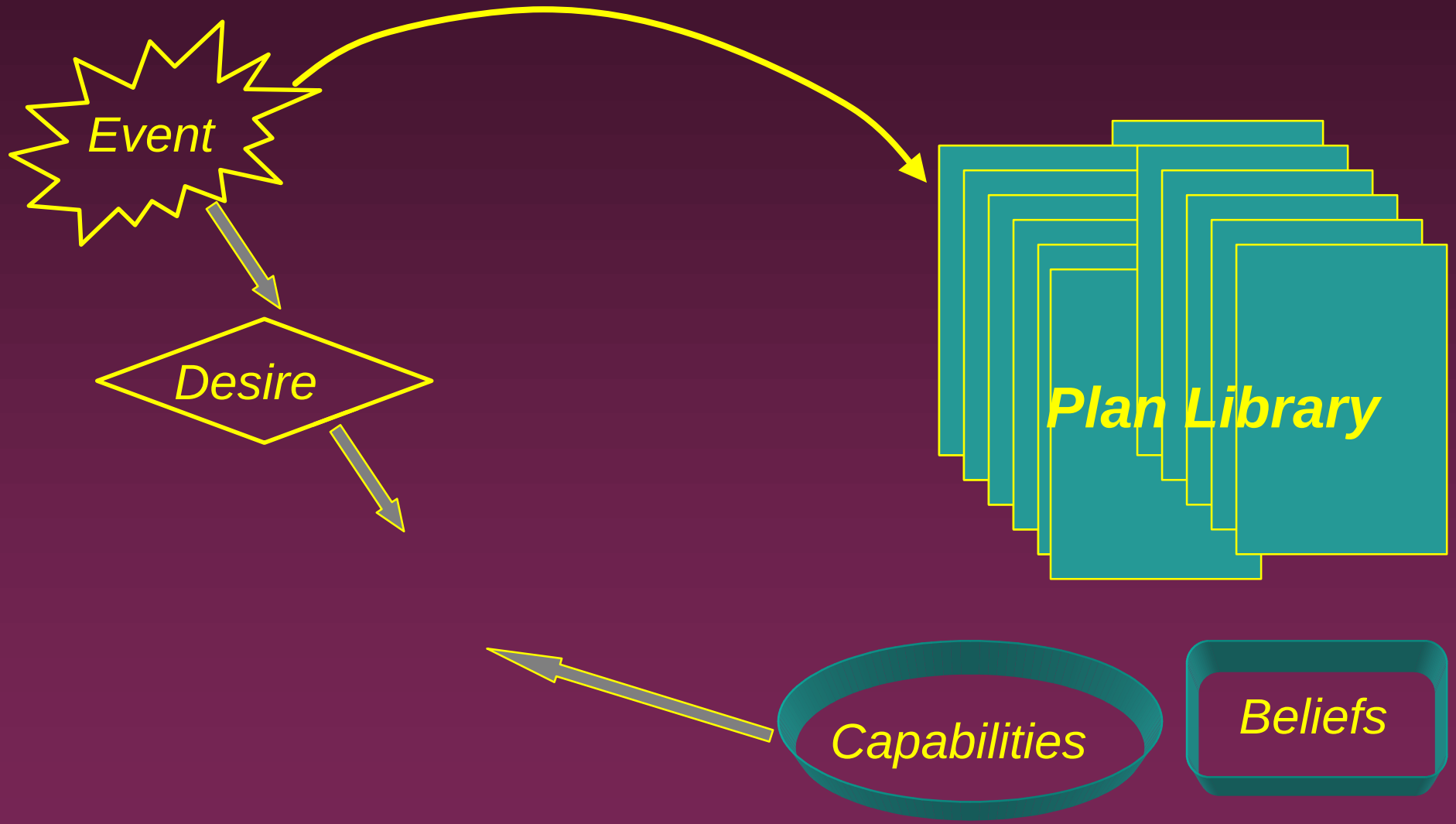
Our model



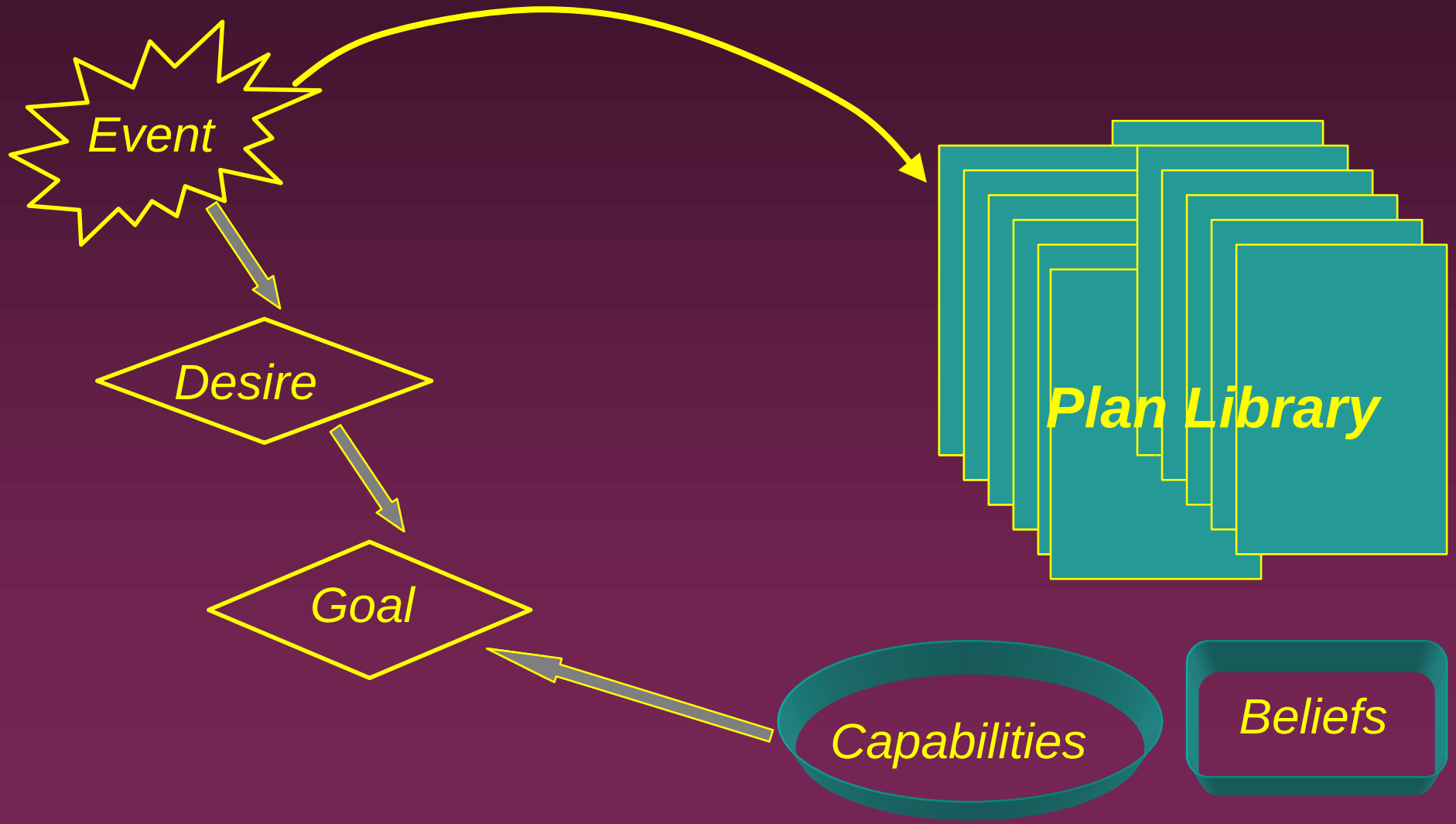
Our model



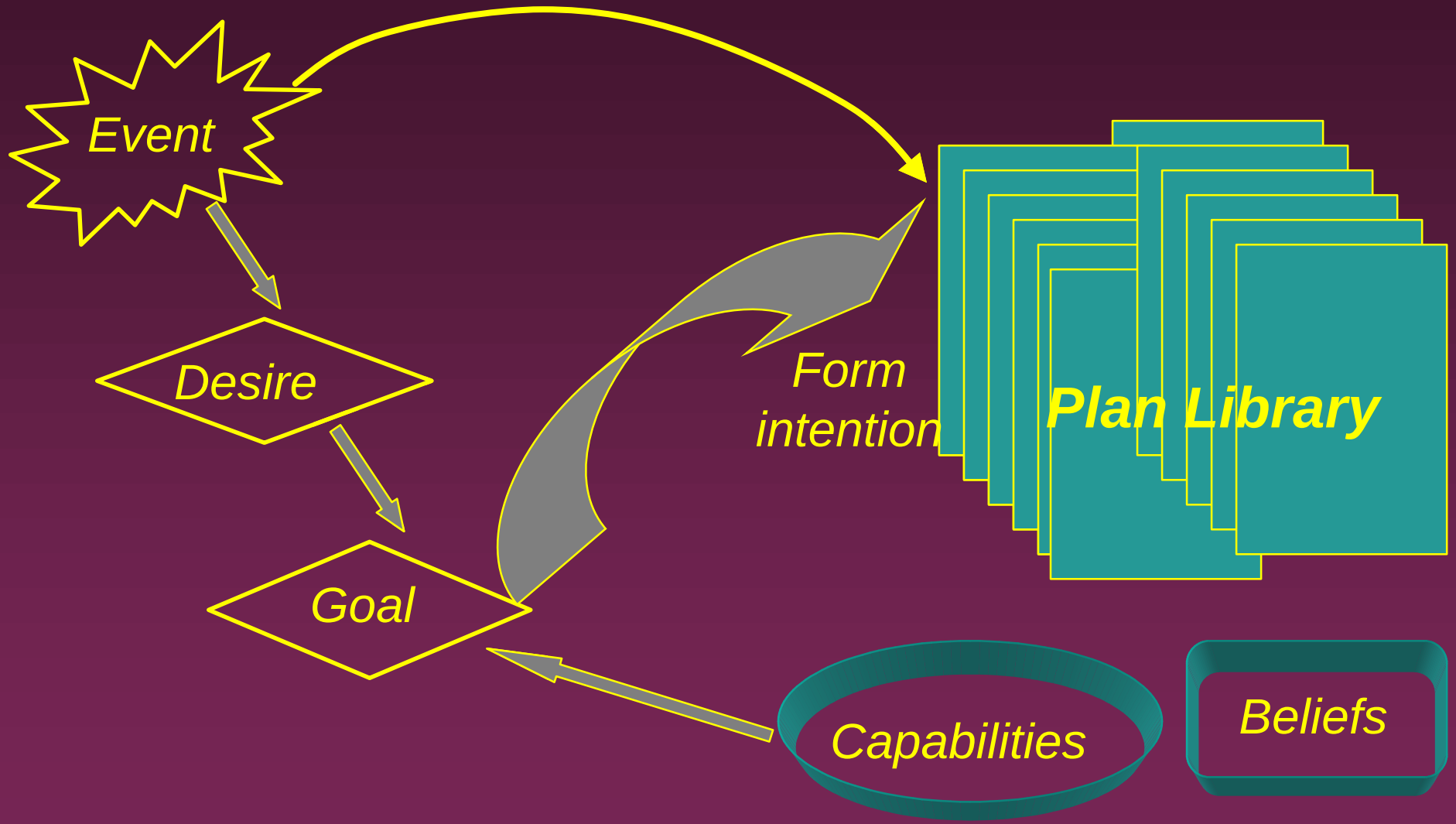
Our model



Our model



Our model


Event
Desire
Goal
Form intention
Plan Library
Capabilities
Beliefs

Goals

- Constrained by Desires and by Capabilities
- Are achieved by execution of plans
- Agent tries all possible means to accomplish
  - I.e. if one plan fails another will be tried

Checking capabilities

- Before a desire X can become a goal, agent must check its capabilities
- Depending on definition of capability this means
  - Do I have some plan to achieve X? or
  - Do I have some plan to achieve X that I can execute now?

Capability nuances

- May need different nuances of capability
- Capability X = have some plan to achieve X
- Current capability X = have plan to achieve X which is currently executable

- Former is static - most useful for advertising to others
- Latter prevents agent adopting goals it can't act on

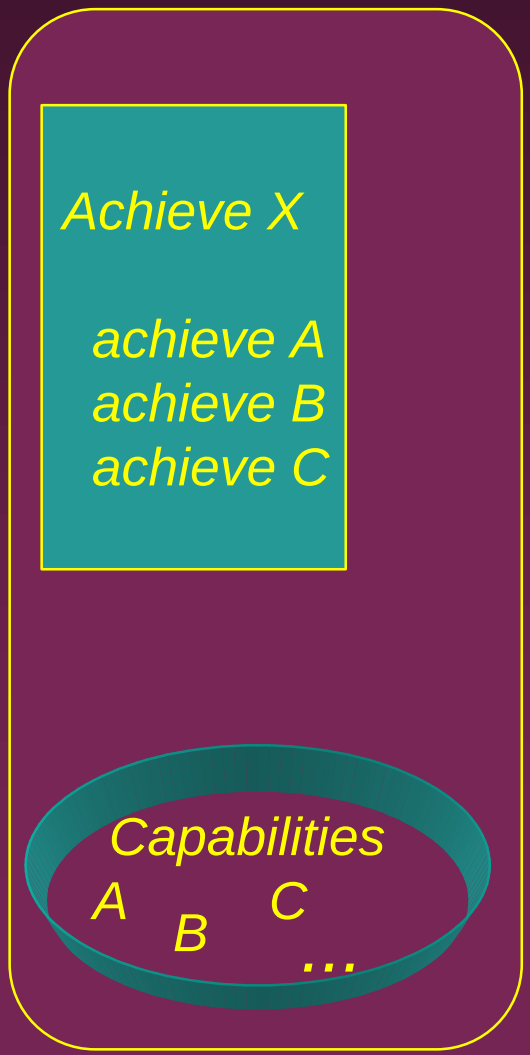
Capability = plan or {plans}?

- Theoretically each plan achieves some goal so capability = single plan
- But many plans achieve low-level goals, relevant only in context
- Possibly *package* plan sets that each achieve a “relevant” goal

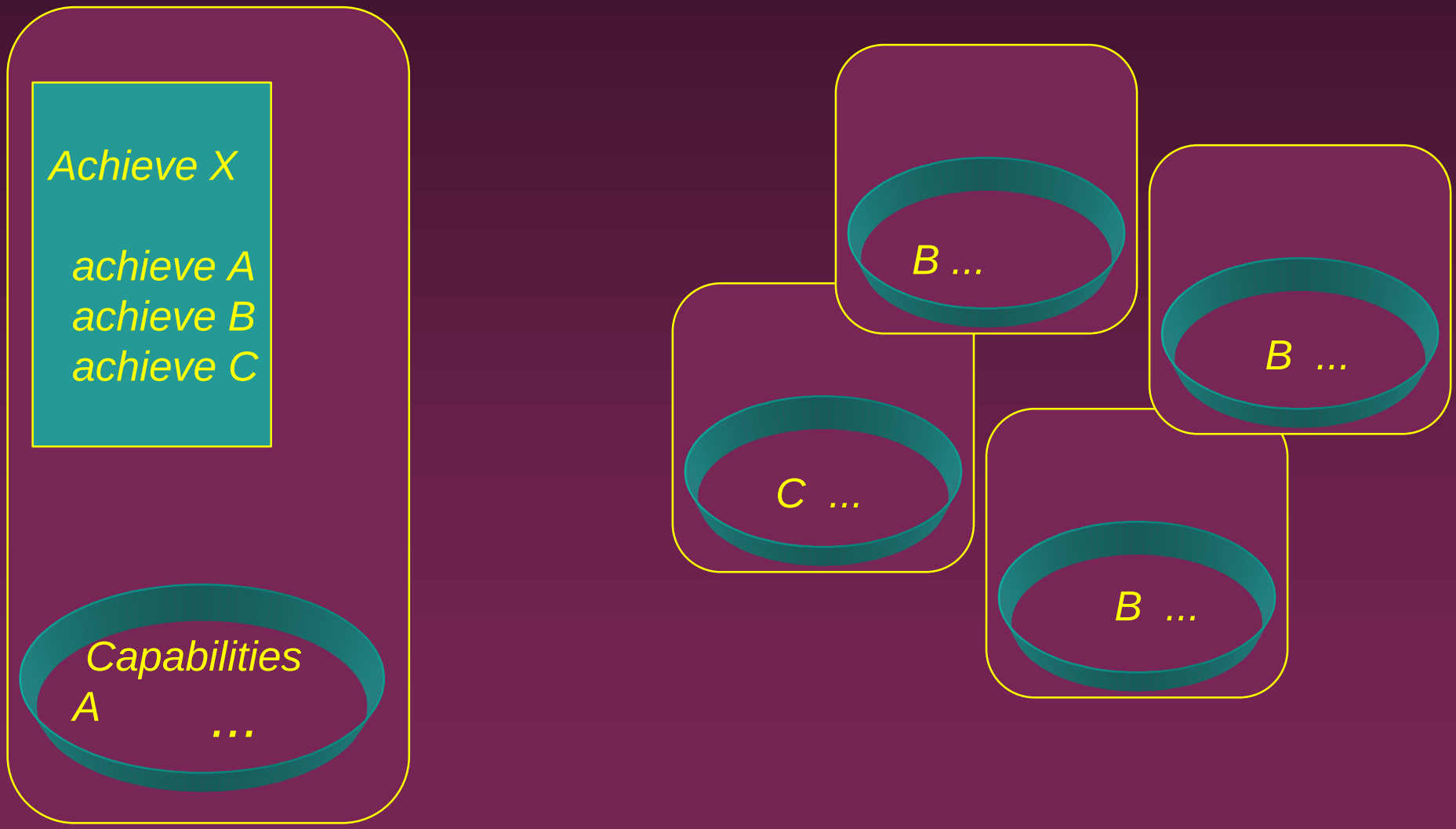
Capabilities as packages in JACK

- Introduced primarily as a tool for modular program development
- Capabilities provide a level at which code can be re-used
- Also provide a level at which an agent can be described
- A single capability may handle multiple events (I.e. achieve multiple goals)

Multi-agent systems



Multi-agent systems



(Co-operative) multi-agent systems

- Many use team plans and shared goals
- Doesn’t cover sub-contractor model
- Advertising capabilities could be used in both team plans and sub-contractor model

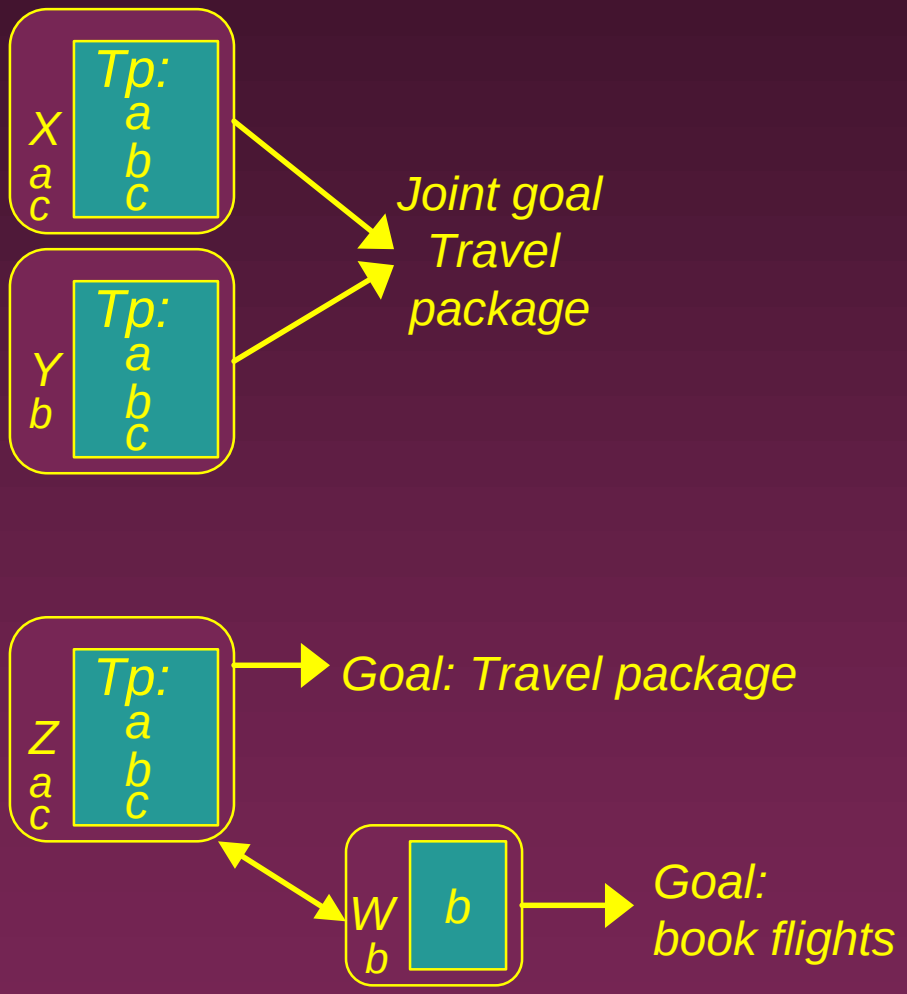
Example

Travel package

a: book accommodation

b: book flights

c: package info



External representation

- Descriptive name of service or function
- Goal(s) it can achieve
- Event(s) it can handle
- Events it posts
- (Sub)goals it requires assistance with
- Resources it uses
- Access/request information
- ???

Summary

- Capabilities begin to be explicit in intelligent agents - much less than goals and intentions
- Better match of theory to systems if capabilities included in theory
- Useful level for agents to advertise their abilities to others
- Details of appropriate representation unclear