DELIBERATIVE ARCHITECTURE: ACCOUNTING FOR OBSERVATIONS
Recap

- Domain description: Define a transition function $\Phi$ from actions and states to sets of states.
  - Effect of actions: $a$ causes $f$ if $p_1, \ldots, p_m$
  - Causality: $p_1, \ldots, p_n$ causes $f$
  - Executability conditions: executable $a$ if $q_1, \ldots, q_r$.
  - $\Phi(a, s) = \emptyset$ means $a$ is not executable in the state $s$.

- Observations: $f$ after $a_1, \ldots, a_n$.
  - These observations (except of the kind initially $f$) are actually hypothetical statements.
  - Together with a domain description (or a $\Phi$) they define possible initial states.
  - Real observations vs hypothetical statements (or oracles)
    * $a$ causes $f$; $b$ causes $g$; $f$ after $a$; $g$ after $b$ is consistent.
    * But only one of $a$ and $b$ happened in the initial state!
A language for observations

• Motivation:
  – Initially the car was parked. The next day the car was gone. Conclusion: The car must have been stolen.
  – I am at home, I have a car, and my suitcase is unpacked. I need to go to the airport with a packed suitcase. I make the plan to pack and then drive to the airport. After packing I observe that my car is no longer there. I can not just execute the remaining part of my old plan. I need to make a new plan from the current situation I am in.
  – I left home and my light switch was working. I came back and turned on the light and it does not work. I need to diagnose what went wrong; but not just a surface level diagnosis about what component is faulty but also a deeper diagnosis explaining what actions caused the fault.

• Syntax of the language:
  – $f \text{ at } s$
  – $s_1 \text{ precedes } s_2$
Deliberative architecture: accounting for observations

- $\alpha$ occurs at $a$
- $\alpha$ between $s_1, s_2$
- $s$ and $s_i$s are situations (or time points), $\alpha$ is a sequence of actions.

• Example 1:
- $O = \{\text{has car at } s_0; \neg\text{has car at } s_1; s_0 \text{ precedes } s_1\}$
- $D = \{\text{steal car causes } \neg\text{has car}\}$
- From $(D, O)$ we want to conclude that steal car between $s_0, s_1$.

• Example 2:
- $O_1 = \{\text{at home at } s_0; \text{has car at } s_0; \neg\text{packed at } s_0; \}$
- $D_1 = \{\text{pack causes packed}; \text{executable pack if at home};$
  $\text{steal car causes } \neg\text{has car}; \text{rent car causes } \text{has car};$
  $\text{drive causes at airport}; \text{executable drive if has car};$
  $\text{at home causes } \neg\text{at airport}; \text{at airport causes } \neg\text{at home}\}$
- From $(D_1, O_1)$ we can conclude that at airport, packed after pack, drive.
- I do the packing.
- $O_2 = O_1 \cup \{\text{pack occurs at } s_1; s_0 \text{ precedes } s_1\}$. 
- But then I notice that my car is gone.
  \( O_3 = O_2 \cup \{ \neg \text{has\_car at } s_2; s_1 \text{ precedes } s_2 \} \)
- From \((D_1, O_3)\) I can no longer conclude that \(\text{at\_airport, packed after drive} \).
- But from \((D_1, O_3)\) I should be able to conclude that
  \(\text{at\_airport, packed after rent\_car; drive} \).

- Query language: The language of observations plus
  
  \( f \text{ after } a_1, \ldots, a_n \text{ at } s. \)

- If \(s = s_C\), the current situation then we just write
  
  \( f \text{ after } a_1, \ldots, a_n \)
- If \(n = 0\) then we write \(f \text{ at } s.\)
- If \(n = 0\) and \(s = s_C\) then we write \(\text{currently } f.\)
- If \(n = 0\) and \(s = s_0\) then we write \(\text{initially } f.\)

- \(s_C\) is a special constant and observations are not allowed with respect to this constant.
- A pair \((D, O)\) will now be referred to as a narrative.
Semantics of narratives

• Defining $\Phi$
  
  – An action $a$ is said to be possibly executable in a state $s$ w.r.t. a $D$ if there exists executable $a$ if $q_1, \ldots, q_r$ in $D$ such that $q_1, \ldots, q_r$ hold in $s$.
  
  – $\Phi(a, s) = \emptyset$ if $a$ is not possibly executable in $s$.
  
  – Else, $\Phi(a, s) = \{s' : s' = Cn_R((s \cap s') \cup E_a(s))$ and $s'$ is an interpretation $\}$.

• Valid states (w.r.t. $D$): States that are closed under the static causal rules in $D$.

• Semantics is defined in terms of a causal model $\Psi$ and a situation assignment $\Sigma$.

• A causal model $\Psi$ of a domain description $D$ is a function from action sequences to valid states such that

  – $\Psi([ ])$ is a valid state.
  
  – If $\Phi(a, \Psi(\alpha)) \neq \emptyset$ then $\Psi(\alpha \circ a) \in \Phi(a, \Psi(\alpha))$ else $\Psi(\alpha \circ a)$ is undefined.

Intuitively $\Psi$ selects one of the possible trajectories defined by $\Phi$ as a possible evolution of the world.
A *situation assignment* with respect to $D$ is a mapping $\Sigma$ from situations into the set of action sequences of $D$ that satisfy the following properties:

- $\Sigma(s_0) = []$;
- for every $s \in S$, $\Sigma(s)$ is a prefix of $\Sigma(s_C)$.

An *interpretation* $M$ of $(D, O)$ is a pair $(\Psi, \Sigma)$, where $\Psi$ is a causal model of $D$, $\Sigma$ is a situation assignment, and $\Sigma(s_C)$ is defined.

For an interpretation $M = (\Psi, \Sigma)$ of $(D, O)$:

- $\alpha$ occurs_at $s$ is true in $M$ if the sequence $\Sigma(s) \circ \alpha$ is a prefix of $\Sigma(s_C)$;
- $\alpha$ between $s_1, s_2$ is true in $M$ if $\Sigma(s_1) \circ \alpha = \Sigma(s_2)$;
- $f$ at $s$ is true in $M$ if $f$ holds in $\Psi(\Sigma(s))$;
- $s_1$ precedes $s_2$ is true in $M$ if $\Sigma(s_1)$ is a prefix of $\Sigma(s_2)$.

An interpretation $M = (\Psi, \Sigma)$ is a *model* of a narrative $(D, O)$ if:

(i) observations in $O$ are true in $M$;
(ii) there is no other interpretation $M' = (\Psi, \Sigma')$ such that $M'$ satisfies condition (i) above and $\Sigma'(s_C)$ is a subsequence of $\Sigma(s_C)$.
Observe that these models are minimal in the sense that they exclude extraneous actions.

For a model $M = (\Psi, \Sigma)$, $\Sigma(s_C)$ is referred to as the actual line w.r.t. $M$. It encodes how the world has progressed till the current situation.

- A narrative is *consistent* if it has a model. Otherwise, it is *inconsistent*.
- A query of the form $\varphi$ after $\alpha$ at $s$ is true in a model $M = (\Psi, \Sigma)$ if $\varphi$ is true in $\Psi(\Sigma(s) \circ \alpha)$.
- We already defined when other kind of queries (i.e., observations) are true in a model.
- A query $q$ is entailed by a narrative $(D, O)$, denoted by $(D, O) \models q$, if $q$ is true in every model of $(D, O)$.
The stolen car example revisited

- $D = \{\text{steal\_car causes } \neg \text{has\_car}\}$
- $O = \{\text{has\_car at } s_0; \neg \text{has\_car at } s_1; s_0 \text{ proceeds } s_1\}$
- The semantics:
  - States: $\emptyset, \{\text{has\_car}\}$
  - $\Phi(\text{steal\_car}, \emptyset) = \emptyset; \Phi(\text{steal\_car}, \{\text{has\_car}\}) = \emptyset$
  - $\Psi(\emptyset) = \{\text{has\_car}\}; \Psi(\alpha \circ \text{steal\_car}) = \emptyset$
  - $\Sigma(s_0) = []; \Sigma(s_1) = [\text{steal\_car}]; \Sigma(s_C) = [\text{steal\_car}]$
  - $(D, O)$ has only one model $M = (\Psi, \Sigma)$
  - Why is $(\Psi, \Sigma')$ not a model?
    (where $\Sigma'(s_0) = []; \Sigma'(s_1) = [\text{steal\_car, steal\_car}]; \\
    \Sigma'(s_C) = [\text{steal\_car, steal\_car}].)$
- $D, O \models \text{steal\_car between } s_0, s_1$
  because $\text{steal\_car between } s_0, s_1$ is true in the only model $M$ of $(D, O)$
  because $\Sigma(s_1) \circ \text{steal\_car} = \Sigma(s_0)$. 
An example with multiple models.

- $O = \{\neg f, \neg g \text{ at } s_0; \ f \text{ at } s_1; \ s_0 \text{ precedes } s_1; \ a_4 \text{ occurs at } s_1\}$
- $D = a_1 \text{ causes } g; \ a_2 \text{ causes } f \text{ if } g; \ a_3 \text{ causes } g; \ a_4 \text{ causes } \neg g \text{ if } f\}$

- What are the models of $(D, O)$? What are the actual lines? Does $(D, O) \models \text{ currently } \neg g$?

<table>
<thead>
<tr>
<th>actions</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$a_3$</th>
<th>$a_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>states</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_1$</td>
<td>${\neg f, \neg g}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_2$</td>
<td>${f, \neg g}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_3$</td>
<td>${\neg f, g}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_4$</td>
<td>${f, g}$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- $\Phi$ and $\Psi$ are straightforward.

- $\Sigma_1(s_0) = [\ ], \Sigma_1(s_1) = [a_1, a_2], \Sigma_1(s_C) = [a_1, a_2, a_4].$
- $\Sigma_2(s_0) = [\ ], \Sigma_2(s_1) = [a_1, a_2, a_3], \Sigma_1(s_C) = [a_1, a_2, a_3, a_4].$
- $\Sigma_3(s_0) = [\ ], \Sigma_3(s_1) = [a_3, a_2], \Sigma_3(s_C) = [a_3, a_2, a_4].$