Modal Logic

- Modal operator operates on formulas (sentences in a logic)
- Can be used for expressing belief
  - So system can reason about what a person $m$ believes
    \[ \text{B}_m(\text{raining}) \]
    \[ \text{B}_m(\text{raining} \rightarrow \text{wet}) \]
  - What should we be able to conclude?
- Can be used for expressing wants (user $m$’s goals)
  \[ \text{Want}_m(\text{cleandishes}) \]
  \[ \text{cleandishes} \rightarrow \text{cleaningthem} \]
- Can be used for expressing possibility
  - System can reason about what is possible
    \[ \text{possibly}(\text{raining}) \]
    \[ \text{raining} \rightarrow \text{wet} \]

Possible Worlds
- Each world is like an interpretation
  - Full assignment specifying everything as either true or false
- Unlike an interpretation, can have multiple worlds with the same truths
- But there is structure between the worlds
  - Accessibility relation between worlds $w_iRw_j$
    - $B\phi$ is true in world $w$ if $\phi$ is true in all worlds $w'$ such that $wRw'$
      + $m$ is actually in world $w$ but might not know which world they are in.
Comparison to Datalog/FOPC

- Semantics of FOPC can be thought of as possible worlds
  - Intended interpretation is $w_0$
  - All models of the KB $w'$ are accessible from $w_0$: $w_0Rw'$
    + Including the intended interpretation: $w_0Rw_0$
- For modal logic
  - There is a real world $w_0$ that the agent knows it is in (assumed to have complete knowledge)
  - Accessibility relation is not just from the initial world $w_0$
    + Will allow us to model nested modal operators: $bel(bel(raining))$
  - Might not include $w_0Rw_0$
    + Agents might have incorrect beliefs
  - Worlds (interpretation) can be repeated, but with different accessibility relations

Overview

⇒ Omnipotent
- Model Operator Axioms
- Syntactic Proofs
- Constants and Quantifiers
Epistemic Necessitation

- Does the agent know all necessary truths?
- Example:
  - DeMorgan’s law is always true (due to semantics of $\land$, $\lor$, and $\neg$)
  - $\neg(\phi \land \psi) \leftrightarrow (\neg \phi \lor \neg \psi)$
  - So it is true in every world
  - So it is true in every accessible world from any world
  - So $B(\neg(\phi \land \psi) \leftrightarrow (\neg \phi \lor \neg \psi))$ ($\phi$ and $\psi$ are variables over formulas)
  - Possible worlds semantics is forcing us to make all tautologies true of modal operators
- Can be written as: If $\vdash \phi$ then $\vdash B\phi$
  - This is not the same as $\phi \rightarrow B\phi$
  - That is saying that the agent believes everything that is true in the world
  - If that was true, modal operator would not be very useful

Distribution Axiom

- If $B\phi$ and $B(\phi \rightarrow \psi)$ does $B\phi$ ???

- So agents believe all of the consequences of their beliefs
  - This is forced by the possible worlds semantics
- Possible world semantics is forcing agents to believe all tautologies and logical consequences of their beliefs
  - Not entirely realistic for human agents
- This can equivalently be written as $B(\phi \rightarrow \psi) \rightarrow (B\phi \rightarrow B\psi)$
  - Can distribute $B$ over implication
Properties on R

- Depending on what B is modeling, you might want R to have certain properties
  - Reflexive: for all $w$, $wRw$
  - Symmetric: if $w_1Rw_2$ then $w_2Rw_1$
  - Transitive: if $w_1Rw_2$ and $w_2Rw_3$ then $w_1Rw_3$
  - Serial: for all $w_1$ there exists $w_2$ s.t. $w_1Rw_2$
- R is euclidean if R has what two properties?
Knowledge Axiom

• $B\phi \rightarrow \phi$  Everything that user believes is true
  - Which is why it is called the knowledge axiom
  - Not usually used for belief, but is used for knowledge

• Consider if $R$ is reflexive:  $wRw$ is true for every world
  - Means that $B\phi$ is only true if $\phi$ is true in the current world
    + Person’s beliefs (+ve and -ve) are subset of what is true in current world
  - So if $R$ is reflexive than the knowledge axiom holds (not vice versa)

Positive-Introspection Axiom

• $B\phi \rightarrow B(B\phi)$
  - If a user believes something, the user believes that they believe it

• If $R$ is transitive:  if $w_1Rw_2$ and $w_2Rw_3$ then $w_1Rw_3$
  - Positive-Introspection Axiom holds
  - Proof?
Negative Introspection Axiom

- $\neg B\phi \rightarrow B(\neg B\phi)$
  - If you do not believe something, you believe you do not believe it

- Follows if R is eclidean
  - if $w_1 R w_2$ and $w_1 R w_3$ then $w_2 R w_3$

Different Combinations

- Axioms:
  - Epistemic Neccessitation
  - Distribution Axiom
  - Knowledge Axiom
  - Positive Introspection Axiom
  - Negative Introspection Axiom

- Can pick and choice which ones you want for your logic

<table>
<thead>
<tr>
<th>Modal Logic</th>
<th>Constraints on R</th>
<th>Axioms in Proof Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>no constraint</td>
<td>D</td>
</tr>
<tr>
<td>T</td>
<td>reflexive</td>
<td>D,K</td>
</tr>
<tr>
<td>S4</td>
<td>reflexive, transitive</td>
<td>D,K,P</td>
</tr>
<tr>
<td>S5</td>
<td>reflexive, transitive, symmetric</td>
<td>D,K,P,N</td>
</tr>
<tr>
<td>Weak S4</td>
<td>transitive</td>
<td>D,P</td>
</tr>
<tr>
<td>Weak S5</td>
<td>transitive, euclidean</td>
<td>D,P,N</td>
</tr>
</tbody>
</table>
Example Proof

Nora believes $p \rightarrow q$
Nora does not believe $q$ (believes $\neg q$ or has no belief about $q$)
Show that Nora does not believe $p$
Two Wise Men

The king tells two wise men that at least one of them has a white spot on his forehead. Each man can see the other’s forehead but not his own. The first wise man says “I don’t know whether I have a spot.” The second says ...

• For \(a\) (the first wise man)
  - \(a\) knows whether \(b\) has a spot and that either \(a\) or \(b\) or both have a spot
  - if \(b\) does not have a spot, \(a\) would know this, and would hence know that \(a\) has a spot, and would answer “I know I have a spot”
  - but he didn’t, so \(a\) must know that \(b\) has a spot

• For \(b\) (the second wise man)
  - knows that \(a\) knows whether \(b\) has spot and that either \(a\) or \(b\) has spot
  - \(b\) knows that if \(b\) does not have a spot, \(a\) would know this, and \(a\) would conclude that \(a\) has a spot, and \(a\) would answer “I know I have a spot.”
  - but \(a\) didn’t, so \(b\) must have a spot, and \(b\) answers “I know I have a spot”

Writing the Knowledge

• Says two wise men, so let’s view this as knowledge
  - Computer will be \(b\). How does \(b\) reason?

• \(b\)’s knowledge
  - No need to put inside of a modal operator
  - either \(a\) or \(b\) has a spot (or both have a spot)
  - \(a\) has a spot

• \(b\)’s knowledge of \(a\)’s knowledge (written with \(K_a\))
  - \(a\) knows \(a\) or \(b\) has a spot (or both)
  - \(a\) knows whether \(b\) has a spot
  - \(a\) does not know whether \(a\) has a spot
  - if \(b\) has a spot, \(a\) will know it
  - if \(b\) does not have a spot, \(a\) will know it
Proving that $s(b)$

\[-s(b) \rightarrow K_a(\neg s(b))\]  
\[K_a(s(a) \lor s(b))\]  
\[\neg K_a(s(a))\]  
\[s(b) \lor K_a(\neg s(b))\]  

Can we rewrite 2 as $K_a(s(a)) \lor K_a(s(b))$? No.

\[K_a(\neg s(b) \rightarrow s(a))\]  
\[K_a(\neg s(b)) \rightarrow K_a(s(a))\]  
\[\neg K_a(\neg s(b)) \lor K_a(s(a))\]  
\[s(b) \lor K_a(s(a))\]  
\[s(b)\]

Pitfall of Syntactic Proofs

- Syntactic Proofs
  - Extra axioms
  - Converting to and from CNF
  - What do we need to make it complete?
  - Can we search effectively in this space?
Alternative: Proofs in Possible Worlds

- $B(\alpha)$ in $w$ means $\alpha$ is true in all $w'$ such that $wRw'$
  - Let’s use $T(\alpha, w)$ to mean $\alpha$ is true in $w$
- $T$ and $R$ are used in our semantics to give meaning to the modal operator
- However, could view $T$ and $R$ to be predicates in FOPC logic
  - Worlds could then be objects in that language
  - Translate sentences about $B$ into $R$ and $T$
  - Predicates in modal logic become functions
  - Rather than use axioms, use corresponding restriction on $R$

Example

\[\neg s(b) \rightarrow K_a(\neg s(b))\]
\[\neg s(b) \rightarrow (\forall w R(w_0, w) \rightarrow \neg T(s(b), w))\]
\[K_a(s(a) \lor s(b))\]
\[\neg K_a(s(a))\]
\[K_a(s(a) \lor s(b))\]
\[\forall w R(w_0, w) \rightarrow T(s(a) \lor s(b), w)\]
Overview

- Omnipotent
- Model Operator Axioms
- Syntactic Proofs
  ⇒ Constants and Quantifiers

Constants and Quantifiers

- $B_j(\text{woman(mary)})$
  - Does the system and john agree on who mary is?
- $\text{woman(mary)} \land B_j(\text{woman(mary)})$
  - Now do they agree?
- $\exists x \ B_j(\text{woman(x)})$ versus $B_j(\exists x \ \text{woman(x)})$