MAIN TRACK:

LOGICS OF AGENCY

LOGICS OF KNOWLEDGE AND BELIEF

- logics of knowledge = epistemic logics
- logics of belief = doxastic logics

- Knowledge (naively) understood as true belief.

- Main problem with simplistic approaches: logical omniscience (agents are assumed to have perfect reasoning capability).
Traditional approach

(Hintikka, ...)

- Use the multi-modal logic determined by all reflexive (knowledge) \([\text{serial} (\text{belief})]\), transitive, euclidean multi-relational frames (one relation \(R(i)\) for each agent \(i \in I\)).

- This is (soundly & completely) axiomatized by \(K(I,45) = \bigcup_{i \in I} [K(I,45)_{\text{i} \in I}].\)

- Define \(K_i A \equiv [B, \text{A}]\) (i knows i believes) that \(A\) as \([i] A.\)

reflexivity: \[T:\]

\(\sim_i\) \(\sim_i\) \(\sim_i\) \(\sim_i\)

\(\sim_i\) \(\sim_i\) \(\sim_i\) \(\sim_i\)

seriality: \[D:\]

\(\sim_i\) \(\sim_i\) \(\sim_i\) \(\sim_i\)

\(\sim_i\) \(\sim_i\) \(\sim_i\) \(\sim_i\)

knowledge axiom:

\(\sim_i\) \(\sim_i\) \(\sim_i\) \(\sim_i\)

\(\sim_i\) \(\sim_i\) \(\sim_i\) \(\sim_i\)

what an agent knows, is true

belief axiom:

\(\sim_i\) \(\sim_i\) \(\sim_i\) \(\sim_i\)

\(\sim_i\) \(\sim_i\) \(\sim_i\) \(\sim_i\)

an agent does not believe the negation of what it believes

knowledge - consistent with the reality

beliefs - internally consistent
transitivity: \[ \langle \varepsilon \rangle A \sqsubseteq [\varepsilon] \langle i\rangle A \]

Euclidean: \[ \langle \varepsilon \rangle A \sqsubseteq [\varepsilon] \langle i\rangle A \quad \langle [\varepsilon] A \Rightarrow [\varepsilon] \langle i\rangle A \rangle \]

Positive introspection:
- an agent knows [believes]
- it knows [believes]
- what it knows [believes]

Negative introspection:
- an agent knows [believes]
- it doesn't know [believes]
- what it doesn't know [believes]

(Do we always want it??)

Logical omniscience

K:
\[ [\varepsilon] (A \rightarrow B) \Rightarrow [\varepsilon] A \rightarrow [\varepsilon] B \]

an agent's knowledge [belief] is closed under implications it knows [believes]

RN:
\[ \frac{A}{[\varepsilon] A} \]

an agent knows [believes] all tautologies
Example (of the power of these things!):

Suppose
\[ K_a q \]
Then
\[ K_a K_a q \] Alice knows she doesn't know that \( q \) (since she can do neg. introspection).
\[ K_a K_b K_a q \] Alice knows that Bob doesn't know she knows that \( q \) (Bob's knowledge is consistent with the reality to Alice's knowledge).
\[ K_a K_b K_b K_a q \] Alice knows that Bob in fact knows he doesn't know she knows that \( q \) (Alice knows Bob can do neg. introspection).

Common knowledge [belief]

- Introduce a pseudo-agent \( c \) ("any fool").

- Add this frame condition:
  For any \( w, w' \in W \),
  \[ w R(c) w' \] iff, for some \( \langle w_0, \ldots, w_n \rangle \in W, i_0, \ldots, i_n-1 \in I, w = w_0, w_0 R(i_0) w_1, \ldots, w_{n-1} R(i_{n-1}) w_n, w_n = w' \]

- Define \( K_c A \) [\( B_c A \)] (it is commonly known [believed] that \( A \), any fool knows [believes]) that \( A \) as \( [c] A \).
c behaves quite well as an agent:

From the assumption that all $R_i$'s (i.e. I) are reflexive (serial), it follows that so is $R(c)$; besides $R(c)$ is clearly transitive.

Check this!

- The naive logic of knowledge [belief] with "any fool" is (soundly & completely) axiomatized by $K_i(T4S)_i \in I$ [ $K_i(D4S)_i \in I$ ] plus

\[
[c]A \Rightarrow \land_{i \in I} [i](A \land [c]A)
\]

\[
B \Rightarrow \land_{i \in I} [i](A \land B)
\]

\[
| B \Rightarrow [c]A
\]
Distributed knowledge

- Introduce a pseudo-agent $d$ ("a wise man").

- Add the frame condition:
  for any $w, w' \in W$,
  \[ wR(d)w' \iff \text{for all } i \in I, \ wR(i)w', \]

- Define \( K_d A \) (a wise man would know that $A$) as \([d] A\).

- $d$ behaves very well as an agent wrt. knowledge:
  From the assumption that all $R_i$'s (i.e. $I$) are reflexive, transitive and euclidean, it follows that so is $R(d)$.

- The naive logic of knowledge with "a wise man" is (soundly & completely) axiomatized by
  \( K_i u[d] (T45) \) for $i \in I \cup \{d\}$ plus
  \[ [i] A \supset [d] A, \]
  if $|I| > 2$. 
Other approaches

- We shall consider three "fine" approaches:
  - Fagin and Halpern's logic of (general) awareness;
  - Fagin and Halpern's logic of local reasoning;
  - Lewisque's logic of only-believing (Duc's version of this).

Logic of (general) awareness

(Fagin, Halpern)

- In addition to the usual notion of implicit belief, captures a notion of explicit belief based on the notion of awareness.

Syntax

Assumed are a denumerable set $\text{FmA} = \{ p_0, p_1, \ldots \}$ of prop. letters and a (non-empty) set $\text{I}$ of agent identifiers. The set $\text{FmA}$ of formulas is defined as follows:

- If $p \in \text{FmA}$, then $p \in \text{FmA}$;
- $T, \bot \in \text{FmA}$;
- If $A \in \text{FmA}$, then $\neg A \in \text{FmA}$;
- If $A \in \text{FmA}$, then $\forall A \in \text{FmA}$.

- If $A \in \text{FmA}$, then $A; A$ (it's aware of $A$);
- $\mathbf{B}^i_A$ (i believes $A$ implicitly), $\mathbf{B}^{\exp}_i A$ (i believes $A$ explicitly) $\in \text{FmA}$. 
Strategies for (general) awareness

A structure for (general) awareness is a quadruple \( M = (W, R, A, V) \) where
* \( W \) is a non-empty set of worlds;
* \( R \in \text{[I\rightarrow P(W \times W)]} \) is a function from agent identifiers to \( \text{relational} \) (relations between \( W \) and \( W \) (accessibility relations);
* \( A \in \text{[I\rightarrow [\text{FmA} \rightarrow P(W)]]} \) is a function from agent identifiers to \( \text{functions} \) (functions) from \( W \) to \( \text{sets of \( W \)} \) (functions);
* \( V \in \text{[FmA} \rightarrow P(W) \] \) is a function from prop. letters to \( \text{sets of \( W \)} \) (valuation).

A formula is a tautology in the logic of (general) awareness, if it is valid in all structures for (general) awareness.

Satisfaction in structures for (general) awareness

Given a structure \( M = (W, R, A, V) \) for (general) awareness, the interpretation function \( \text{[I\rightarrow G]} \in \text{[FmA} \rightarrow P(W) \] \) is defined as follows (for we \( \text{[I\rightarrow G]} \) M):  
* If p \in \text{FmA}, then: \( k^M_A \text{ p if } p \in V(p) \);
* \( k^M_A \text{ T if } \text{p is true} \);
* \( k^M_A \text{ F if } \text{p is false} \);
* \( k^M_A \text{ if } \text{not } k^M_A \);
* \( k^M_A \text{ if } \text{A}; A \);
* \( k^M_A \text{ if } \text{A}; A \text{ for any } w \in W \text{ s.t. } w R; w', k^M_A \);  
* \( k^M_A \text{ if } \text{A}; k^M_A \text{ if } \text{A}; A \text{ and } \text{A}; A \text{ for any } w \in W \text{ s.t. } w R; w', k^M_A \).
Observations

The following hold for any structure $M$ for (general) awareness:

1. $\models^M B_i \exp A \equiv A; A \land B; A$

2. $\models^M B_i \exp (A \land B) \equiv (B_i \exp A \land B \implies B_i \exp B)$

3. If $\models^M A_i$, then $\models^M A; A \land B; A$

4. $\models^M \neg (B_i \exp A \land B_i \neg A)$

5. $\models^M B_i \exp A \land A; B_i \exp A \implies B_i \exp B_i \exp A$

6. $\models^M \neg B_i \exp A \land A; B_i \exp A \implies B_i \exp \neg B_i \exp A$

Axiomatization

The logic of (general) awareness is (soundly and completely axiomatized by $K_i$ plus

$$B_i \exp A \equiv A; A \land B; A$$
Logic of local reasoning (Fagin, Halpern)

- Agents as societies of minds, each mind with its own beliefs, possibly contradicting those of the fellow minds.

Syntax

Assumed again are a denumerable set \( \text{Fma} = \{ p, p', p'' \} \) of prop letters and a (non-empty) set \( \text{I} \) of agent identifiers. The set \( \text{Fma} \) of formulas is defined as follows:

- if \( p \in \text{Fma} \), then \( p \in \text{Fma} \);
- \( T, I \in \text{Fma} \);
- if \( A \in \text{Fma} \), then \( \neg A \in \text{Fma} \);
- \ldots
- if \( A \in \text{Fma} \), then \( B_i A \) (i believes A in some frame of mind), \( B_i \text{imp} A \) (i believes A implicitly, "between" its different frames of mind)
Structures for local reasoning

A structure for local reasoning is a triple \( M = (W, N, V) \) where:

* \( W \) is a non-empty set of worlds;
* \( N : \{ 1 \rightarrow \mathcal{P}(W \times \mathcal{P}(W)) \} \) is a function from agent identifiers to relations between worlds and sets of worlds (neighborhood relations), such that:
  * for any \( w \in W \), there exists an \( X \in W \) st \( wN X \);
  * for no \( w \in W \), \( wN \emptyset \);
* \( V \in \{ \text{true}, \text{false} \} \) is a function from agent identifiers to subsets of \( W \) (valuation).

Satisfaction in structures for local reasoning

Given a structure \( M = (W, N, V) \) for local reasoning, the interpretation function \( \llbracket \cdot \rrbracket^M : \{ \text{true}, \text{false} \} \rightarrow \mathcal{P}(W) \) for this structure is defined as follows:

\( \llbracket \phi \rrbracket^M \) (short for \( \llbracket \phi \rrbracket^M \)):

* If \( \phi \in \text{true} \), then: \( \llbracket \phi \rrbracket^M \) is true;
* If \( \phi \in \text{false} \), then: \( \llbracket \phi \rrbracket^M \) is false;
* \( \llbracket p \rrbracket^M \) is true if \( \phi \in V(p) \);
* \( \llbracket \neg A \rrbracket^M \) is true if \( \llbracket A \rrbracket^M \) is false;
* \( \llbracket A \lor B \rrbracket^M \) is true if \( \llbracket A \rrbracket^M \) or \( \llbracket B \rrbracket^M \) is true;
* \( \llbracket A \land B \rrbracket^M \) is true if \( \llbracket A \rrbracket^M \) and \( \llbracket B \rrbracket^M \) are true;
* \( \llbracket A \rightarrow B \rrbracket^M \) is true if \( \llbracket A \rrbracket^M \) is false or \( \llbracket B \rrbracket^M \) is true;
* \( \llbracket A \iff B \rrbracket^M \) is true if, for some \( X \in W \) st \( wN X \),
  * for any \( w \in X \), \( \llbracket A \rrbracket^M \);
  * for any \( w \in \{ X \mid wN X \} \), \( \llbracket B \rrbracket^M \).
Observations

- For any structure \( M \) for local reasoning, the following hold:
  - \( \text{if } \models^n A \Rightarrow B, \text{ then } \models^n B ; A \Rightarrow B ; B \)
  - \( \models^n B ; T \)
  - \( \text{if } \models^M A, \text{ then } \models^n B ; A \)
  - \( \models^n B ; A \Rightarrow B ; A \).

- There is a structure \( M \) s.t. \( \models^n B ; (p \Rightarrow T) \Rightarrow (B ; p \Rightarrow B ; p) \).
- There is a structure \( M \) s.t. \( \models^n \neg (B ; p \land B ; \neg p) \) and \( \models^n B ; \bot \).

Axiomatization

The logic of local reasoning \( \mathcal{N} \) (soundly and completely) axiomatized by \( E^I \) (\( M,N,P^i \)) \( i \in I \), i.e.,

\[
\begin{align*}
M_i &: B; (A \land B) \Rightarrow B; A \land B; B \\
N_i &: B; T \\
P_i &: \neg B; \bot
\end{align*}
\]

plus

\[
B; A \Rightarrow B; A \quad (i \in I)
\]

(\( |I| \geq 2 \)).
Logic of only-believing

(Levesque / Duc)

• Besides the usual believing-that (believing-at-least-that) operator, I study a believing-that and-only-that (believing-just-that) operator.
• For simplicity, assume just one agent.

Syntax

From a denumerable set \( \mathcal{F}_{\text{in}} = \{ p_0, p_1, \ldots \} \) of prop. letters, define the set \( \mathcal{F}_{\text{in}} \) of formulae as follows:

* if \( p \in \mathcal{F}_{\text{in}} \), then \( p \in \mathcal{F}_{\text{in}} \);
* \( T, \bot \in \mathcal{F}_{\text{in}} \);
* if \( A \in \mathcal{F}_{\text{in}} \), then \( \neg A \in \mathcal{F}_{\text{in}} \);
* \( \ldots \)
* if \( A \in \mathcal{F}_{\text{in}} \), then \( \mathcal{B}A \) (the system believes at least \( A \)), \( \mathcal{N}A \) (the system believes not at most \( A \)) \in \mathcal{F}_{\text{in}}.

For \( \mathcal{B}A \land \mathcal{N}A \) (the system believes exactly \( A \)), write \( \mathcal{D}A \).
Structures for only-believing

A structure for only-believing is a (uni)relational structure \( \mathcal{M} = (W, R, V) \) which is transitive and euclidean (note: that reflexivity is not required).

Satisfaction in structures for only-believing

For a structure \( \mathcal{M} = (W, R, V) \) for only-believing, the interpretation function \( [-]_\mathcal{M} : [\text{Formula}] \to \mathcal{P}(W) \) is defined as follows:
- If \( p \in \text{Formula} \), then: \( [p]_\mathcal{M} \) if \( we \in V(p) \);
- \( [\top]_\mathcal{M} = \{ w \in W \} \) is true;
- \( [\bot]_\mathcal{M} = \{ w \in W \} \) is false;
- \( [\neg A]_\mathcal{M} \) iff \( [A]_\mathcal{M} \) is false;
- \( [\land]_\mathcal{M} \) for any \( w \), \( \forall w' : wRw' \), \( [A]_\mathcal{M} \);,
- \( [\lor]_\mathcal{M} \) for any \( w \), \( \forall w' : wRw' \), \( [A]_\mathcal{M} \).

Some observations

- For any structure \( \mathcal{M} \) for only-believing, the following hold:
  - If \( [A]_\mathcal{M} \), then \( [\neg A \land BA]_\mathcal{M} \) (i.e., \( [\neg (\neg A \land BA)]_\mathcal{M} \));
  - If \( [A]_\mathcal{M} \), then \( [\neg BA \lor \neg A]_\mathcal{M} \);
Axiomatization

The logic of only-believing is (soundly & completely) axiomatized by \( R45 \) plus

\[
N(A \Rightarrow B) \Rightarrow (N \neg A \Rightarrow N \neg B)
\]

\[
\frac{A}{\neg A}
\]

\[
\neg S \subseteq (S \subseteq A \land \neg B) \Rightarrow \subseteq \neg (A \lor B)
\]

where \( S, S' \) are arbitrary finite (possibly empty) sequences of \( B's \) and \( \neg B's \)

(the Humberstone axiom scheme)