KARO logic (Van Linder et al.)

- Knowledge & Belief:
  - epistemic logic
- Abilities, Results & Opportunities:
  - dynamic logic
- Modalities for Desires & Goals

Epistemic logic

- $K\phi \rightarrow \phi$
- $K\phi \rightarrow KK\phi$
- $\neg K\phi \rightarrow K\neg K\phi$
- $\neg B\bot$
- $B\phi \rightarrow BB\phi$
- $\neg B\phi \rightarrow B\neg B\phi$

Dynamic Logic

- Syntax
  - Operator $[\alpha]$ with reading:
    - $[\alpha]\psi$ : after execution of $\alpha$ it holds (nec.) that $\psi$
    - $<\alpha>\psi = [\alpha]\psi$
- Semantics
  - Accessibility relation $R_\alpha$ for every action $\alpha$
    - $R_{\alpha^+} = R_\alpha \cup R_{\alpha}$
    - $R_\alpha = R_\alpha^*$

Dynamic Logic

- Interpretation formulas
  - $M, s \vDash [\alpha]\psi$ for all $s'$ with $R_\alpha(s, s')$:
    $M, s' \vDash \psi$
  - $M, s \vDash <\alpha>\psi$ for some $s'$ with $R_\alpha(s, s')$:
    $M, s' \vDash \psi$

Dynamic Logic

- Basic property (K)
  - $[\alpha](\psi \rightarrow \psi) \rightarrow ([\alpha]\psi \rightarrow [\alpha]\psi)$
- Structure of actions
  - $[\alpha_1; \alpha_2]\psi \leftrightarrow [\alpha_1]([\alpha_2]\psi)$
  - $[\alpha_1 + \alpha_2]\psi \leftrightarrow [\alpha_1]\psi \land [\alpha_2]\psi$
  - $[\alpha^*]\psi \rightarrow \psi$
  - $[\alpha^*]\psi \rightarrow [\alpha][\alpha^*]\psi$
  - $[\alpha^*](\psi \rightarrow [\alpha]\psi) \rightarrow (\psi \rightarrow [\alpha^*]\psi)$

KARO

- K: Knowledge (Belief)
  - epistemic logic:
    - Knowledge $K$ : the logic $S_5$
    - Belief $B$ : the logic weak $S_5$
- A: Abilities
  - ability operator $A$
**KARO**

- **R: Results** - dynamic logic ("multi-modal K"):
  - $[\alpha]\phi$

- **O: Opportunities** - dynamic logic:
  - $<\alpha>true$

**KARO: formal syntax (omitting agent indexes)**

- Set $A$ of atomic actions
- Set $P$ of atomic propositions

- Formulas $\phi ::= p (\in P) | \neg \phi | \phi_1 \land \phi_2 | ... |
  K_\alpha \phi | B_\phi | D_\psi | [\alpha]_\psi | A_\alpha$

- Actions $\alpha ::= a (\in A) | \alpha_1 ; \alpha_2 | \phi \ ? |
  \text{if} \ \phi \ \text{then} \ \alpha_1 \ \text{else} \ \alpha_2 \ \text{fi} |
  \text{while} \ \phi \ \text{do} \ \alpha \ \text{od}$

**KARO: model for knowledge/beliefs/desires**

- Kripke models of the form:
  $<W, \theta, R_K, R_B, R_D>$

  where:
  - $W$ is a non-empty set of states
  - $\theta$ truth assignment function per state
  - $R_K, R_B, R_D$ accessibility relations on $W$

**KARO: constraints on models**

- $R_K, R_B, R_D$ are accessibility relations on $W$
- $R_K$ is assumed to be an equivalence relation
- $R_B$ is assumed to be euclidean, serial and transitive
- $R_B \subseteq R_K$
- No special constraints on $R_D$

**KARO: modelling actions**

- Structures of the form:
  $<\Sigma, (R_a | a \in A), C, Ag>$

  where:
  - $\Sigma$ set of model/state pairs
  - $R_a (a \in A)$ accessibility relations on $\Sigma$
  - $C, Ag$ functions yielding set of actions the agent is capable to do, and the agent’s agenda, resp., per model/state pair

**KARO: constraints on structures**

- $R_a$ is an accessibility relation on (model, state) pairs!
- $R_a$ is taken to be deterministic:
  - $\{(M,w) | R_a(M,w)(M',w')\} \leq 1$ for all $M, w$
- If $R_a$ is deterministic we may write
  - $R_a(M,w) = \{(M',w')\}$ if $R_a(M,w)(M',w')$
  - $\emptyset$ otherwise
KARO: semantics of actions

- $R_{\alpha} (M, w) = \{(M', w')\}$ if $M, w \vdash \psi$
- $= \emptyset$ otherwise

$R_{\alpha_1} R_{\alpha_2} (M, w) = R_{\alpha_2} (R_{\alpha_1} (M, w))$

$R_{\alpha_1}$ then $\alpha_1$ else $\alpha_2$ if $M, w \not\vdash \psi$ otherwise

$R_{\alpha_1}$ otherwise.

$R_{\alpha} (X) = \bigcup_{(M, w) \in X} R_{\alpha} (M, w)$

- Note $R_{\alpha} \{(M, w)\} = R_{\alpha} (M, w)$

KARO: constraints on structures

- $C$ is a function of type $\Sigma \rightarrow P(\text{Actions})$
- $C(M, w)$ is the set of actions that the agent is capable of to perform in $(M, w)$
- One might impose conditions on $C$ regarding the structure of actions analogous to that of $R$
  - E.g., $\alpha_1, \alpha_2 \in C(M, w)$ if $\alpha_1 \in C(M, w)$ and $\alpha_2 \in C(R_{\alpha_1} (M, w))$

KARO: interpretation of formulas

- $M, w \vdash \phi \iff \theta(w)(\phi) = \text{true}$, for $\phi \in P$
- $M, w \vdash \neg \phi \iff M, w \not\vdash \phi$

- $M, w \vdash \phi_1 \land \phi_2 \iff M, w \vdash \phi_1$ and $M, w \vdash \phi_2$
- $M, w \vdash K_{\phi} \iff M, w' \vdash \phi$ for all $w'$ such that $R_{\phi} (w, w')$
- $M, w \vdash B_{\phi} \iff M, w' \vdash \phi$ for all $w'$ such that $R_{\phi} (w, w')$

KARO: interpretation of formulas (2)

- $M, w \vdash D_{\phi} \iff M, w' \vdash \phi$ for all $w'$ such that $R_{\phi} (w, w')$
- $M, w \vdash [\alpha]_{\phi} \iff M', w' \vdash \phi$ for all $M', w'$ such that $R_{\alpha} ((M, w), (M', w'))$
- $M, w \vdash A_{\alpha} \iff \alpha \in C(M, w)$
- $M, w \vdash Com_{\alpha} \iff \alpha \in Ag(M, w)$

Validities

- $\langle \phi \rangle \phi \iff (\phi \land \phi)$
- $\langle \alpha \rangle \phi \iff (\langle \alpha \rangle \phi) \land (\phi \land \langle \alpha \rangle \phi)$
- $\langle \alpha \rangle \phi \iff (\langle \alpha \rangle \phi) \land (\phi \land \langle \alpha \rangle \phi)$
- $\langle \phi \rangle \phi \iff (\phi \land \phi) \land (\phi \land \langle \alpha \rangle \phi)$
- $\langle \alpha \rangle \phi \iff \alpha \land \langle \alpha \rangle \phi$
KARO: correctness, feasibility

- **Correct**: $\alpha(\psi) = \langle \alpha \rangle \psi$
- **Feasible**: $\alpha(\psi) = \text{Feasible}(\alpha)$
- **PractPoss**: $\text{PractPoss}(\alpha, \psi) = \text{Correct}(\alpha, \psi) \land \text{Feasible}(\alpha)$
- **Can**: $\text{Can}(\alpha, \psi) = K\text{PractPoss}(\alpha, \psi)$
- **Cannot**: $\text{Cannot}(\alpha, \psi) = \neg K\text{PractPoss}(\alpha, \psi)$

Remark

- Note that $\langle \alpha \rangle \psi \rightarrow \langle \alpha \rangle \text{true}$
- I.e. correctness implies opportunity
- If $\alpha$ is deterministic:
  \[\langle \alpha \rangle \psi \leftrightarrow [\alpha] \psi \land \langle \alpha \rangle \text{true}\]
- So, for deterministic actions the diamond is stronger than the box:
  \[\langle \alpha \rangle \psi \text{ expresses both result and opportunity!}\]

KARO: properties of Can

- **some properties:**
  - $\text{Can}(\psi, \psi) \rightarrow K(\psi \rightarrow \psi)$
  - $\text{Can}(\alpha, \alpha) \Rightarrow \text{Can}(\alpha, \text{PractPoss}(\alpha, \psi))$
  - $\text{Can}(\alpha, \alpha, \psi) \rightarrow \langle \alpha \rangle \text{Can}(\alpha, \psi)$, if $\alpha$ accordant, i.e. $K(\alpha) \psi \rightarrow [\alpha] K(\psi)$
  - $\text{Can}(\text{if } \psi \text{ then } \alpha_1 \text{ else } \alpha_2, \psi) \land K(\psi) \leftrightarrow \text{Can}(\alpha_1, \psi)$
  - $\text{Can}(\text{if } \psi \text{ then } \alpha_1 \text{ else } \alpha_2, \psi) \land K(\psi) \leftrightarrow \text{Can}(\alpha_2, \psi)$

KARO: implementability

- **Impl**: $[\text{PractPoss}(a_1; \ldots; a_k, \psi)]$ for some atomic actions $a_1, \ldots, a_k$
  - $\psi$ is implementable/realizable/achievable by the agent by means of a plan $a_1; \ldots; a_k$. The agent has the pract. poss. (so is able and has opportunity) to realize $\psi$ by executing this plan.

KARO: informational attitudes

- **Belief types & belief revision in an agent-oriented setting:**
  - $K = B^c$ (certain) knowledge
  - $B^p$ belief by observation
  - $B^c$ belief by communication
  - $B^d$ belief by default

- **Agn**
  - $\neg B^p \psi \land \neg B^c \psi$ (x $\in \{k, o, c, d\}$)
  - $\text{Saw}\psi = \text{Agn}^o \psi \land B^o \psi$
  - $\text{Heard}\psi = \text{Agn}^c \psi \land B^c \psi$
  - $\text{Jumped}\psi = \text{Agn}^d \psi \land B^d \psi$
KARO: informational actions

- **observe, inform, try_jump**

- Semantics (written as functions):
  
  - $R_{\text{observe}}(M, w) = \text{update\_belief}(\phi, (M, w))$
  
  - $R_{\text{inform}(i)(\phi, (M, w))} = \text{update\_belief}(\phi, (M, w))$
  
  - $R_{\text{try\_jump}(\phi, (M, w))} = \text{update\_belief}(\phi, (M, w))$

KARO: semantics update_belief

\[
\text{update\_belief}(\phi, (M, w)) =
\]

KARO: informational attitudes

- Validities:
  
  - $\langle\text{observe } \phi\rangle \rightarrow \neg \text{Agn} \phi$
  
  - $\phi \rightarrow \langle\text{observe } \phi\rangle B \phi$
  
  - $\neg \phi \rightarrow \langle\text{observe } \phi\rangle B \neg \phi$
  
  - $\phi \land \text{Agn} \phi \rightarrow \langle\text{observe } \phi\rangle \text{Saw} \phi$
  
  - $\neg \phi \land \text{Agn} \phi \rightarrow \langle\text{observe } \phi\rangle \text{Saw} \neg \phi$
  
  - $\phi \land (\text{Heard} \neg \phi \lor \text{Jumped} \neg \phi) \rightarrow$
  
  - $\langle\text{observe } \phi\rangle \text{Saw} \phi$

KARO: informational attitudes

- Further validities:
  
  - $\text{Default}(\phi) \land \text{Agn} \phi \rightarrow$
  
  - $\langle\text{try\_jump } \phi\rangle \text{Jumped} \phi$
  
  - $\text{Default}(\phi) \land \neg \text{Agn} \phi \rightarrow$
  
  - $\langle\text{try\_jump } \phi\rangle \chi \leftrightarrow \chi$
KARO: motivational attitudes

- **Goal**: $\text{Goal}(\phi) = D_0 \wedge \neg \phi \wedge \text{Impl}$
  - A **goal** is desired, not yet true and implementable/realizable
  - **NOT**: $\phi \rightarrow \psi \Rightarrow \text{Goal}(\phi) \rightarrow \text{Goal}(\psi)$

- **Can**: $\text{Can}(\alpha, \phi) = \text{Can}(\alpha, \phi) \wedge \text{Goal}(\phi)$
  - **PossIntend**: $\text{PossIntend}(\alpha, \phi) = \text{PossIntend}(\alpha, \phi) \wedge K_{\text{Goal}}(\phi)$

- An agent **possibly intends** to do $\alpha$ to achieve $\phi$ if it knows that $\phi$ is its goal and can do $\alpha$ with $\phi$ as result.

- **Uncommit**
  - **Special actions**: commit$\alpha$, uncommit$\alpha$
    - **Semantics**:
      - $R_{\text{commit}}(M, w) = \text{update_agenda}(\alpha, (M, w))$ if
        $M, w \vdash \text{PossIntend}(\alpha, \phi)$ for some $\phi$, and $= \emptyset$, otherwise
      - $R_{\text{uncommit}}(M, w) = \text{update_agenda}(\alpha, (M, w))$ if
        $M, w \vdash \text{Com} \alpha$, and $= \emptyset$, otherwise
      - uncommit$\alpha \subseteq C(M, w)$ iff
        $M, w \vdash \lnot \text{PossIntend}(\alpha, \phi)$ for all $\phi$

KARO: motivational actions

- One may specify **specific agent types** that have certain policies of committing and uncommitting, for example:
  - Agents that may commit to one action only, i.e. the agenda $C(M, s)$ contains at most 1 element (‘single-minded agent’)
  - Agents that never uncommit (only implicitly after completing the execution of agenda items)
  - Agents that uncommit to an agenda item as soon as it detects that the item serves no purpose any more (i.e. PossIntend has ceased to hold: the agent cannot do the action any more or it does not lead to achieving any of the known goals)