KARO logic (Van Linder et al.)

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

130

Epistemic logic

- $K\varphi \rightarrow \varphi$
- $K\varphi \rightarrow KK\varphi$ *knowledge*
- $\neg K\varphi \rightarrow K\neg K\varphi$
- $\neg B\bot$
- $B\varphi \rightarrow BB\varphi$ *belief*
- $\neg B\varphi \rightarrow B\neg B\varphi$

132

Dynamic Logic

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

133

Dynamic Logic

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

134

Dynamic Logic

- Basic property (K)
  - $[\alpha](\varphi \rightarrow \psi) \rightarrow ([\alpha]\varphi \rightarrow [\alpha]\psi)$

Structure of actions

- $[\alpha_1 ; \alpha_2]\varphi \leftrightarrow [\alpha_1]([\alpha_2]\varphi)$
- $[\alpha_1 + \alpha_2]\varphi \leftrightarrow [\alpha_1]\varphi \wedge [\alpha_2]\varphi$
- $[\alpha^*]\varphi \rightarrow \varphi$
- $[\alpha^*]\varphi \rightarrow [\alpha][\alpha^*]\varphi$
- $[\alpha^*](\varphi \rightarrow [\alpha]\varphi) \rightarrow (\varphi \rightarrow [\alpha^*]\varphi)$

135

KARO

- **K: Knowledge (& Belief)**
  - epistemic logic:
    - Knowledge K : the logic S5
    - Belief B : the logic weak S5
- **A: Abilities**
  - ability operator A

136

KARO

- **R: Results**
  - dynamic logic ("multi-modal K"):
    - $[\alpha]\varphi$
- **O: Opportunities**
  - dynamic logic:
    - $\langle\alpha\rangle$true

137

KARO: formal syntax (omitting agent indexes)

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

- Formulas $\varphi ::= p (\in P) \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \ldots \mid K\varphi \mid B\varphi \mid D\varphi \mid [\alpha]\varphi \mid A\alpha$
- Actions $\alpha ::= a (\in A) \mid \alpha_1 ; \alpha_2 \mid \varphi? \mid$ if φ then α_1 else α_2 fi I while φ do α od

138

KARO: model for knowledge/beliefs/desires

- Kripke models of the form:

$$\langle W, \theta, R_K, R_B, R_D \rangle$$

where:

- W is a non-empty set of states
- θ truth assignment function per state
- R_K, R_B, R_D accessibility relations on W

140

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

141

KARO: modelling actions

- Structures of the form:

$$\langle \Sigma, \{R_a \mid a \in A\}, C, Ag \rangle$$

where:

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

142

KARO: constraints on structures

- R_a is an accessibility relation on *(model, state) pairs!*
- R_a is taken to be deterministic:
  - $\{(M',w') \mid R_a(M,w)(M',w')\} \leq 1$ for all M, w
- If R_α is deterministic we may write
  - $R_\alpha(M,w) = \{(M',w')\}$ if $R_\alpha(M,w)(M',w')$
    $= \emptyset$ otherwise

143

KARO: semantics of actions

- R_α is defined by induction on α:
  - $R_{\varphi?}(M,w) = \{(M,w)\}$ if M, w $\models \varphi$
    $= \emptyset$ otherwise
  - $R_{\alpha_1;\alpha_2}(M,w) = R_{\alpha_2}(R_{\alpha_1}(M,w))$
  - $R_{\text{if } \varphi \text{ then } \alpha_1 \text{ else } \alpha_2 \text{ fi}}(M,w) = R_{\alpha_1}(M,w)$ if M, w $\models \varphi$
    $= R_{\alpha_2}(M,w)$ otherwise

144

KARO: semantics of actions

- $R_{\text{while } \varphi \text{ do } \alpha \text{ od}}(M,w) = \{(M',w')\}$ iff
  exists $k \geq 0$ exists $M_0,w_0, \ldots, M_k,w_k$:
  $(M_0,w_0) = (M, w)$ and $(M_k,w_k) = (M',w')$ and
  $\{(M_{j+1},w_{j+1})\} = R_{\varphi?;\alpha}(M_j,w_j)$ and $M',w' \not\models \varphi$.
- $R_{\text{while } \varphi \text{ do } \alpha \text{ od}}(M,w) = \emptyset$ otherwise.

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

145

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. $\alpha_1; \alpha_2 \in C(M, w)$ iff $\alpha_1 \in C(M, w)$ and $\alpha_2 \in C(R_{\alpha_1}(M, w))$

146

KARO: interpretation of formulas

- $M,w \models p \Leftrightarrow \theta(w)(p) = \text{true}$, for $p \in P$
- $M,w \models \neg\varphi \Leftrightarrow M,w \not\models \varphi$
- $M,w \models \varphi_1 \wedge \varphi_2 \Leftrightarrow M,w \models \varphi_1$ and $M,w \models \varphi_2$
- $M,w \models K\varphi \Leftrightarrow M,w' \models \varphi$ for all w' such that $R_K(w, w')$
- $M,w \models B\varphi \Leftrightarrow M,w' \models \varphi$ for all w' such that $R_B(w, w')$

147

KARO: interpretation of formulas (2)

- $M,w \models D\varphi \Leftrightarrow M,w' \models \varphi$ for all w' such that $R_D(w, w')$
- $M,w \models [\alpha]\varphi \Leftrightarrow M',w' \models \varphi$ for all M',w' such that $R_\alpha((M,w), (M',w'))$
- $M,w \models A\alpha \Leftrightarrow \alpha \in C(M, w)$
- $M,w \models \text{Com}\alpha \Leftrightarrow \alpha \in Ag(M, w)$

148

Validities

- $\langle\varphi?\rangle\psi \leftrightarrow (\varphi \wedge \psi)$
- $\langle\alpha_1;\alpha_2\rangle\psi \leftrightarrow \langle\alpha_1\rangle(\langle\alpha_2\rangle\psi)$
- $\langle$if φ then α_1 else α_2 fi $\rangle\psi \leftrightarrow ((\varphi \wedge \langle\alpha_1\rangle\psi) \vee (\neg\varphi \wedge \langle\alpha_2\rangle\psi))$
- $\langle$while φ do α od$\rangle\psi \leftrightarrow ((\neg\varphi \wedge \psi) \vee (\varphi \wedge \langle\alpha\rangle\langle$while φ do α od$\rangle\psi))$
- $A(\alpha_1;\alpha_2) \leftrightarrow A\alpha_1 \wedge [\alpha_1]A\alpha_2$

149

KARO: correctness, feasibility

- Correct(α, φ) = $<\alpha>\varphi$
- Feasible(α) = $A\alpha$
- PracPoss(α,φ) = Correct(α,φ) $\wedge$ Feasible(α)
- Can(α,φ) = KPracPoss(α,φ) = K($<\alpha>\varphi \wedge A\alpha$)
- Cannot(α, φ) = K$\neg$PracPoss(α, φ)
  - An agent *can* do an action α with result φ iff it *knows* that it has the *pract. poss.* to do α, i.e. that α has φ as *result* and that it is *able* to do α.

150

Remark

- Note that $\models <\alpha>\varphi \rightarrow <\alpha>true$
  - I.e. correctness implies opportunity
- If α is *deterministic*:

  $\models <\alpha>\varphi \leftrightarrow [\alpha]\varphi \wedge <\alpha>true$
- So, for deterministic actions the diamond is *stronger* than the box:
  - $<\alpha>\varphi$ expresses both result and opportunity!

151

KARO : properties of Can

- some properties:
  - Can($\varphi?, \psi$) $\leftrightarrow$ K($\varphi\wedge\psi$)
  - Can($\alpha_1; \alpha_2, \varphi$) $\leftrightarrow$ Can(α_1, PracPoss(α_2, φ))
  - Can($\alpha_1; \alpha_2, \varphi$) $\rightarrow$ $<\alpha_1>$Can(α_2, φ) , if α_1 *accordant*, i.e. K$[\alpha_1]\varphi \rightarrow [\alpha_1]K\varphi$
  - Can(if φ then α_1 else α_2 fi, ψ) $\wedge$ Kφ $\leftrightarrow$ Can(α_1, ψ) $\wedge$ Kφ
  - Can(if φ then α_1 else α_2 fi, ψ) $\wedge$ K$\neg\varphi$ $\leftrightarrow$ Can(α_2, ψ) $\wedge$ K$\neg\varphi$

152

KARO: implementability

- Implφ = "PracPoss(a_1; ...; a_k, φ) for some atomic actions a_1, ..., a_k"
  - φ is *implementable / realizable / achievable* by the agent by means of a *plan* a_1; ...; a_k! The agent has the pract. poss. (so is able and has opportunity) to realize φ by executing this plan.

153

KARO : informational attitudes

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

154

KARO: informational attitudes

- $Agn^x = \neg B^x\varphi \wedge \neg B^x\neg\varphi$ $(x \in \{k, o, c, d\})$
- $Saw\varphi = Agn^k\varphi \wedge B^o\varphi$
- $Heard\varphi = Agn^o\varphi \wedge B^c\varphi$
- $Jumped\varphi = Agn^c\varphi \wedge B^d\varphi$

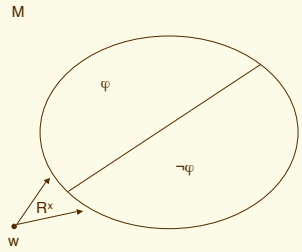
155

KARO: informational actions

- observe, inform, try_jump
- Semantics (written as functions):
  - $R_{observe\varphi}$ (M, w) = *update_belief*^{o}(φ, (M, w))
  - $R_{inform(\varphi,i)}$ (M, w) = *update_belief*$_i^{c}$(φ, (M, w))
  - $R_{try\_jump\varphi}$ (M, w) = *update_belief*^{d}(φ, (M, w))

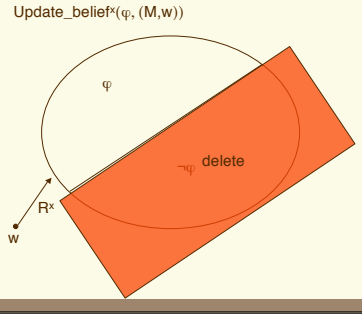
156

KARO: semantics update_belief



157

KARO: semantics update_belief



158

KARO : informational attitudes

- validities:
  - <observe φ>¬Agn^{o}φ
  - φ → <observe φ> B^{o}φ
  - ¬φ → <observe φ> B^{o}¬φ
  - φ ∧ Agn^{k}φ → <observe φ>Sawφ
  - ¬φ ∧ Agn^{k}φ → <observe φ>Saw¬φ
  - φ ∧ (Heard¬φ ∨ Jumped¬φ) → <observe φ>Sawφ

159

KARO : informational attitudes

- $B_j^{d}\varphi \wedge \neg D_{i,j}\varphi \rightarrow (<do_j(inform\ (\varphi, i)>\chi \leftrightarrow \chi)$
- $D_{i,j}\varphi \wedge B_j^{o}\varphi \rightarrow <do_j(inform\ (\varphi, i)> B_i^{c}\varphi$
- $D_{i,j}\varphi \wedge B_j^{o}\varphi \wedge Agn_i^{o}\varphi \rightarrow <do_j(inform\ (\varphi, i))>Heard_i\varphi$
- $D_{i,j}\varphi \wedge Heard_j\varphi \wedge Agn_i^{d}\varphi \rightarrow <do_j(inform\ (\varphi, i))> Heard_i\varphi$
- $D_{i,j}\varphi \wedge Heard_j\varphi \wedge \neg Agn_i^{d}\varphi \rightarrow (<do_j(inform\ (\varphi, i))>\chi \leftrightarrow \chi)$

[$D_{i,j}\varphi$: dependency of agent i on agent j w.r.t. φ]

160

KARO : informational attitudes

- Further validities:
  - Default(φ) ∧ Agn^{d}φ → <try_jump φ>Jumpedφ
  - Default(φ) ∧ ¬Agn^{d}φ → (<try_jump φ>χ ↔ χ)

161

KARO : motivational attitudes

- Goal(φ) = Dφ ∧ ¬φ ∧ Implφ
  - A *goal* is desired, not yet true and implementable/realizable
  - NOT: ⊨ φ → ψ ⇒ ⊨ Goal(φ) → Goal(ψ)
- CanG(α, φ) = Can(α, φ) ∧ Goal(φ)
- PossIntend(α,φ) = Can(α, φ) ∧ KGoal(φ)
  - An agent *possibly intends* to do α to achieve φ iff it knows that φ is its goal and *can* do α with φ as result.

KARO : motivational attitudes

- An agent that *possibly intends* to do action α to achieve result φ may choose to commit to action α, which means that the agent puts α on its agenda
- How an agent uses its agenda to determine the *actual next* action that it will perform is dependent on the type of agent and is *not* specified in KARO

KARO : motivational attitudes

- An agent that is committed to an action α may discover later that PossIntend(α, φ) *does not hold any more for any* φ.
- This may happen if (for all φ)
  - ¬Can(α, φ) holds: the agent cannot do α (with φ as result) any longer, or
  - ¬KGoal(φ) holds: φ isn't (known to be) a goal any more
- In this case the commitment to α has become *useless* and the agent may choose to *uncommit* to it.

KARO : motivational actions

- Special actions: commitα, uncommitα
  - Semantics:
    - $R_{commit\alpha}$(M, w) = *update_agenda*$^{+}$(α, (M, w)) if M,w ⊨ PossIntend(α, φ) for some φ, and = ∅, otherwise
    - $R_{uncommit\alpha}$(M, w) = *update_agenda*$^{-}$(α, (M, w)) if M,w ⊨ Com α, and = ∅, otherwise
    - uncommitα ∈ C(M, w) *iff* M,w ⊨ ¬PossIntend(α, φ) for *all* φ

KARO : motivational attitudes

- PossIntend(α, φ) → <commitα> Com(α)
- PossIntend(α, φ) → ¬Auncommitα
- Com(α) → <uncommitα> ¬Com(α)
- Com(α) → KCom(α)
- Com(α_1 ; α_2) → Com(α_1) ∧ K[α_1]Com(α_2)
- Com(α) ∧ ¬Can(α,true) → Can(uncommitα, ¬Com(α))

KARO : motivational attitudes

- 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)