COHEN AND LEVESQUE ON BDI

Camilo Thorne

22/5/2005

Introduction - Main Features

- There are only two primitive BDI modalities: **GOAL** (desire) and **BEL** (belief). The intention modality, i.e. **INT**, is defined by imposing persistence conditions on **GOAL**.
- The language includes variables (and constants) ranging over agents.
- There are action modalities together with action connectives. Control structures (iterative and conditional) can be defined.
- Possible worlds in models are discrete linear orders (finite or inifite) with a least lower bound. Furthermore, the accessiblity relations satisfy strong realism.
- No formal system is given.
- They can be called CL logics. In them, commitment is modelled in two different ways: blind-mindedness and open-mindedness.

Introduction – Desiderata

- Intentions provide a filter for adopting other intentions, which must not enter in conflict (coherency).
- Agents track the success of their intentions and are inclined to try again if their attempts fail.
- Agents believe their intentions are possible. They adopt intentions believe to be feasible.
- Agents need not hold their intentions forever. Only as long as they believe that the action purported has not, as yet, obtained.
- Intentions imply commitment to a certain action, which can be of various kinds.
- Agents need not intend all the possible consequences of their intentions.
- Agents may modify an intention due to belief-revision.
- Agents may adopt new intentions while planning.
- Commitment and intention is derived from the persistence, so to speak, of desires.

The Formal Language

Definition 1 *The set F' of CL formulae is recursively defined as follows:*

- $< numeral >::= 1|...|k,\ k \geq 0.$
- $< object - var >::= x_0|...|x_n,\ n \geq 0.$
- $< action - var >::= e_0|...|e_m,\ m \geq 0.$
- $< agent - var >::= a_0|...|a_p,\ p \geq 0.$
- $< pred >::= P_0^0|...|P_l^s,\ l, s \geq 0.$
- $< var >::=< object - var >$
  $| < event - var >$
  $| < agent - var >.$
- $< atom >::=< pred > (< var >, ..., < var >).$
- $< action - ex >::=< action - var >$
  $| < action - exp >; < action - exp >$
  $| < action - exp > | < action - exp >$
  $| < form >?$
  $| < action - exp >^*.$
- $< time - prop >::=< numeral >.$
- $< form >::=< atom >$

$$|\neg < form >$$
$$| < form > \wedge < form >$$
$$|\exists < var >< form >$$
$$|\textbf{Happ} < action - exp >$$
$$|\textbf{Done} < action - exp >$$
$$|\textbf{Ag} < agent - var >< action - var >$$
$$|\textbf{Bel} < agent - var >< form >$$
$$|\textbf{Go} < agent - var >< form >$$
$$| < time - prop >$$
$$| < action - var > \leqslant < action - var >.$$

Remarks

We omit the definition of the derived operators (such as the temporal operators), because unessential to this short summary, except for the following. We send the reader to [4] for further details.We further remark that the following assertion in true: $\exists! e \forall e' (e \leqslant e')$. We name it thus **NIL** – the empty action or sequence (of actions).

Some Important Derived Operators

Definition 2 *We put:*

- $\Diamond A =_{df} \exists e(\mathbf{Happ}\, e; A?)$.
- $\mathbf{Before} A\, B =_{df} \forall e(\mathbf{Happ}\, e; B?) \rightarrow \exists e'(e' \leqslant e) \wedge (\mathbf{Happ}\, e'; A?)$.
- $\mathbf{Single}\, e =_{df} (e \neq \mathbf{NIL}) \wedge (\forall(e' \leqslant e) \rightarrow (e' = e) \vee (e' = \mathbf{NIL}))$.
- $\mathbf{Done}_a e =_{df} (\mathbf{Done}\, e) \wedge (\mathbf{Ag}_a e)$.
- $\mathbf{Happ}_a e =_{df} (\mathbf{Happ}\, e) \wedge (\mathbf{Ag}_a e)$.

Formal Semantics - Models

Definition 3 *A model is a structure* $M = (\Theta, P, E, Agt, T, B, G, \Phi)$ *where:*

- Θ *is a set of* objects*.*
- P *is a set of* persons*.*
- E *is a set of* events*.*
- $Agt \in E \rightarrow P$ *is a function specifying the* agent *of a given event.*
- $T \subseteq \mathbb{Z} \rightarrow E$ *is a set of* worlds*.*
- $B \subseteq T \times P \times \mathbb{Z} \times T$ *is a* belief accessiblity *relation.*
- $G \subseteq T \times P \times \mathbb{Z} \times T$ *is a* goal accessiblity *relation.*
- $\Phi : Pred^k \times T \times \mathbb{Z} \rightarrow \Theta \cup P \cup E^*$ *is an* interpretation *function for predicate symbols.*

Formal Semantics – Domains

Definition 4 *Let M be a model. Then:*

- $D = \Theta \cup P \cup E^*$ *is called a* domain*.*
- $AGT \subseteq E^* \times P$ *is the set of* agents*.*

Formal Semantics – Satisfaction

Definition 5 *The* satisfaction *relation is defined by recursion on F':*

- $MP(x_1, ..., x_n)$ *iff* $(v(x_1), ...v(x_n)) \in \Phi(P, w, t)$.
- $M\neg A$ *iff* MA.
- $MA \vee B$ *iff* $M \models^{v}_{w_n} A$ *or* $M \models^{v}_{w_n} 2B$.
- $M\exists xA$ *iff* $M \models^{v^\star}_{w_n} A$ *for some* $\theta \in D$.
- $M\exists eA$ *iff* $M \models^{v^\star}_{w_n} A$ *for some* $\epsilon \in D$.
- $M\exists aA$ *iff* $M \models^{v^\star}_{w_n} A$ *for some* $\pi \in D$.
- $M\nu$ *iff* $v(\nu) = n$ *(for* $n \in \mathbb{N}$*)*.
- $Me \leqslant e'$ *iff* $v(e)$ *is an initial segment of* $v(e')$.
- $M\mathbf{Ag}_a\, e$ *iff* $Agt[v(e)] = \{v(e)\}$.
- $M\mathbf{Happ}\, \alpha$ *iff for some* $m \geq n$ *such that* $M, w, n[|\alpha|]m$.
- $M\mathbf{Done}\, \alpha$ *iff for some* $m \leq n$ *such that* $M, w, n[|\alpha|]m$.
- $M\mathbf{Bel}_a A$ *iff for any w' such that B(w,n,v(a),w') holds, we have that* $M \models^{v}_{w'_n} A$.
- $M\mathbf{Go}_a A$ *iff for any w' such that G(w,n,v(a),w') holds, we have that* $M \models^{v}_{w'_n} A$.

Formal Semantics – The Occurs Relation

Definition 6 *The* occurrence *of an action expression is defined by structural induction on action expressions as follows:*

- $M, w, m[|e|]n$ *iff* $v(e) = \epsilon_1 ... \epsilon_{n-m}$ *and* $w_{n+i} = \epsilon_i$ *for* $1 \leq i \leq n - m$.
- $M, w, m[|\alpha|\beta|]n$ *iff* $M, w, n[|\alpha|]m$ *or* $M, w, n[|\beta|]m$.
- $M, w, m[|\alpha; \beta|]n$ *iff for some* $m \leq k \leq n$, $M, w, m[|\alpha|]k$ *and* $M, w, k[|\beta|]n$.
- $M, w, m[|A?|]n$ *iff* $M \models^v_{w_n} A$.
- $M, w, m[|\alpha^*|]n$ *iff for some finite sequence* $, n_1, ..., n_k >$ *where* $n = n_1$, $m = n_k$ *and for every* $1 \leq j \leq k$, $M, w, n_j[|\alpha|]n_{j+1}$.

General Remarks on the Semantics and the Syntax

- *Truth* and *validity* are defined in a manner analogous to that of BDI logics.
- Regarding prrof theory, soundeness and completeness, the same remarks we made on BDI logics hold for Cohen and Levesque's sytem.

CL Logics – General Axioms

We will not enter into details regarding the axioms characterizing accessiblity in models. We remark only that we have the same axioms as in BDI logics. The following are more important:

Name	**Axiom**	**Property**
AX1	$\Diamond\neg(\mathbf{Go}_a(\mathbf{Later}\,A))$	Non infinite deferral
AX2	$\forall a \forall e[(\mathbf{Ag}_a e) \rightarrow \{(\mathbf{Done}\,e) \leftrightarrow (\mathbf{Bel}_a(\mathbf{Done}\,e))\}]$	Competence w.r.t. primitive actions
AX3	$\forall a \forall e[(\mathbf{Bel}_a(\mathbf{Happ}_a e; \alpha?)) \rightarrow (\mathbf{Bel}_a(\mathbf{Happ}_a e; (\mathbf{Bel}_a \alpha)?))]$	Beliefs about complex actions
AX4	$\forall a \forall e[(\mathbf{Ag}_a e) \wedge \mathbf{Single}\,e) \rightarrow (\mathbf{Bel}_a(\mathbf{Happ}\,e)) \vee(\mathbf{Bel}_a \neg(\mathbf{Happ}\,e))]$	No doubts relative to a course of action

CL Logics – Intention with Fanatical Commitment

1. Persistent goals:

Definition 7

$$\mathbf{P\text{-}Go}_a A =_{df} \begin{cases} \mathbf{Go}_a(\mathbf{Later} A) \wedge (\mathbf{Bel}_a \neg A) \wedge \\ \{\mathbf{Before}((\mathbf{Bel}_a A) \vee (\mathbf{Bel}_a \Box \neg A)) \\ \neg(\mathbf{Go}_a(\mathbf{Later} A))\}. \end{cases}$$

2. Intention:

Definition 8

$$\mathbf{In}^1_a \alpha =_{df} \quad (\mathbf{P\text{-}Go}_a\{\mathbf{Done}_a(\mathbf{Bel}_a(\mathbf{Happ}\,\alpha))?; a\}).$$

$$\mathbf{In}^2_a A =_{df} \begin{cases} (\mathbf{P\text{-}Go}_a \exists e(\mathbf{Done}_a\{(\mathbf{Bel}_a \exists e'(\mathbf{Happ}_a\, e'; A?)) \\ \wedge \neg(\mathbf{Go}_a \neg(\mathbf{Happ}_a\, e; A?))\}?; e; A?)). \end{cases}$$

CL Logics – Intention with Single-minded Commitment

1. Persistent relativized goals:

Definition 9

$$\textbf{PR-Go}_a A\, B =_{df} \begin{cases} \textbf{Go}_a(\textbf{Later} A) \wedge (\textbf{Bel}_a \neg A) \wedge \\ \{\textbf{Before}((\textbf{Bel}_a A) \vee (\textbf{Bel}_a \Box \neg A) \vee (\textbf{Bel}_a \neg B)) \\ \neg(\textbf{Go}_a(\textbf{Later} A))\}. \end{cases}$$

2. Intention:

Definition 10

$$\textbf{In}^1_a \alpha\, A =_{df} \quad (\textbf{PR-Go}_a \{\textbf{Done}_a(\textbf{Bel}_a(\textbf{Happ}\,\alpha))?; a\}\, A).$$

$$\textbf{In}^2_a A\, B =_{df} \begin{cases} (\textbf{PR-Go}_a \exists e(\textbf{Done}_a\{(\textbf{Bel}_a \exists e'(\textbf{Happ}_a\, e'; A?)) \\ \wedge \neg(\textbf{Go}_a \neg(\textbf{Happ}_a\, e; A?))\}?; e; A?)\, B). \end{cases}$$

References

[1] Amal El Fallah-Seghrouchni Alejandro Guerra-Hernandez and Henry Soldano. Learning in bdi multi-agent systems. http://centria.di.fct.unl.pt/ jleite/climalV/12.pdf, 2004.

[2] Michael Bratman. *Intentions, Plans and Practical Reason.* Harvard U. Press, 1987.

[3] Stuart Chalmers and Peter M.D.Gray. Bdi agents and constraint logic. *Artificial Intelligence and Simulation of Behaviour*, (1), 2001.

[4] Philip Cohen and Hector Levesque. Intention is choice with commitment. *Artificial Intelligence*, (42), 1990.

[5] Roberta Ferraio and Alessandro Oltramari. Towards a computational ontology of mind. In *Formal Ontology in Information Systems'*, 2004.

[6] Alejandro Guerra Hernandez. *Apprentissage d'agents rationnels BDI dans un univers multi-agents.* PhD thesis, Université de Paris 13, 2003.

[7] James Harland and Michael Winikoff. Agents via mixed-mode computation in linear logic. http://citeseer.ist.psu.edu/harland01agents.html, 2003.

[8] Andreas Herzig and Dominique Longin. Beliefs, goals and intentions. http://www.irit.fr/ Adreas.Herzig, 2003.

[9] Lin Padgham John Thangarajah and James Harland. Representation and reasoning for goals in bdi agents. In *Proceedings of the 25th Australasian Conference on Computer Science*, pages 259–265. Australian Computer Society, 2002.

[10] Vinet Chand PADMANABHAN NAIR. *On Extending BDI Logics.* PhD thesis, Faculty of Enginneering and Information Technology, Griffith University, Queensland, 2003.

[11] Willem VISSER Rafael H. BORDINI, Michael FISHER and Michael WOOLDRIDGE. Verifiable multi-agent programs. http://www.cs.uu.nl/ProMAS/2003/papers/paper1.pdf, 2003.

[12] Anand RAO and Michael GEORGEFF. Modelling rational agents within a bdi-architecture. http://citeseer.ist.psu.edu/122564.html, 1991.

[13] Anand S. RAO. Agentspeak(l): Bdi agents speak out in a logical computable language. In *Seventh European Workshop on Modelling Autonomous Agents in a Multi-Agent World*, 1996. http://citeseer.ist.psu.edu/rao96agentspeakl.html.

[14] John SEARLE. *Intentionality.* Cambridge U. Press, 1999.

[15] Guido GOVERNATORI Vineet PADMANABHAN and Abdul SATTAR. Actions made explicit in bdi. In *Advances in Artificial Intelligence*, pages 390–401. Springer-Verlag, 2001.

[16] Michael WOOLDRIDGE. *An Introduction to Multi-Agent Systems.* Wiley and Sons, 2002.

[17] Michael WOOLRIDGE. *Reasoning about Rational Agents.* The MIT Press, 2000.