THE BDI MODEL OF AGENCY AND BDI LOGICS

Camilo Thorne

22/4/2005

Introduction – Beliefs, Desires, Intentions (Bratman)

- Desires and beliefs range over states of affairs, while intentions range over actions and by extension, plans.
- Intentions are persistent, whereas desires can be dropped at any time.
- Intentions need not be holded forever.
- Intentions drive means-end reasoning.
- Beliefs constrain desires.
- Intentions constrain future deliberation and planning.
- Intentions influence beliefs upon which future practical reason is based.
- Intentions imply a degree of commitment to a goal.
- Intentions and beliefs are required to be consistent, i.e. not to imply some kind of pragmatical contradiction. This condition is assumed to imply that of rationality.
- Intentions, beliefs and desires need not be complete or, to put it simply, all-encompassing.
- Beliefs are subject to revision.
- Intentions and hence plans can be reconsidered.

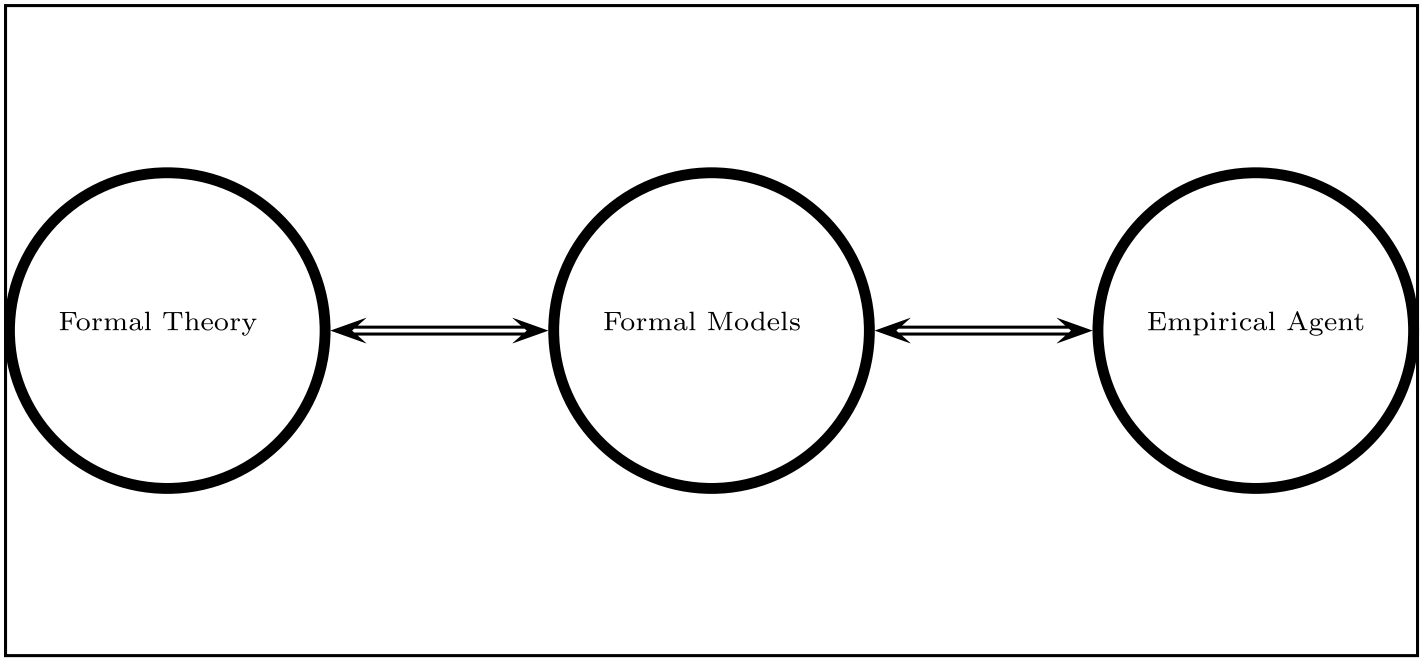
Introduction – Agents (Bratman)

Definition 1 *We understand by an* agent *an entity (a moral or a legal person, a computer program) that is capable of reacting to a certain environment through its performing a certain number of actions over which it can exert some kind of control. We say further that an agent is* rational *if his actions, decisions, plans and intentions are consistent or coherent with his beliefs and desires as well as between themselves.*

Introduction – Kinds of Agents (Bratman)

- *Blindly-minded agents* are agents that are blindy over-commited to their basic beliefs and intentions or desires, which they never put in question nor revise. They can be seen as fanatical agents.
- *Single-minded agents* are agents whose (derived) intentions may change due to belief revision. They are thus cautiously commited to their intentions. They can be seen as cautious agents. They are able to modify a plan if needed.
- *Open-minded agents* are agents that revise their beliefs and that change both their desires and derived intentions accordingly. They are thus under-commited to their intentions.

Introduction – Modelling Desiderata



BDI Logic Language Primitives

Definition 2 *The set $\mathcal{F}$ of the* formulae *of BDI logic is defined by the grammar:*

- $< o - var >::= x_1|...|x_n,\ n \geq 0.$
- $< e - var >::= e_1|...|e_m,\ m \geq 0.$
- $< pred >::= P_0^0|...|P_k^l,\ k, l \geq 0.$
- $< var >::=< o - var > | < e - var >.$
- $< atom >::=< pred > (< var >, ..., < var >).$
- $< state - form >::= succeeded(< event - var >)|failed(< event - var >)$
  $| < atom > |\neg < state - form > | < state - form > \vee < state - form >$
  $|\exists < var >< state - form > |\mathbf{Bel} < state - form >$
  $|\mathbf{Go} < state - form > |\mathbf{In} < state - form > |optional < path - form > .$
- $< path - form >::=< state - form > |\neg < path - form >$
  $| < path - form > \vee < path - form > | < path - form > \mathbf{U} < path - form >$
  $|\diamond < path - form > |\bigcirc < path - form > .$
- $< form >::=< state - form > | < path - form >.$

Formal Semantics – Models

Definition 3 *A* model *for BDI logics is a Kripke branching-time temporal model with three distinct accessiblity relations. Let $\mathcal{R}$ denote the set of propositional symbols. Then a model is a structure $M = (D, E, T, W; \prec, I, B, G; \Phi)$ where:*

- D *is a non-empty set called* domain of objects.
- E *is a non-empty set called* domain of events.
- T *is a non-empty set of* time points.
- $\prec \subseteq T \times T$ *is the branching time relation.*
- W *is a non empty set of worlds over* T*.*
- $I \subseteq W \times T \times W$ *an* intention accessibility *relation.*
- $B \subseteq W \times T \times W$ *a* belief accessibility *relation.*
- $G \subseteq W \times T \times W$ *a* goal accessibility *relation.*
- $\Phi : \mathcal{R} \times W \times T \to \bigcup_{i \in \mathbb{N}} \wp(D^i)$ *is an* interpretation *function for predicate symbols.*

Formal Semantics – Satisfaction

Definition 4 *The* satisfaction *relation is defined by induction on* $\mathcal{F}$ *as follows – first on path formulas and then on state formulas. Let* v *be an assignation and* $v*$ *be an assignation identical to* v *but for some object or event variable* x *or* e*. Then:*

- $M \models^{v}_{w_t} P(x_1, ..., x_n)$ *iff* $(v(x_1), ...v(x_n)) \in \Phi(P, w, t)$.
- $M \models^{v}_{w_t} \neg A$ *iff* $M \not\models^{v}_{w_t} A$.
- $M \models^{v}_{w_t} A \vee B$ *iff* $M \models^{v}_{w_t} A$ *or* $M \models^{v}_{w_t} B$.
- $M \models^{v}_{w_t} \exists x^O A$ *iff* $M \models^{v^\star}_{w_t} A$ *for some* $d \in D_O$.
- $M \models^{v}_{w_t} \exists x^E A$ *iff* $M \models^{v^\star}_{w_t} A$ *for some* $e \in D_E$
- $M \models^{v}_{<w_{t_0}, w_{t_1}, ...>} A$ *iff* $M \models^{v}_{w_{t_0}} A$.
- $M \models^{v}_{<w_{t_0}, w_{t_1}, ...>} \bigcirc A$ *iff* $M \models^{v}_{<w_{t_1}, ...,>} A$
- $M \models^{v}_{<w_{t_0}, w_{t_1}, ...>} \Diamond A$ *iff for some* $i \geq 0$ *such that* $M \models^{v}_{<w_{t_i}, ...,>} A$.
- $M \models^{v}_{<w_{t_0}, w_{t_1}, ...>} A\mathbf{U}B$ *iff either of these conditions hold:*
  1. *For some* $i \geq 0$ *such that* $M \models^{v}_{<w_{t_i}, ...>} B$ *and for all* $0 \leq j < i, M \models^{v}_{<w_{t_j}, ...,>} A$.
  2. *For any* $j \geq 0, M \models^{v}_{<w_{t_j}, ...>} A$.
- $M \models^{v}_{w_{t_0}}$ *optional*A *iff for some fullpath* $< w_{t_0}, w_{t_1}, ... >$, $M \models^{v}_{<w_{t_0}, w_{t_1}, ...>} A$.

- $M \models^v_{w_t}$ *succeded*(e) *iff for some time point* t', $S_w(t', t) = v(e)$.
- $M \models^v_{w_t}$ *failed*(e) *iff for some time point* t', $F_w(t', t) = v(e)$.
- $M \models^v_{w_t} \mathbf{Bel}A$ *iff for any* $w' \in \mathcal{B}^w_t$, $M \models^v_{w'_t} A$.
- $M \models^v_{w_t} \mathbf{In}A$ *iff for any* $w' \in \mathcal{I}^w_t$, $M \models^v_{w'_t} A$.
- $M \models^v_{w_t} \mathbf{Go}A$ *iff for any* $w' \in \mathcal{G}^w_t$, $M \models^v_{w'_t} A$.

Example of Model

The following model M_1 satisfyies the formula

$$\mathbf{In}(optional(\Diamond(does(eating)))) \rightarrow \mathbf{Go}(optional(\Diamond(does(eating)))$$

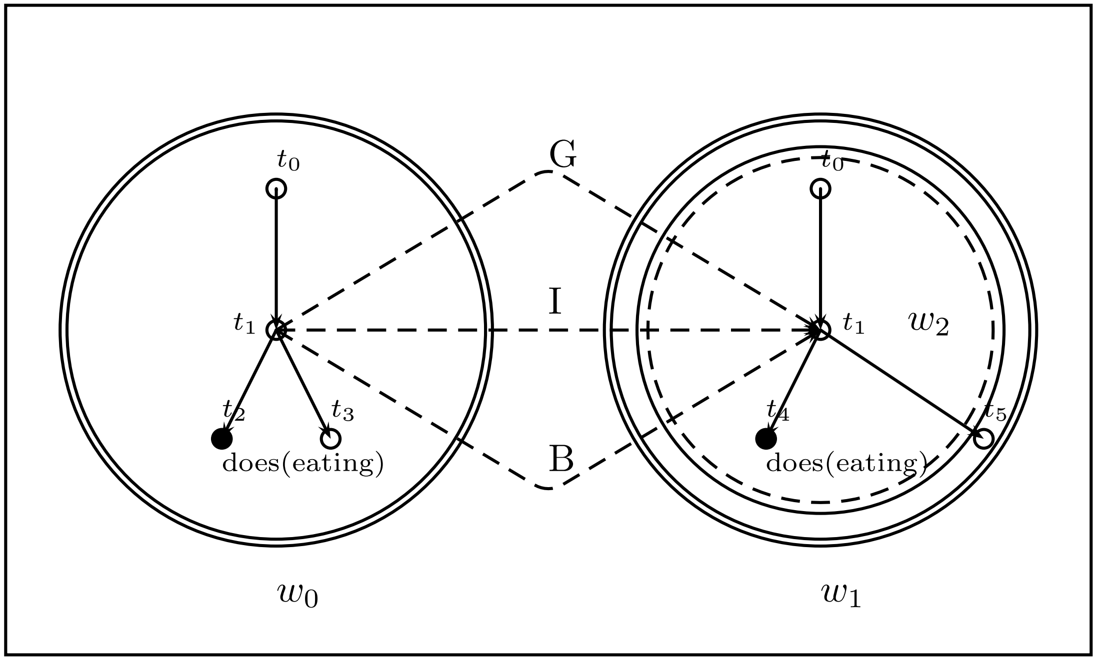
For indeed we have that:

$$M_1 \models_{w_{0_{t_1}}} \mathbf{In}(optional(\Diamond(does(eating))))$$

implies that:

$$M_1 \models_{w_{0_{t_1}}} \mathbf{Go}(optional(\Diamond(does(eating))).$$

The Model M_1



M_1

Correspondence Theory

Name	Modal Formula Scheme	Condition on Model
BK	$\mathbf{Bel}(A \to B) \to (\mathbf{Bel}A \to \mathbf{Bel}B)$	B non empty.
BD	$\mathbf{Bel}A \to \neg\mathbf{Bel}\neg A$	B is serial.
B4	$\mathbf{Bel}A \to \mathbf{BelBel}A$	B is transitive.
B5	$\neg\mathbf{Bel}\neg A \to \mathbf{Bel}\neg\mathbf{Bel}\neg A$	B is euclidian.
IK	$\mathbf{In}(A \to B) \to (\mathbf{In}A \to \mathbf{In}B)$	I non empty.
ID	$\mathbf{In}A \to \neg\mathbf{In}\neg A$	I is serial.
GK	$\mathbf{Go}(A \to B) \to (\mathbf{Go}A \to \mathbf{Go}B)$	G non empty.
GD	$\mathbf{Go}A \to \neg\mathbf{Go}\neg A.$	G is serial.
G-B	$\mathbf{Go}A \to \mathbf{Bel}A$	$G \subseteq B$.
I-G	$\mathbf{In}A \to \mathbf{Go}A$	$I \subseteq G$.
G-B*	$\mathbf{Go}\alpha \to \mathbf{Bel}\alpha$	$G \subseteq_{struct} B$.
I-G*	$\mathbf{In}\alpha \to \mathbf{Go}\alpha$	$I \subseteq_{struct} G$.

Basic Axioms

Name	Modal Formula Scheme	Intuitive property
BK	$\mathbf{Bel}(A \to B) \to (\mathbf{Bel}A \to \mathbf{Bel}B)$	Belief implication closure
BD	$\mathbf{Bel}A \to \neg\mathbf{Bel}\neg A$	Belief consistency
B4	$\mathbf{Bel}A \to \mathbf{BelBel}A$	Belief positive introspection
B5	$\neg\mathbf{Bel}\neg A \to \mathbf{Bel}\neg\mathbf{Bel}\neg A$	Belief negative introspection
IK	$\mathbf{In}(A \to B) \to (\mathbf{In}A \to \mathbf{In}B)$	Intention implication closure
ID	$\mathbf{In}A \to \neg\mathbf{In}\neg A$	Intention consistency
GK	$\mathbf{Go}(A \to B) \to (\mathbf{Go}A \to \mathbf{Go}B)$	Goal implication closure
GD	$\mathbf{Go}A \to \neg\mathbf{Go}\neg A.$	Goal consistency
G-B*	$\mathbf{Go}\alpha \to \mathbf{Bel}\alpha$	Desire-belief compatibility
I-G*	$\mathbf{In}\alpha \to \mathbf{Go}\alpha$	Intention-desire compatibility

Intention Axioms

Name	Axiom Scheme	Intuitive property
A11	$\mathbf{In}A \rightarrow \mathbf{Bel}(\mathbf{In}A)$	Intentions about beliefs
A12	$\mathbf{Go}A \rightarrow \mathbf{Bel}(\mathbf{Go}A)$	Beliefs about goals
A13	$\mathbf{In}A \rightarrow \mathbf{Go}(\mathbf{In}A)$	Desires about intentions
A14	$\forall e(\mathbf{In}(does(e)) \rightarrow does(e)$	Intentions leading to actions
A15	$\forall e(done(e)) \rightarrow \mathbf{Bel}(done(e))$	Awareness of primitive events
A16	$\mathbf{In}A \rightarrow inevitable(\Diamond(\neg\mathbf{In}A))$	No infinite deferral property

Commitement Axioms

Name	**Axiom Scheme**	**Intuitive property**
C1	$\mathbf{In}(inevitable(\Diamond A)) \rightarrow$ $inevitable(\mathbf{In}(inevitable(\Diamond A))\mathbf{U}A)$	Blind-mindedness property
C2	$\mathbf{In}(inevitable(\Diamond A)) \rightarrow$ $inevitable(\mathbf{In}(inevitable(\Diamond A))$ $\mathbf{U}(A \vee \neg\mathbf{Bel}(optional(\Diamond A)))$	Single-mindedness property
C3	$\mathbf{In}(inevitable(\Diamond A)) \rightarrow$ $inevitable(\mathbf{In}(inevitable(\Diamond A))$ $\mathbf{U}(A \vee \neg\mathbf{Go}(optional(\Diamond A)))$	Open-mindedness property

The Systems

Definition 5 *The axioms without the commitment axioms constitute the I system: the* Basic Intention System*. They model basic agents. Commitement axioms extend it to cover different kinds of agents.*

Properties

- $I \vdash \mathbf{In}\alpha \rightarrow \mathbf{Bel}\alpha$
- $I \vdash (\mathbf{Go}\alpha \wedge \mathbf{In}(\alpha \rightarrow \beta)) \rightarrow \mathbf{Go}\beta$.
- $I \vdash (\mathbf{Bel}\alpha \wedge \mathbf{In}(\alpha \rightarrow \beta)) \rightarrow \mathbf{Bel}\beta$.
- $I \vdash (\mathbf{Bel}\alpha \wedge \mathbf{Go}(\alpha \rightarrow \beta)) \rightarrow \mathbf{Bel}\beta$.

Extensions

Name	Modal Formula Scheme	Intuitive property
E1	$\neg(\mathbf{In}A \wedge \mathbf{Bel}\neg A)$	Intention-belief consistency
E2	$\mathbf{In}A \wedge \neg BelA$	Intention-belief incompleteness
E3	$\mathbf{Bel}A \wedge \neg\mathbf{Go}A$	Transference property
E4	$\mathbf{In}A \wedge \mathbf{Bel}(A \rightarrow B)$	Side effects property
	$\wedge\neg\mathbf{In}A$	

Limitations

Some extensions do not work, for example:

$$I + \text{E2} \vdash \bot$$

(i.e. Bratman's asymmetry thesis) Since:

1. $\mathbf{In}A \rightarrow \mathbf{Bel}A$ – Proposition 2.3
2. $\neg(\mathbf{Go}A \wedge \neg\mathbf{Bel}A)$ – 1,PL
3. $\mathbf{In}A \wedge \neg\mathbf{Bel}A$ – E2
4. $\bot$ – 2,3,PL

Cohen and Levesque

- 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.

Conclusions

- We ignore if full BDI logics are sound or complete, although we assume that they are sound. They are, anyway, undecidable.
- Different axioms convey different properties of agents.
- Formal constrainment of future desires, beliefs and intentions by present intentions is not very intuitive.
- They are a good specification tool (used to develop expremental software agents).
- Not all of Bratman's *provisi* and thesis hold.
- BDI logics model only deliberation and not means-end reasoning.

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.