BeliefDesireIntention

Andre Vellino's lecture slides available here [2].

(adapted from Ramiro Liscano's BeliefDesireIntentionLecture and the Weiss book. See SoftwareAgentBooks)

motivation: we want to model and design agents that act rationally. A rational agent is one that can come up with a PLAN OF ACTIONS consistent with its BELIEFS in order to achieve the GOALS it has been designed for.

"If I have a goal of staying dry, and I believe it is raining, then it is rational of me to take an umbrella when I leave my house" (Wooldridge in "Reasoning about Rational Agents")

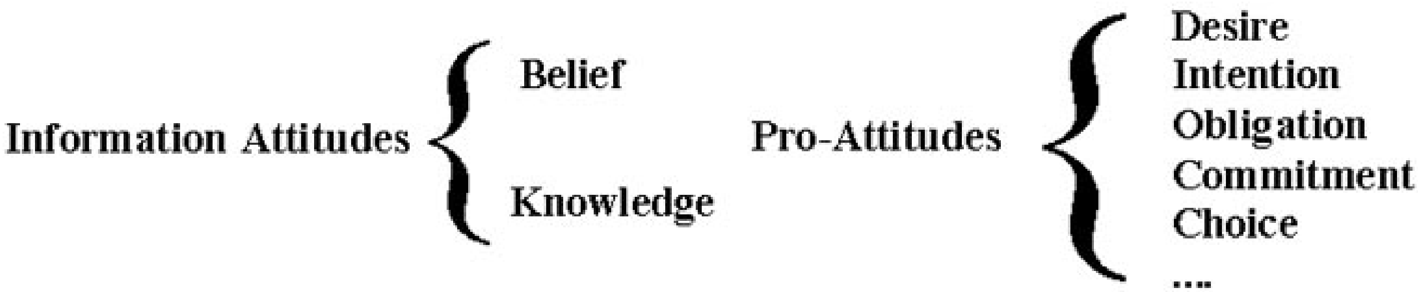
Note that for the agent to behave rationally, it does not matter whether it is indeed raining or not.

See also Dennett's intentional stance and the Mc Carthy quote in AgentDefinition.

3 components to BeliefDesireIntention:

- a philosophical theory of practical reasoning (Bratman)
- a logical foundation (Shoham, Cohen & Levesque, Wooldridge...)
- an implementation framework (Shoham, Rao & Georgeff)

mental constructs: Practical reasoning vs theoretical reasoning



- *Beliefs* are a set of statements that an agent has about the world and itself.
- *Desires* are a set of beliefs that an agent would like to reach
- *Intentions* are a subset of desires that the agent selects, through a deliberation process, and commits to achieving.

properties:

- consistency (between intentions, between intentions and beliefs, between beliefs...)
- persistence
  - *Intention is choice with commitment.* (Cohen & Levesque)
  - *beliefs persist by default*, and their absence as well, until the belief is learnt (Shoham) [4]
- good faith: only commit to what you believe you are capable of
- introspection

- intentions drive means-ends reasoning
- more: see TheoryOfIntentions

formal background: Epistemic logic, based on ModalLogic.

implementation: BdiArchitecture

issues:

- mapping of modal logic to the proposed architecture?
  - omniscience and other theorem proving issues
  - resulting performance
- finding a good balance between overcommitment to an intention and being overly cautious

refs:

Stanford's encyclopedia of philosophy: [3]

(last edited September 22, 2005)
Find Page by browsing or searching

BeliefDesireIntentionLecture

(this is Ramiro Liscano's lecture on BeliefDesireIntention, given in fall 99)

Simple Non-Operational Definition (David Kinny-Australian Artificial Intelligence Institute)

1. Is embedded in an environment from which it receives "events" that carry information about the state of the environment, and within which it can perform "actions" that modify that environment.

2. Contains representations (as objects, data structures, or whatever) of:
   - "beliefs", which constitute its knowledge of the state of its environment (and perhaps also some internal state),
   - "desires", which determine its motivation - what it is trying to bring about, maintain, find out, etc., and
   - "intentions", which capture its decisions about how to act in order to fulfil its desires. It should be possible, by an internal examination, to say at any instant precisely what the "mental state" (B's, D's, and I's) of an agent is.

3. Has a control mechanism which ensures that:
   - Its beliefs change over time in response to external events (and perhaps due to internal actions),
   - Its intentions determine and cause sequences of actions to be taken,
   - Its intentions change over time as a result of its beliefs changing, its desires becoming fulfilled or failing to be fulfilled, actions being taken, and new events being received.

Topics

- Introduction (BdiIntroduction).
- Intentional Systems (IntentionalAgents).
- Intentional Attitudes (IntentionalAttitudes).
- Intentional Notions (IntentionalNotions).
- Possible Worlds & Modal Logic (PossibleWorlds).
- Normal Modal Logic (NormalModalLogic).
- Theory of Intentions (TheoryOfIntentions).
- Possible Worlds Example (PossibleWorldsExample).
- BDIM Agent Toolkit (BdimAgentToolkit).
- BDI Agent Issues (BdiAgentIssues).

References

Cohen, P. R. and Levesque, H. J., 1990a. "Intention is choice with comitment". *Artificial Intelligence*, **42**, pp. 213-

61.

Cohen, P. R. and Levesque, H. J., 1990b. "Rational interaction as the basis for communications". *Intentions in Communication*, pp. 221-56, MIT Press.

Rao, A. S. and Georgeff, M. P., 1991. "Modelling rational agents within a BDI architecture". *In Proc. of Knowledge Representation and Reasoning (KR&R-91)*, pp. 473-84, Morgan Kaufman.

Rao, A. S. and Georgeff, M. P., 1992. "An abstract architecture for rational agents". *In Proc. of Knowledge Representation and Reasoning (KR&R-92)*, pp. 439-49.

Shoham, Y., 1993. "Agent oriented programming", *Artificial Intelligence*, **60(1)**, pp. 51-92.

Woolridge, M. and Jenningd, N. R., 1995. "Intelligent agents: theory and practice". *The Knowledge Engineering Review*, **10(2)**, pp. 115-52.

Links to other Sources

- Agents in Complex Environments (BDI) [1]

[SoftwareAgentCourse | SoftwareAgentCourse | BdiIntroduction]

(last edited November 1, 2000)
Find Page by browsing or searching

BdiIntroduction

BDI Control Mechanism

Much theoretical work has occured that explores the consequences of different decisions and offer frameworks for reasoning about formal abstract models of mental states, there is still a large gap between this work and implemented BDI architectures.

System Complexity

BDI agent architectures tend to be very complex systems, just in terms of the amount of functionality that is built in to the architecture itself. But BDI systems appear to have significant advantages in complex multi-agent applications characterised by uncertainty, uncontrolled environmental change, and ongoing changes in specification.

BDI Advantages

- Explicit representations of the B's, D's, and I's that you can point to and reason about.
- The agents can be viewed at a level of abstraction that facilitates understanding of what it is doing and why.
- It is possible to meaningfully make statements like "Agent X wrongly believes that the fuel manifold is leaking".
- Specifications of desired system behaviour are much more easily developed into implementations of that behaviour, because the BDI architecture provides a rich, high-level "virtual machine" for the programmer that operates at a level of abstraction much closer to the specification.

AOP over OOP

- The level of abstraction for OOP is much lower than AOP and most of the design needs to be redeveloped for each application.
- We hope AOP will be easy to design and develop agents that behave in the intended manner, robustly and predictably, are easily modified, extended and controlled, and whose design can be re-used..

[BeliefDesireIntention | BeliefDesireIntention | IntentAgents]

(last edited October 4, 1999)
Find Page by browsing or searching

IntentionalAgents

Agents as Intentional Systems (M. Woolridge & N. Jennings, 1995)

Folk Psychology Statements

- "Janine took her umbrella because she believed it was going to rain."
- "Michael worked hard because he wanted to posses a PhD."

Intentional Notions & Systems (Dennett, 1987)

- A system whose behaviour can be predicted by the method of attributing belief, desires, and rational acumen.
- All bodies can be described by intentional stance, but it doesn't make sense for those systems we know sufficiently enough about that can be modeled in a simple manner. The more we know about a system the less we have to rely on intentional behaviour. However some complex systems cannot easily resolve a diagnostic query.

References

Dennett, D. C., 1987. *The intentional stance*, MIT Press.

[BdiIntroduction | BeliefDesireIntention | IntentionalAttitudes]

(last edited October 4, 1999)
Find Page by browsing or searching

IntentionalAttitudes

- Information Attitudes - Relate to information about the world the agent resides in.
- Pro-attitudes - Guide the agent's actions.

[IntentionalAgents | BeliefDesireIntention | IntentionalNotions]

(last edited October 31, 2000)
Find Page by browsing or searching

IntentionalNotions

Propositional Logic Limitations (Truth Functional)

- Bel(Janine, Father (Zeus, Cronos))
- (Zeus = Jupiter)
- Bel(Janine, Father (Jupiter, Cronos))
- Janine believing p is not dependent on the truth of p.

Syntatic Component

- Modal Language Formalism
- Meta-Language Formalism

Semantic Component

- Possible Worlds Semantics (More Common)
- Interpreted Symbolic Structures (Use of Belief Data Structures)

[IntentionalAttitudes | BeliefDesireIntention | PossibleWorlds]

(last edited October 31, 2000)
Find Page by browsing or searching

PossibleWorlds

Background

- Originally proposed by Hintikka in 1962.
- Now is commonly modeled using modal logic developed by Kripke in 1963.

Possible Worlds Example

An agent is playing poker. Complete knowledge of the opponents hand is impossible to determine. The ability to play is determined partially by the agent's belief in the opponents hand. Suppose agent has Ace of Spades. First compute all the possible ways the cards could be dealt to the opponent. These are possible worlds. Then eliminate those worlds that are not possible given what the agent knows. What is left over are the epistemic alternatives (worlds possbile given ones beliefs). Something TRUE in all worlds is said to be believed by the agent (It is TRUE that the agent has the Ace of Spades).

Normal Modal Logic for Possible Worlds

- Classical propositional logic extended by 2 operators. "" (necessarily) and "" (possibly).
- The semantics of modal connectives define what worlds are considered accessible from other worlds.
- The formulae is then TRUE if is TRUE for all worlds accessible from the current world.
- The formulae is TRUE if is TRUE for in at least one world accessible from the current world.
- The two modal operators are duals of each other:

[IntentionalNotions | BeliefDesireIntention | NormalModalLogic]

(last edited October 31, 2000)
Find Page by browsing or searching

NormalModalLogic

The K Axiom (Kripke)

- An agent's knowledge is closed under implication.

Necessitation Rule

- If is VALID then is VALID.
- An agent knows all valid formulaes.

Reflexive Accessibility Relation (T) or Knowledge Axiom

- What is known is TRUE.

Serial Accessibility Relation (D)

- Agent's beliefs are non-contradictory.

Transitive Accessibility Relation (4) or Positive Introspection Axiom

- Examining ones beliefs.

Euclidean Accessibility Relation (5) or Negative Introspection Axiom

- Agent is aware of what it doesn't know.

KTD45 is idealised knowledge and KD45 as idealised belief.

[PossibleWorlds | BeliefDesireIntention | TheoryOfIntentions]

(last edited October 31, 2000)
Find Page by browsing or searching

TheoryOfIntentions

Theory of Intentions (Cohen & Levesque, 1990a)

Seven properties to be Satisfied

1. Intentions pose problems for agents, who need to determine ways to achieving them.
2. Intentions provide a "filter" for adopting other intentions, which must not conflict.
3. Agents track the success of their intentions, and are inclined to try again if their attempts fail.
4. Agents believe their intentions are possible.
5. Agents do not believe they will not bring their intentions.
6. Under certain circumstances, agents believe they will bring about their intentions.
7. Agents need not intend all the expected side effects of their intentions.

[NormalModalLogic | BeliefDesireIntention | PossibleWorldsExample]

(last edited October 4, 1999)
Find Page by browsing or searching

PossibleWorldsExample

Events

```
d1 - go to the dentist 1
d2- go to dentist 2
b - go shopping
```

Facts

```
p - pain
f - tooth filled
```

[TheoryOfIntentions | BeliefDesireIntention | BdimToolkit]

(last edited October 4, 1999)
Find Page by browsing or searching

BdimAgentToolkit

The BDIM Agent Toolkit Design (Busetta & Ramamohanarao, 1997)

Objective

- A collection of JAVA classes for the development of agents based on the BDI model.
- Components are simple and independent of specific architecture.

The Belief, Desire, Intention, Message Model

BDIM Components

- Mental - Beliefs, Goals, Messages, Events, and Intentions.
- Capabilities - Plan and Intention Instantiator.
- Excecution - Event Intention Manager and Plan Interpreters.

References

Busetta, P. and Ramamohanarao, K., 1998, "An architecture for mobile BDI agents.", "In Proc. of the 1998 ACM Symposium on Applied Computing (SAC'98)", pp. 445-452.

Busetta, P. and Ramamohanarao, K., 1998, "The BDIM agent toolkit design.", *The Department of Computer Science, University of Melbourne, Technical Report 97/15*, Melbourn, Australia.

Links to other Sources

- BDIM Toolkit [1]

Other BDI Implementations

- JAM [3]
- LALO [4]

[PossibleWorldsExample | BeliefDesireIntention | BdiAgentIssues]

BdiAgentIssues

Shortcomings

- Lacks paradigm for concurrency control among intentions performing conflicting operations.
- Are missing an exception handling mechanism.
- There are not conducive to mobility.

Theory => Practicality

- What does modal logic have to do with the procedural implementations of BDIM based on data structures?

[BdimAgentToolkit | BeliefDesireIntention | End]

(last edited October 31, 2000)
Find Page by browsing or searching

BdiArchitecture

(source: Weiss book, chapt. 1)

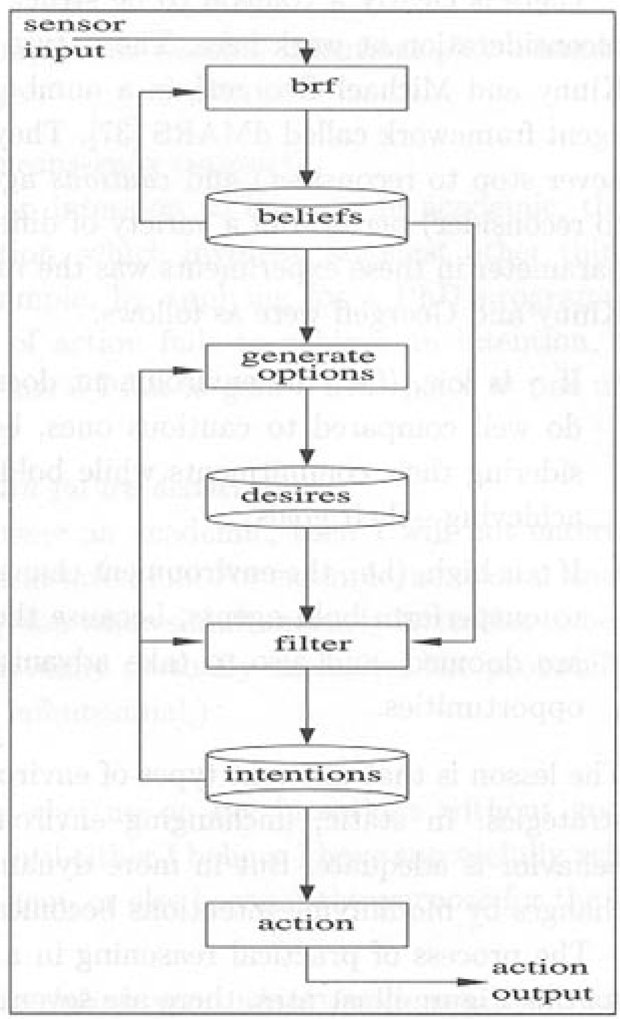


Figure 1.5 Schematic diagram of a generic belief-desire-intention architecture.

- function *action*(perceptions): action
- begin
  - B := *brf*(B, perceptions)
  - D := *options*(D, I)
  - I := *filter*(B, D, I)
  - return *execute*(I)
- end function *action*

note:

- need for belief revision
- form of planning

known BDI architectures, frameworks and languages:

- Agent0 (Shoham)
- PRS, dMARS, Agent Speak (Georgeff, Rao)
- LALO
- Jason (http://jason.sourceforge.net), a Java implementation of AgentSpeak?

(currently edited by 65.39.186.83?)
Find Page by browsing or searching

Agents with
Beliefs Desires & Intentions

Andre Vellino
Cognitive Sciences
Carleton University

September 24, 2005

Overview

- Motivation for BDI
- Logical models for BDI
- BDI Agent Implementations

Folk Psychology of BDI

- Actions, choices and decisions in human beings can be “explained” in folk psychology with BDI
  - i.e. They are made based on a mental representation of the world (beliefs) and goals to be achieved (desires) and rational deliberations and commitments for achieving them (intentions)

Desiderata for Communicating Distributed Agents

- Make autonomous decisions;
- React to a changing environment;
- Collaborate with other agents towards a common goal;
- Reason about (problem solve) attaining its own <u>and</u> other Agents' objectives;
- Act rationally;
- Account for choices and actions.

BDI Model for Agents: Assumptions

- There exists a “mental representation language” which can express BDIs and the way the world is;
- The means for achieving goals (actions, behaviours) can be deduced from a logic and an ATP;
- BDI agent architectures can best meet desiderata of distributed, collaborative multi-agents;

The Planning Problem

Given

- Initial conditions (I)
- Goal (G)

Find

- Sequences of intermediate states (S) to achieve (G)

i.e. design a theorem prover for deducing (G) from (I). The proof is (S)

In this model Computation <u>is</u> Deduction

Knowledge + Question = Answer

Some Properties of BDI Model

- Intentions (Plans of Action) must be believed to be achievable, given what the agent knows
- Intentions and Beliefs must be compatible with actions

(more later ….)

Expressing & Reasoning w/ BDIs

- Logic for changing beliefs and plans (Temporal and Modal Logic)
  - Agents need to reason about its own and other agents BDIs (Modal Logic) and do so over time (Temporal Logic).
- Decision procedure
  - Needs to be effective and resource-bounded

What's in a Logic?

- Syntax
  - Rules for constructing WFFs
    - e.g. $(p \rightarrow (q \rightarrow p))$
- Model Theory
  - "Interpretations" for satisfying WFFs
    - e.g. {T,F} assignments to boolean formulas
- Proof Theory
  - Mechanisms for inference
    - Axiom-systems, Natural Deduction, Resolution, Tableaux

Varieties of “Modal Logics”

- Modal Logic
  - $\Box$ It is necessary that…
  - $\Diamond$ It is possible that…

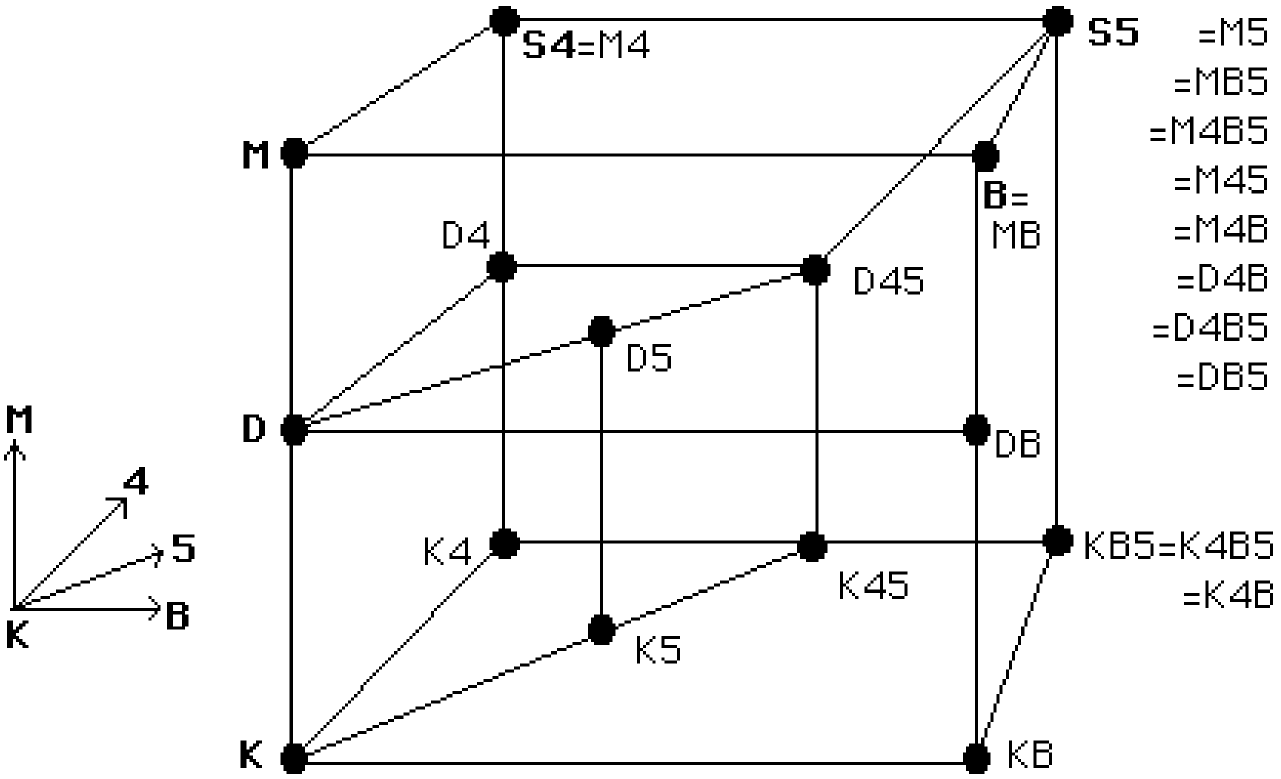
  $\Box A = \sim \Diamond \sim A$
- Deontic Logic
  - O It is obligatory that ..
  - P It is permitted that .. P(A) = ~O(A)
  - F It is forbidden that .. F(A) = O~(A)
- Temporal (Tense) Logic
  - G It will always be the case that ..
  - F It will be the case that ..
  - H It has always been the case that ..
  - P It was the case that..
- Doxastic (Epistemic) Logic
  - Bx x believes that ..
  - Kx x knows that ...

“Classical” Modal Logic

- Symbols / axioms of 2-valued propositional calculus plus… $\Diamond$ + $\Box$ +
- Axioms for system K, M, S4, S5, B
  - K) $\Box(P \rightarrow Q) \rightarrow (\Box P \rightarrow \Box Q)$ (K - Kripke)
  - M) $\Box P \rightarrow P$ (M [or T] - Modal)
  - 4) $\Box P \rightarrow \Box \Box P$ (S4 = M + 4)
  - 5) $\Diamond P \rightarrow \Box \Diamond P$ (S5 = M + 5)
  - B) $P \rightarrow \Box \Diamond P$ (B - Brouwer)

    - S5 = S4 + B; 5 is equivalent to $\Diamond\Box P \rightarrow P$
    - In deontic logic, replace M by axiom (D) : $O(A) \rightarrow P(A)$, i.e. $\Box P \rightarrow \Diamond P$

Map of Modal Logics



Questions

- Can you interpret □ as "it ought to be"?
- If you interpret □ as "knows" do you think it is true that

$$\frac{\Box P \quad P \rightarrow Q}{\Box P}$$

  or even $\Box P \rightarrow \Box\Box P$?
- Exercise: devise a logic for modal operator:

♥

George Dubya Bush - Modal Logician

4) $\square\ P \rightarrow \square\ \square\ P$

"*I know what I believe. I will continue to articulate what I believe and what I believe - I believe what I believe is right*"

G. W. Bush, Rome, July 22, 2001

“Possible World” Semantics for Modal Logic

A sentence A is true in a possible world α in a model $\mathcal{M} = <\mathcal{W},\mathcal{P}>$

$$\models^{\mathcal{M}}_{\alpha} A$$

Where $\mathcal{P} = \{\mathcal{P}_1, \mathcal{P}_2, \ldots \mathcal{P}_n\}$ is a sequence of subsets of possible worlds in $\mathcal{W}$ such that the $\mathcal{P}_1$ is the set of worlds at which the atomic formula P_1 is true, $\mathcal{P}_2$ is the set of worlds at which the atomic formula P_2 is true, etc…

Some Of The Truth Conditions

$\models^{\mathcal{M}}_{\alpha} P_n$ iff $\alpha \in \mathcal{P}_n$

$\models^{\mathcal{M}}_{\alpha} A \,\&\, B$ iff $\models^{\mathcal{M}}_{\alpha} A$ and $\models^{\mathcal{M}}_{\alpha} B$

.....

$\models^{\mathcal{M}}_{\alpha} \Box A$ iff for every $\beta \in \mathcal{M}$ $\models^{\mathcal{M}}_{\beta} A$

$\models^{\mathcal{M}}_{\alpha} \Diamond A$ iff for some $\beta \in \mathcal{M}$ $\models^{\mathcal{M}}_{\beta} A$

Note: Referential Opacity of Modal Operators

$\mathcal{O}(P)$ & $P = Q$ does not imply $\mathcal{O}(Q)$

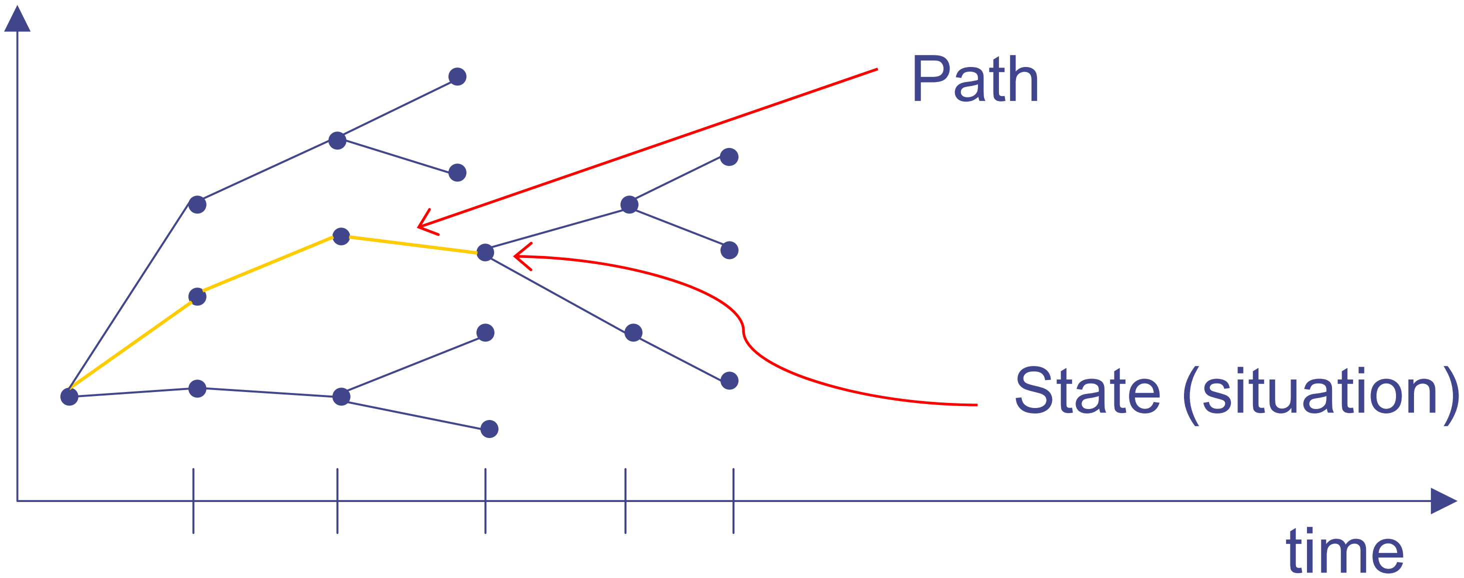
e.g.
□(9>3)
~□(Number of Planets > 3)

BEL(wrote('Mark Twain', 'Huck Fin'))
~BEL(wrote('Samuel Clemens', 'Huck Fin'))

Standard Models (Kripke)

- Add "Accessibility Relation" to possible world model $\mathcal{M} = <\mathcal{W}, \mathcal{R}, \mathcal{P}>$ where $\mathcal{R}$ is a binary relation on $\mathcal{W}$.
- The meaning of $\mathcal{R}$ is "relative possibility", "relevance" or "accessibility"
- $\alpha \mathcal{R} \beta$ is interpreted as β is accessible from α
- $\Box(\Diamond)$ P is true at world α iff P is true at every (some) world β that is $\mathcal{R}$-accessible from α
- Needed for temporal modalities

Temporal Logic (Computation Tree Logic CTL*)



Branching Time (model concurrent distributed systems)

- State formulas
- Path formulas
- CTL modalities + "Optional" and "Inevitable"
- Discrete time

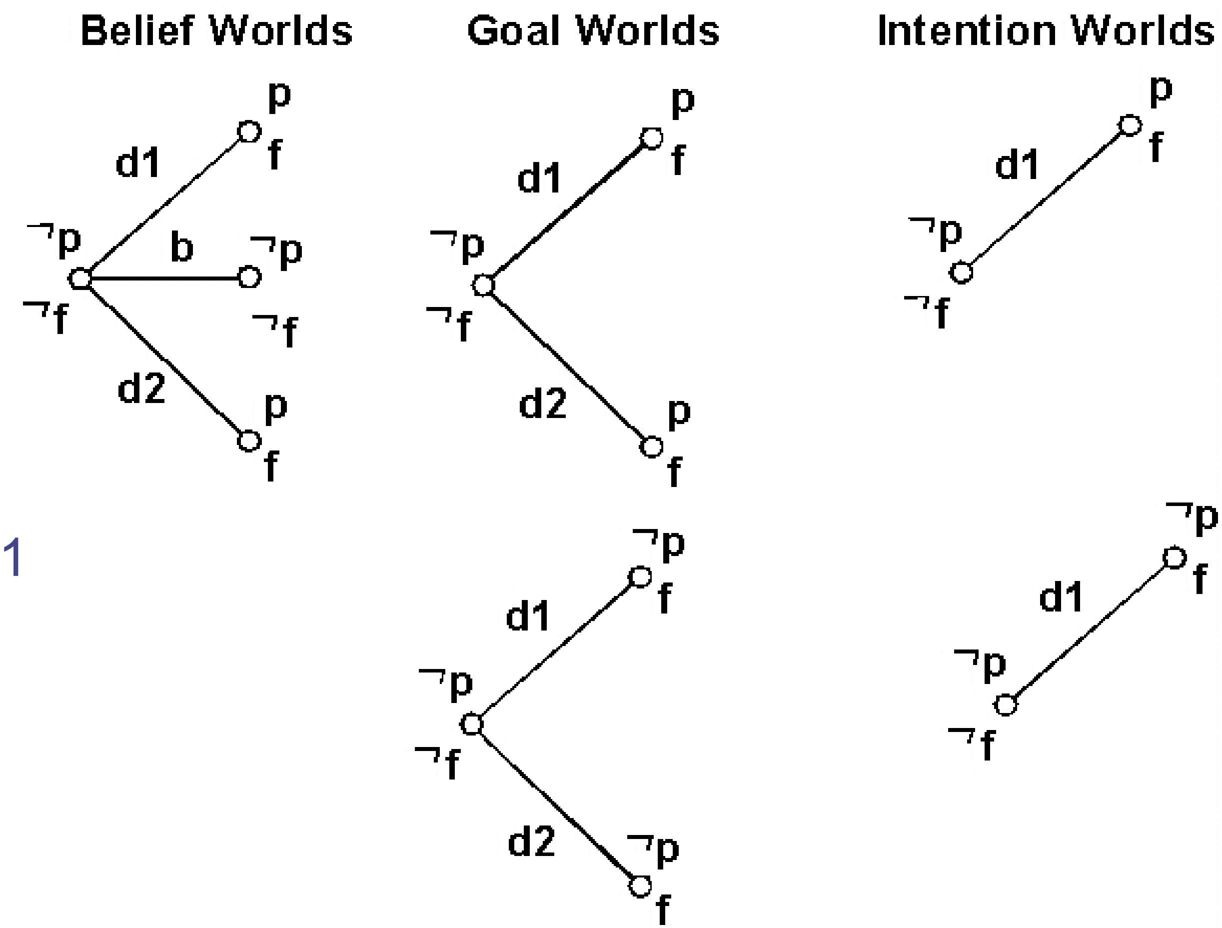
CTL* Modalities

- 0 “at the next moment in time” - Next
- ◊ “at some future point” - Eventually
- □ “always in the future” - Always
- **U** Until
- Optional - “on some future path”
- Inevitable - “on all future paths”

BDI Characteristics in CTL*

- Agents must believe they can optionally achieve their goals
  - i.e. for each belief-accessible world there is a goal-accessible world
- However, inevitabilities need not be goals or intentions
  - Inevitable(filling $\rightarrow$ pain) & GOAL(filling) & ~GOAL(pain)

Example



Events

d1 - go to the dentist 1

d2- go to dentist 2

b - go shopping

Facts

p - pain

f - tooth filled

BEL(INTEND(f) → inevitable(◊p)), INTEND(f) & ~INTEND(p)

BDI Axioms

- $GOAL(\psi) \rightarrow BEL(\psi)$
  - goals are believed.
- $INTEND(\psi) \rightarrow GOAL(\psi)$
  - intentions are goals.
- $INTEND(\psi) \rightarrow BEL(INTEND(\psi))$
- $GOAL(\psi) \rightarrow BEL(GOAL(\psi))$
  - intentions (and goals) are believed
- $INTEND(\psi) \rightarrow GOAL(INTEND(\psi))$
  - intentions must be goals
- $INTEND(\psi) \rightarrow$ inevitable $(\Diamond \sim INTEND(\psi))$
  - don’t defer indefinitely (i.e. do something)

Commitment Strategies (1)

<u>Blind Commitment</u>

- $INTEND(inevitable \lozenge \psi) \rightarrow inevitable(INTEND(inevitable (\lozenge \psi)) \, \mathbf{U} \, BEL(\psi)$

  If ψ is an action-statement and if an agent intends that inevitably ψ will eventually be true then the agent will inevitably maintain her intentions (for ψ) until she believes ψ.

Commitment Strategies (2)

<u>Single-Minded Commitment</u>

- INTEND(inevitable $\Diamond\ \psi$) $\rightarrow$ inevitable(INTEND(inevitable ($\Diamond\ \psi$)) **U** (BEL(ψ) $\vee$ ~BEL(optional $\Diamond\ \psi$))

  An agent maintains her intentions as long as she believes that they are still options.

Commitment Strategies (3)

Open-Minded Commitment

- INTEND(inevitable $\Diamond\ \psi$) $\rightarrow$ inevitable(INTEND(inevitable ($\Diamond\ \psi$)) **U** (BEL(ψ) $\lor$ ~GOAL(optional $\Diamond\ \psi$))

  An agent maintains her intentions as long as those intentions are still her goals.

Exercise:

Express wedding vows as a commitment strategy in a BDI logic!

Complexity Problem

- Semantics and Proof-theory for Modal Logics is complex
- Automated Theorem Provers (planners) run afoul of feasibility problem
- Two response:
  - Simplify your logic to make proofs feasible
  - Limit what you can conclude

Satisfiability / Unsatisfiability

a set of clauses $\Sigma = \{C_1, C_2, ...C_n\}$ is <u>*satisfiable*</u> if $\exists$ an assignment of truth values to literals in Σ such that

C_1 & C_2 & ...&C_n is *true*

SAT

NP-complete

a set of clauses $\Sigma = \{C_1, C_2, ...C_n\}$ is <u>*unsatisfiable*</u> if <u>no</u> assignments of truth values to literals in Σ are such that

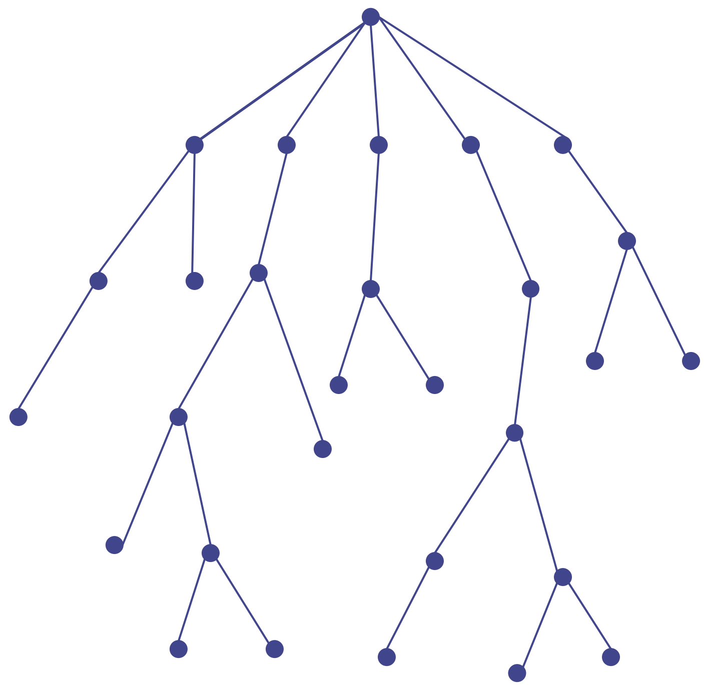
C_1 & C_2 & ...&C_n is *true*

co-SAT

co-NP-complete

Search-Space vs. Proof Length

- For problems in NP (SAT), the search space is exponentially large but the proof is polynomial
- For problems in co-NP (co-SAT), the minimal length proof is exponential and the search space even larger



Other Complexity Classes

- PSPACE-complete
  - Class of problems that can be solved by a polynomial-space bounded, Deterministic Turing Machine (DTM)
  - All NP-complete problems can be solved in PSPACE but is PSPACE = PTIME ? PSPACE not likely to be in NP
- EXPTIME
  - Class of problems with complexity bounded by $2^{p(n)}$ for some polynomial *p* of input length *n*

Complexity of Modal Logic

- S5: (co)SAT is (co)NP-complete
- T,K4, S4: SAT is PSPACE-complete
- K: SAT is EXPTIME-complete

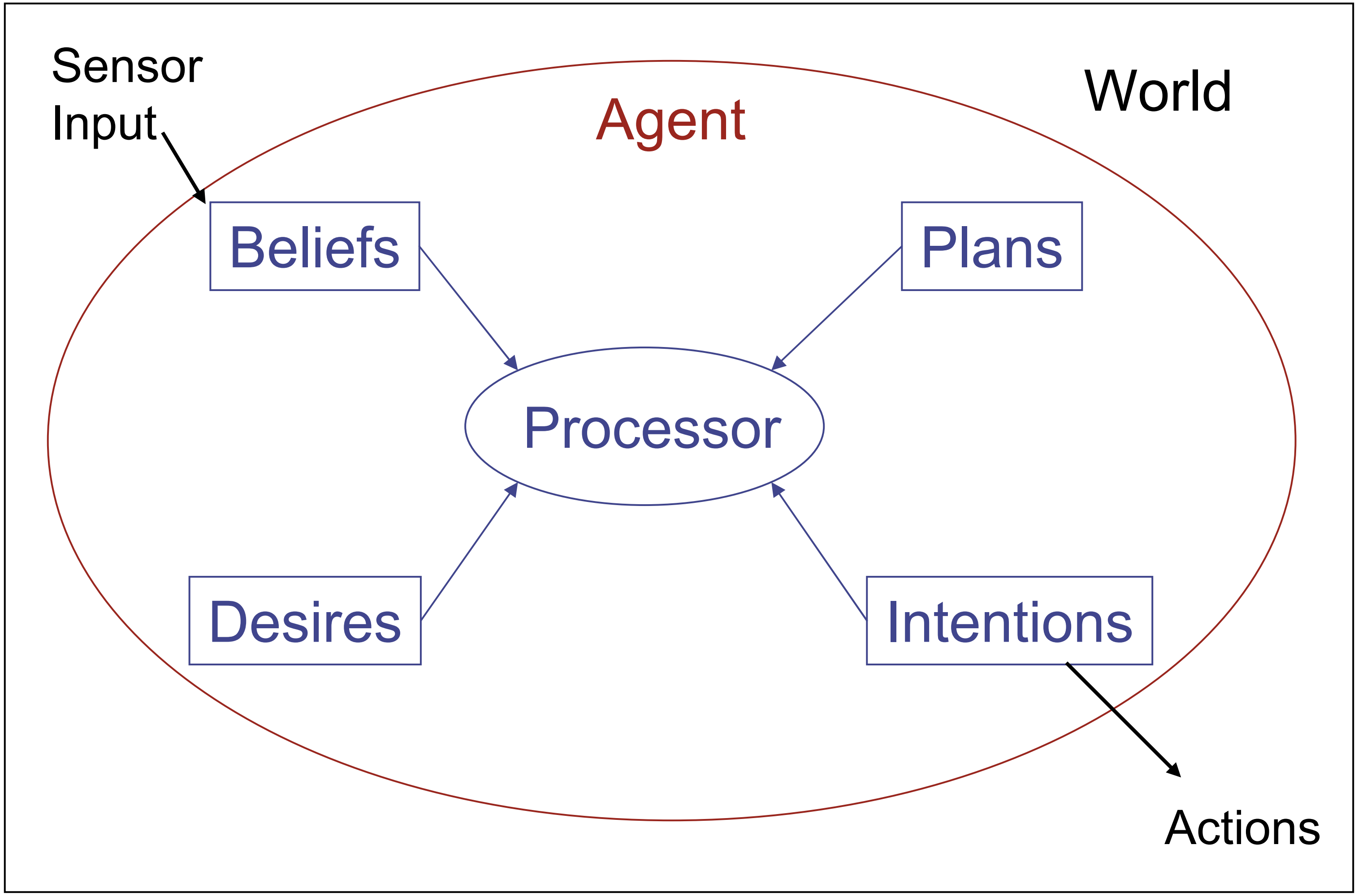
  (see Marx '97)

K) $\Box(P \rightarrow Q) \rightarrow (\Box P \rightarrow \Box Q)$ (K - Kripke)
M) $\Box P \rightarrow P$ (M [or T] - Modal)
4) $\Box P \rightarrow \Box \Box P$ (S4 = M + 4)
5) $\Diamond P \rightarrow \Box \Diamond P$ (S5 = M + 5)

Dealing with Complexity

- Simplify your logic to make proofs feasible
- Limit what you can conclude
  - In PRS:
    - Only represent beliefs about current state of the world
    - Consider only ground terms (no variables)
    - No disjunctions or implications
    - Plans are obtained from plan-libraries that represent accessible future states
    - Plans are treated implicitly on the “goal stack”

Procedural Reasoning System



Georgeff & Lansky '87

PRS Interpreter

initialize,

repeat,

  generate-options(event-queue,options),
  select-options(options,selected-options),
  update-intentions(selected-options),
  execute,
  get-new-external-events,
  drop-successful-intentions,
  drop-impossible-intentions,

end repeat

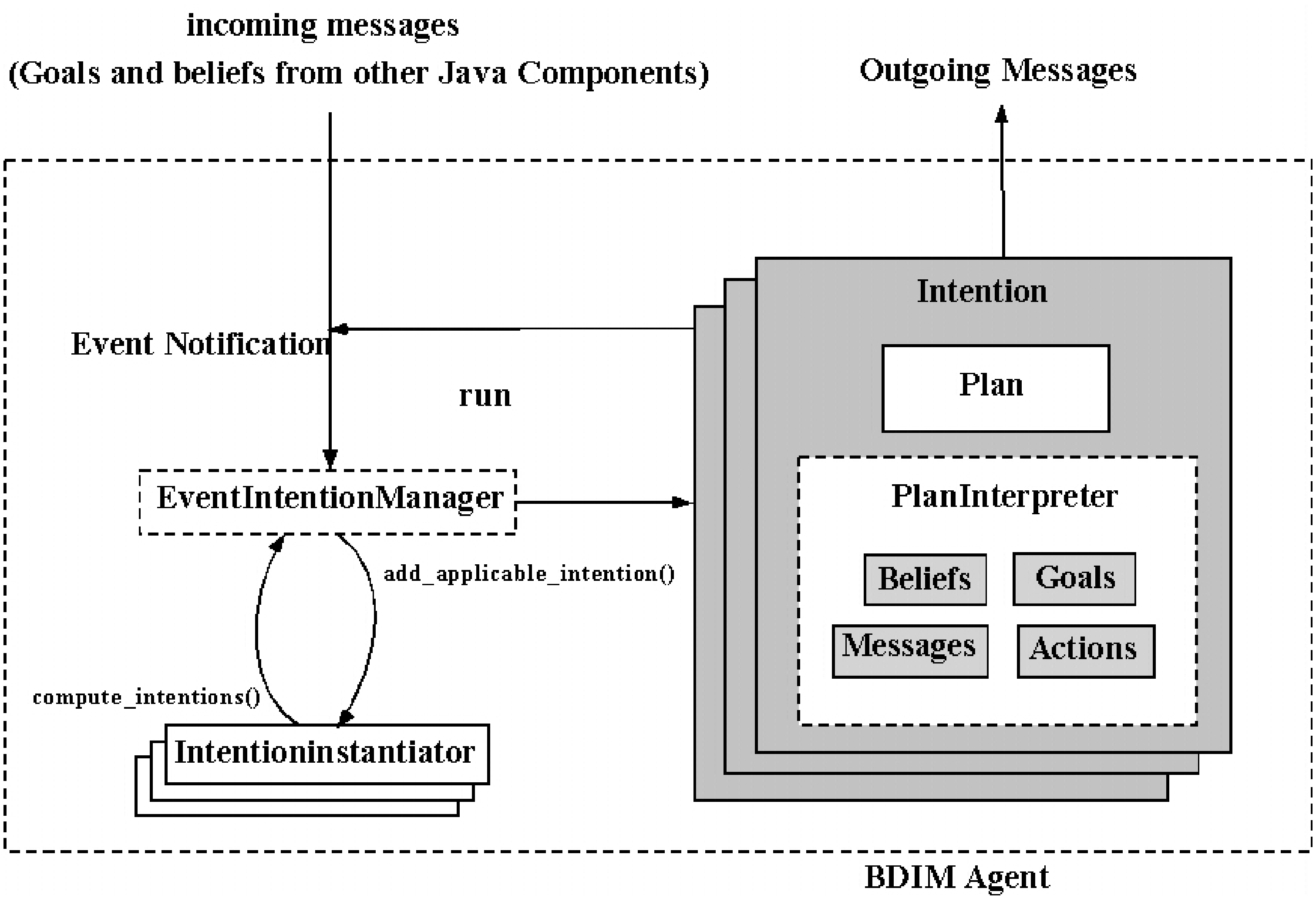
dMARS (Distributed Multi-Agent Reasoning System)

- Based on PRS
- Paired-down version (PRS-lite) used in space shuttle Reaction Control System
  (diagnosis of malfunction and automatic system reconfiguration)
- No first-principles planning
- Only ground formulae and negations

BDIM & TOMAS

- BDI + Messages Toolkit
- Adds concurrency control in BDI
- Addresses problems of multiple agents attempting to collaboratively achieve the same goal
- Potentially useful for mobile BDI agents
- **T**ransaction **O**riented **M**ulti-**A**gent **S**ystem
  - Concurrent BDIMs for teams of BDI agents

BDIM Agent Architecture



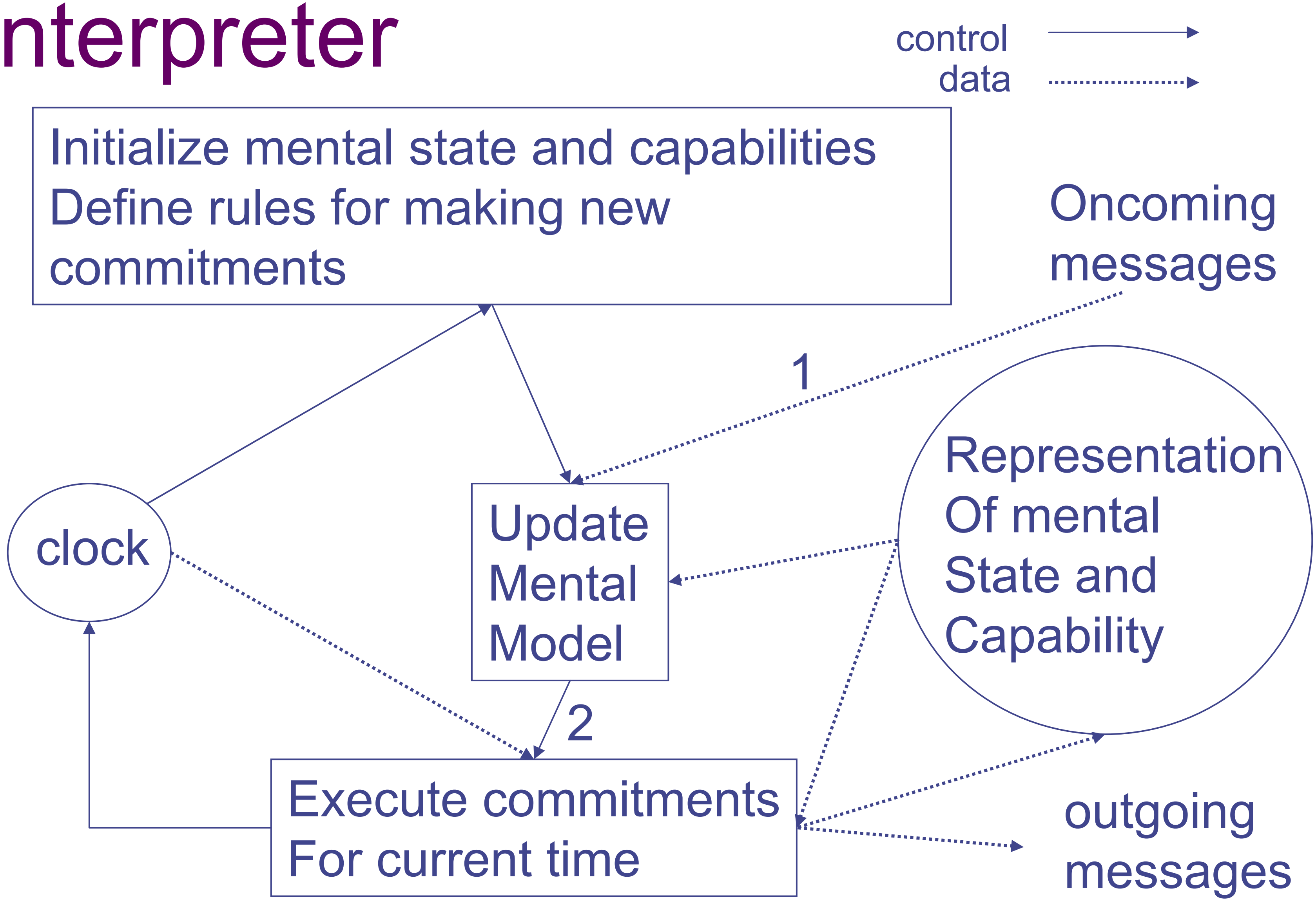
Agent0 (Shoham '93)

- Time-indexed-states (facts) - $p(a,b)^t$
- Action - states w/ effects - $q(a,b)^t$
- Belief - "mental state" modality - $B^t_a(\psi)$
- Obligation - 2-ary deontic modality - $OBL^t_{ab}(\psi)$
- Choice - Self-obligation - $DEC^t_a(\psi)=OBL^t_{aa}(\psi)$
- Capability - $CAN^t_a(\psi)$, ψ may be time-indexed as well $ABLE^t_a(\psi)$

Agent0 Properties

- Consistency (between intentions, between intentions and beliefs, between beliefs...)
- Good faith: only commit to what you believe you are capable of
- Introspection
- Persistence
    - beliefs (obligations, capabilities) persist by default, and their absence as well, until the belief is learned
- Complexity is dealt with by disallowing connectives other than ~ and disallowing nested modal operators.

Flow Diagram for Agent0 Interpreter



Conclusions

- Theory of BDI is conceptually rich, well-developed and provides fertile ground for AI research
- Successful BDI implementations in reactive systems don't take full advantage of theory (for practical reasons)
- Jury is still out on whether BDI model is better than "representation-free" rational agents

References

- P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42:213–261, 1990.
- A. S. Rao and M. Georgeff. BDI Agents: from theory to practice. In *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS-95)*, pages 312–319, San Francisco, CA, June 1995.
- A. S. Rao and M. P. Georgeff. Modeling rational agents within a BDI-architecture. In R. Fikes and E. Sandewall, editors, *Proceedings of Knowledge Representation and Reasoning(KR&R-91)*, pages 473–484. Morgan Kaufmann Publishers: San Mateo, CA, April 1991.
- Busetta, P. and Ramamohanarao, K., 1998, "An architecture for mobile BDI agents.", In Proc. of the 1998 ACM Symposium on Applied Computing (SAC'98)", pp. 445-452.

References (cont.)

- d'Inverno, M., Kinny, D., Luck, M., Wooldridge, M. “A Formal Specification of dMARS” pages 155-176 of: *Intelligent Agents IV: Proceedings of the Fourth International Workshop on Agent Theories, Architectures and Languages* Springer-Verlag. 1365, 1998.
- Shoham, Y., 1993. "Agent oriented programming", *Artificial Intelligence*, **60(1)**, pp. 51-92.
- Hughes and Cresswell *A New Introduction to Modal Logic* Routledge1996
- M. Marx. “Complexity of modal logics of relations:. Technical Report ML-97-02, Institute for Logic, Language and Computation, University of Amsterdam, May 1997.