
Overview of Cohen & Levesque
Intention is Choice with Commitment

Bratman : Intentions play three *functional* roles for resource-bounded agents

1. Intentions pose problems for the agent--the agent needs to determine a way to achieve them
=> intentions *direct future processing*
2. Once intention A is adopted, intentions inconsistent with it will not be adopted
=>Intentions provide a *screen of admissibility* for adopting other intentions
3. Agents track the success of their attempts to achieve intentions
=> the *environment is monitored* with respect to intention persistence; what information is relevant

Further desired properties of intentions

If *agent adopts intention x*

- agent should believe x is possible
- agent does not believe it will not bring about x
- agent believes it will bring about x
- agent doesn't intend all the side effects of intentions
- agent can't give up too soon

Example: going to the dentist
You intend to get tooth filled. You *expect* to feel pain and anxiety.
You do not *intend* the side effects of pain, anxiety, etc.

C&L :Possible world semantics

The world is defined as a discrete sequence of events,
temporally extended into past and future

- 1 (HAPPENS a)
Action expression *a will happen next*
- 1 (DONE a)
Action expression *just happened*

C&L : action (continued)

More notation on actions

- 1 "a" and "b" denote actions sequences; "e" denotes primitive event
- 1 a;b is action composition (action a, then action b)
- 1 a | b is non-deterministic choice
- 1 a || b is concurrent occurrence of a and b
- 1 p? is a test
 - p?;a "when p is true, action a occurs next"
 - a;p? "action a occurs, after which p holds"
 - p?; a ; q? could be viewed as a specification of pre and post conditions on action a

(AGT x a)

- 1 x is (the only) agent for an action sequence

C&L define these useful abbreviations

- (DONE x a)
abbreviation for (DONE a) & (AGT x a)
- (HAPPENS x a)
abbreviation for (HAPPENS a) & (AGT x a)
agent x is the one that has made a done or that is going to make a happen
- (AFTER a p) abbreviation for (HAPPENS a;p?)

Temporal modalities...

$\Box p$ abbreviation for $\exists e$ (HAPPENS e; p?)

- 1 p is true *eventually*
- 1 ... that there is some event, such that when that event happens, the next thing is that p is true
- 1 only events can cause the truth value of something to change

(LATER p) = $\neg p \ \& \ \Box p$

- 1 p is not true now but eventually it will be true

C&L rational agency: More abbreviations

$\Box p$ abbreviation for $\sim (\Diamond (\neg p))$ *always*

- 1 "it is false that eventually $\neg p$ can become true in the future"
- 1 "p is always true given any course of events"
- 1 "p is necessarily so in the future"

(PRIOR p q) abbreviation for

- 1 $\forall c$ (HAPPENS c;q?) $\Rightarrow \exists a$ (a <= c) & (HAPPENS a;p?)
- 1 for all action sequences c, when that sequence happens and q holds afterwards, then there is some sub-action sequence of c called "a" and after "a" happens, p becomes true
- 1 "p will become true no later than q"

BEL(x p)

pg. 231 Proposition 3.15:

Weak S5 system

- 1 if agent believes p \rightarrow q, then if agent believes p, agent believes q
- 1 if agent believes p, then it is false that the agent believes $\neg p$
- 1 if agent believes p, then agent believes it believes p
- 1 if the agent does not believe something, then agent believes it does not believe it

Knowledge is "true belief" 3.18

- 1 (KNOW x p) \Leftrightarrow p & (BEL x p)

On choice and GOAL(x p)

Goals

- 1 "agents chose worlds they would like (most) to be in"
- 1 "the [worlds] in which their goals are true"
- 1 "BEL and GOAL characterize what is implicit in an agent's beliefs and goals (chosen desires) rather than what he actively or explicitly believes or what he has a goal."

(GOAL x p)

- 1 p follows from the agent's goals
- 1 agent is choosing a world in which p is true

agents choose entire worlds

- 1 the trick is to get them not to intend everything in those worlds...

semantics of GOAL

goals must be consistent (D axiom)

- 1 prop 3.21 (GOAL x p) \rightarrow \neg GOAL (x \neg p)

if an agent chooses a world in which p is true and also chooses a world in which q has consequence, it chooses a world in which that consequent is true... (K axiom)

- 1 prop 3.22 (GOAL x p) & (GOAL x (p \rightarrow q)) \rightarrow (GOAL x q)

Goals and beliefs: Realism

prop 3.26 (BEL x p) \Rightarrow (GOAL x p)

- 1 BEL and GOAL are evaluated at the same point in time
- 1 if agent believes now "light is off", he cannot choose a world now in which "light is not off" at the same time

sometimes justified as accepting the inevitable

- 1 if I believe that (inevitably) the sun will rise tomorrow, I cannot choose a world in which the sun does not rise tomorrow.

a weaker form of realism restates this directly with appeal to temporal considerations

- 1 BEL (x p) \Rightarrow \neg GOAL(x \neg p)
(BEL x q) \Rightarrow \neg (GOAL x \neg q)
(BEL x q) \Rightarrow \neg (GOAL x \blacklozenge \neg q)

Goals and chosen worlds (2)

p. 235, proposition 3.28

$(GOALx p) \ \& \ (BEL \ x \ (p \rightarrow q)) \ \rightarrow \ (GOAL \ x \ q)$

if agent-x chooses worlds in which p holds and agent-x believes that p implies q, then agent-x chooses worlds in which q is true

- 1 note that this is Bratman's "honest" rationality. The choice is coming from a deliberation of all consequences of actions.
- 1 but we see that this is not the same as *intending q*...

On side effects of goals

proposition 3.28. The agent *expects* side-effects of goals...

1 $GOAL \ (x \ p) \ \& \ BEL \ (x \ (p \rightarrow q)) \ \rightarrow \ GOAL \ (x \ q)$

- 1 if p follows in a world chosen by x and x believes p → q,
- 1 then q follows in a world chosen by x

Achievement Goals

Achievement goals

1 $(A-GOAL \ x \ p) \ =_{def} \ (BEL \ x \ \neg p) \ \& \ (GOAL \ x \ (LATER \ p))$

- 1 x believes that p is false [in the current world] and chooses worlds in which p is eventually true

Following this definition, they present the assumption that agents do not persist forever on a goal, and eventually drop it

1 $\models \ \blacklozenge \ \neg \ (GOAL \ x \ (LATER \ p))$

Strong Persistent Goal : Definition 4.1
a kind of achievement goal....

(P-GOAL x p) =_{def}
 (BEL x ~p) &
 (GOAL x (LATER p)) &
 [BEFORE [(BEL x p) or (BEL x ~p)]
 ~(GOAL x (LATER p))]

1. agent x believes p is currently false
2. and chooses worlds in which it will be true later
3. And before dropping the goal,
 agent-x must either believe that p is true or believe that p can never be true

notion of commitment....

Relativized Persistent Goal (Definition 8.1)

(P-R-GOAL x p q) =_{def}
 (BEL x ~p) &
 (GOAL x (LATER p)) &
 [BEFORE [(BEL x p) or (BEL x ~p) or (BEL x ~q)]
 ~(GOAL x (LATER p))]

q is the "escape" clause...
 background conditions

persistent goals & side effects

(GOAL x p) & (BEL x (p -> q)) -> (GOAL x q)
 requires only that agents choose the consequences that follow in their goal worlds

but these consequences are not persistent goals (in most cases)

(P-goal x p) & (BEL x (p -> q)) $\not\rightarrow$ (P-goal x q)
 (P-goal x p) & (BEL x (p -> q)) $\not\rightarrow$ (P-goal x q)

problem case

(P-goal x p) & (BEL x (p -> q)) -> (P-goal x q)
sometimes..

Persistent goals constrain future beliefs

$(P\text{-GOAL } x \ p \ q) \Rightarrow \diamond [BEL \ x \ p \text{ or } (BEL \ x \ \neg p) \text{ or } (BEL \ x \ \neg q)]$

Persistent goals and (future) beliefs must be consistent

↓

(All goals are eventually given up for one of 3 reasons)

$(P\text{-GOAL } x \ p \ q) =_{def}$
 $(BEL \ x \ \neg p) \ \&$
 $(GOAL \ x \ (LATER \ p)) \ \&$
 $[BEFORE \ [(BEL \ x \ p) \ \text{ or } (BEL \ x \ \neg p) \ \text{ or } (BEL \ x \ \neg q)]$
 $\quad \neg(GOAL \ x \ (LATER \ p))]$

Intending to take an action is a kind of persistent goal

$(INTEND \ x \ a \ q) =_{def}$
 $(P\text{-GOAL } x \ ((DONE \ x \ (BEL \ x \ (HAPPENS \ a))?) : a) \ q)$

x has the persistent goal of reaching a state at which it has just occurred that x believes it will do action a, after which it does in fact do action a.

=> agent is committed to arriving at a state in which it is about to do the intended action next
=> agent intends to pick up a cup =

Agent has a persistent goal to be in a state where it has come to pass, that (a) agent believed agent was about to pick up the cup and (b) agent did actually pick it up

recall P-R-GOAL =
 $(BEL \ x \ \neg p) \ \&$
 $(GOAL \ x \ (LATER \ p)) \ \&$
 $[BEFORE \ [(BEL \ x \ p) \ \text{ or } (BEL \ x \ \neg p) \ \text{ or } (BEL \ x \ \neg q)]$
 $\quad \neg(GOAL \ x \ (LATER \ p))]$

Intend-2 (intending to bring about a state)

$(INTEND_2 \ x \ p) =_{def} (P\text{-GOAL } x \ \exists e \ (DONE \ x \ s?; e ; p?)$

the agent is committed to doing some event e, such that p is true afterwards

Prior to e, s holds

s is defined as
 $(BEL \ x \ \exists e' \ (Happens \ x \ e'; p?)) \ \text{ and } \ \neg GOAL(x \ \sim (Happens \ x \ e'; p?))$

agent believes he is about to do something (event sequence e') to bring about p [this means the agent has a PLAN...]

and

agent does not have as a goal that e' would not bring about p
[agent would not select e' if it was thought to lead to -p]

Relativized persistent goals, plans, and goal stacks....

by allowing an "escape clause" or "background conditions"
this analysis allows for the notion of goals and subgoals

(P-R-GOAL agent-x action-a background-beliefs)

(P-R-GOAL x (buy air line ticket to Calgary) (GOAL x (be in Calgary))

Some observations about C & L formalism

The utility of continued action is not addressed (directly) as a reason to abandon a goal

An agent might not even attempt p, even if he intends it!

- 1 the escape clause is unrestricted--really an escape from doing anything

is there really any causal relationship between agent's action and an action occurring?

- 1 he believes he's about to do the action next, and then the action happens next
- 1 ok if he is the only force in the environment (no other agent, no divine intervention)

focuses on (internal) commitment of single agent

- 1 joint commitment of multiple agents? joint persistent goals?
- 1 communication implications when one agent abandons a joint persistent goal?

Implications for architectures

architecture should distinguish

- 1 plans from intentions
- 1 intentions from beliefs
- 1 allow for intentions to persist until one of three possible conditions are met

success | failure | "escape clause"
implies monitoring for success | failure | "escape clause"

architecture should therefore allow specification of

- 1 conditions under which intention is suspended
- 1 conditions under which intention is resumed
- 1 conditions under which intention is abandoned

Linking semantics to architectures....

Let plan p be a data structure

- 1 precondition (p)
- 1 body(p) ... a graph of steps (executable actions or a subgoal)
- 1 postcondition (q)

Whenever an agent intends the body of a plan [the action sequence], then it must have postcondition (q) as a goal and must have precondition (p) as a belief

If agent has plan(p) and has belief precondition (p) and has goal postcondition(q) and the deliberation function has chosen this plan over competing plans, then the agent intends the actions that are the body of the plan.

- 1 if body of plan involves non-primitive actions (sub steps, sub goals) then agent intends appropriate plans to achieve those sub-steps
