Refinement of intentions

Andreas Herzig[1], Laurent Perrussel[1], Zhanhao Xiao[1,2], Dongmo Zhang[2]

[1] IRIT, University of Toulouse, France
[2] AIRG, Western Sydney University, Australia

Abstract. According to Bratman, future-directed intentions are high-level plans. We view such plans as high-level actions that can typically not be executed directly: they have to be progressively refined until executable basic actions are obtained. Higher- and lower-level actions are linked by the means-end relation, alias instrumentality relation. In this paper we extend Shoham's database perspective of Bratman's theory by the notions of refinement and instrumentality.

1 Introduction

Bratman highlighted the fundamental role of an agent's future-directed intentions: they are *high-level plans* to which the agent is committed [3, 5]. Such high-level plans cannot be executed directly: they have to be *refined* as time goes by, resulting in more and more elaborate plans. The lower-level intentions that are inserted are *instrumental* for the high-level intention they refine [4]. At the end of the refinement process plans only have in *basic actions*: actions the agent can perform intentionally. Bratman's theory is at the basis of the by now huge literature on Belief-Desire-Intention (BDI) agents. However and as more extensively discussed in [12], the literature only contains few BDI logics where refinement is a central ingredient: essentially [17, 2, 14]. It is notably absent from Cohen and Levesque's logic [6] and Shoham's *database perspective* [18, 15, 19]. The latter is a simple account that is based on databases of time-indexed basic actions and beliefs. We believe it to be a promising basis for a logical analysis of intentions.

In order to extend Shoham's approach by an account of intention refinement, the first thing to do is to add high-level, temporally extended actions. We are also going to tackle another of its shortcomings, viz. that it does not solve the frame problem: the beliefs at time point t together with the intention at t fail to determine the beliefs at $t+1$. The reason is that Shoham's databases do not account for environment actions, alias events. We here add them to the picture. Just as in PDDL+ planning [10], we suppose that while the planning agent is proactive, the environment is reactive (or, more precisely, the planning agent believes so). Indeed, without such reactive events we would not be able to refine intentions. For example, consider the refinement of my intention to submit a paper to JELIA, which involves clicking Easychair's 'upload' button: I have to believe that my click action triggers the upload event in order to believe that clicking is a means for submitting. It is within this framework of high-level actions and reactive events we then study relations of refinement and instrumentality between intentions.[3]

2 Belief-intention databases

Let $\mathsf{Act} = \{\alpha, \beta, \dots\}$ be a finite set of actions. It contains a set of basic actions $\mathsf{Act}_0 = \{a, b, \dots\}$: actions that can be directly executed by the planning agent. Let $\mathsf{Evt}_0 = \{e, f, \dots\}$ be a finite set of basic events. Basic events and basic actions take one time unit to be executed. Let $\mathbb{P} = \{p, q, \dots\}$ be a finite set of propositional variables. The language of boolean formulas built from $\mathbb{P}$ is noted $\mathscr{L}_{\mathbb{P}}$.

Definition 1. *A dynamic theory is a tuple $\mathcal{T} = \langle pre, post \rangle$ with $pre, post : \mathsf{Act} \cup \mathsf{Evt}_0 \longrightarrow \mathscr{L}_{\mathbb{P}}$, such that that the postconditions of basic actions and events are conjunctions of literals: there are functions $\mathit{eff}^+, \mathit{eff}^- : \mathsf{Act}_0 \cup \mathsf{Evt}_0 \longrightarrow 2^{\mathbb{P}}$ such that for every $x \in \mathsf{Act}_0 \cup \mathsf{Evt}_0$, $post(x) = \Big(\bigwedge_{p \in \mathit{eff}^+(x)} p \Big) \wedge \Big(\bigwedge_{p \in \mathit{eff}^-(x)} \neg p \Big)$.*

We extend the functions $pre, post, \mathit{eff}^+$ and eff^- to sets: $pre(X) = \bigwedge_{x \in X} pre(x)$. We say that a dynamic theory $\mathcal{T}$ is *coherent* if and only if for every $a \in \mathsf{Act}_0$ and $E \subseteq \mathsf{Evt}_0$, if $pre(\{a\} \cup E)$ is consistent then $post(\{a\} \cup E)$ is consistent.

Example 1. Alice has a high-level action **buy** of buying a movie ticket and a basic action of buying a ticket online buyWeb. There is an event deliver of the website delivering the electronic ticket. Here is its coherent dynamic theory:

$pre(\mathbf{buy}) = \top$	$post(\mathbf{buy}) = \mathsf{Ticket}$
$pre(\mathsf{wait}) = \top$	$post(\mathsf{wait}) = \top$
$pre(\mathsf{buyWeb}) = \top$	$post(\mathsf{buyWeb}) = \mathsf{PaidWeb}$
$pre(\mathsf{deliver}) = \mathsf{PaidWeb} \wedge \neg\mathsf{Delivered}$	$post(\mathsf{deliver}) = \mathsf{Ticket} \wedge \mathsf{Delivered}$

In the rest of the paper we suppose a fixed background dynamic theory $\mathcal{T}$.

An agent's database contains her (incomplete) beliefs about the facts and about event occurrences together with her intentions. Occurrence of event $e \in \mathsf{Evt}_0$ at time point $t \in \mathbb{N}$ is noted (t, e). We also consider the agent's beliefs about non-occurrence of events. For that we define the set of event complements $\overline{\mathsf{Evt}_0} = \{\bar{e} : e \in \mathsf{Evt}_0\}$ and write $(t, \bar{e})$ for non-occurrence of e at t. An *intention* is a triple $i = (t, \alpha, d) \in \mathbb{N} \times \mathsf{Act} \times \mathbb{N}$ with $t < d$. It represents that the agent wants to perform α in the time interval $[t, d]$: it should start at or after t and end before or at deadline d. When $\alpha \in \mathsf{Act}_0$ then i is a *basic intention.*

Definition 2. *A belief-intention database is a finite set*

$$\Delta \subseteq (\mathbb{N} \times \mathscr{L}_{\mathbb{P}}) \cup (\mathbb{N} \times \mathsf{Evt}_0) \cup (\mathbb{N} \times \overline{\mathsf{Evt}_0}) \cup (\mathbb{N} \times \mathsf{Act} \times \mathbb{N}).$$

For example, $\Delta_A = \{(0, \mathbf{buy}, 3)\}$ is a database describing Alice's intention to buy a movie ticket within the temporal interval $[0, 3]$.

We define $\mathsf{end}(t, p) = t$, $\mathsf{end}(t, e) = t{+}1$, and $\mathsf{end}(t, \alpha, d) = d$.

[3] This work was partially supported by ANR-11-LABX-0040-CIMI within the program ANR-11-IDEX-0002-02 and CSC (Chinese Scholarship Council). A long version is available at `www.irit.fr/~Andreas.Herzig/P/Jelia16db.html`.

3 Semantics

The semantics of dynamic theories and belief-intention databases is in terms of *paths* defining for each time point which propositional variables are true, which basic events will occur, and which (single) basic action the agent will perform.

Definition 3. *A* path *is a triple* $\pi = \langle V, H, D\rangle$ *with* $V : \mathbb{N} \longrightarrow 2^{\mathbb{P}}$, $H : \mathbb{N} \to 2^{\mathsf{Evt}_0}$*, and* $D : \mathbb{N} \to \mathsf{Act}_0$*. It is a* $\mathcal{T}$-model *iff for every* $t \in \mathbb{N}$*:*

$$\begin{aligned} V(t{+}1) &= \big(V(t) \setminus \mathit{eff}^{-}\big(H(t){\cup}\{D(t)\}\big)\big) \cup \mathit{eff}^{+}\big(H(t){\cup}\{D(t)\}\big)\big) \\ H(t) &= \{e \in \mathsf{Evt}_0 \mid V(t) \models \mathit{pre}(e)\} \\ D(t) &\in \{a \in \mathsf{Act}_0 \mid V(t) \models \mathit{pre}(a)\} \end{aligned}$$

So in a $\mathcal{T}$-model: (1) the state at $t{+}1$ is determined by the state at t and the basic action and events occurring at t; (2) event e occurs *iff* $\mathit{pre}(e)$ is true (the environment is reactive); (3) basic action a occurs *implies* that $\mathit{pre}(a)$ is true (the agent is autonomous and may or may not perform executable actions).

Definition 4. *A* $\mathcal{T}$*-model* $\pi = \langle V, H, D\rangle$ satisfies *intention* $i = (t, \alpha, d)$*, noted* $\pi \Vdash_{\mathcal{T}} i$*, if there are* t', d' *such that* $t{\leq}t'{<}d'{\leq}d$*,* $V(t') \models \mathit{pre}(\alpha)$*,* $V(d') \models \mathit{post}(\alpha)$*, and* $\alpha \in \mathsf{Act}_0$ *implies* $D(t') = \alpha$*.*

So π satisfies (t, α, d) if α is executable at some point after t and can end before the deadline at a point where the postcondition of α is true. Moreover, when α is basic then it conforms to the 'do'-function D of π.

Definition 5. *A* $\mathcal{T}$*-model* $\pi = \langle V, H, D\rangle$ *is a* $\mathcal{T}$-model *of* Δ*, noted* $\pi \Vdash_{\mathcal{T}} \Delta$*, if*

- $(t, \varphi) \in \Delta$ *implies* $V(t) \models \varphi$*;*
- $(t, e) \in \Delta$ *implies* $e \in H(t)$*;*
- $(t, \overline{e}) \in \Delta$ *implies* $e \notin H(t)$*;*
- $i \in \Delta$ *implies* $\pi \Vdash_{\mathcal{T}} i$*.*

We say that Δ is $\mathcal{T}$*-satisfiable* if there exists a $\mathcal{T}$-model of Δ. Δ is a $\mathcal{T}$-*consequence of* Δ', noted $\Delta' \models_{\mathcal{T}} \Delta$, if every $\mathcal{T}$-model of Δ' is also a $\mathcal{T}$-model of Δ. We write $\Delta' \models_{\mathcal{T}} i$ when Δ is a singleton $\{i\}$.

Proposition 1. $\mathcal{T}$*-satisfiability and* $\mathcal{T}$*-consequence are decidable, for every* $\mathcal{T}$*.*

4 Refining an intention

A high-level intention cannot be executed directly by the agent: it can only be refined into lower-level intentions, until basic intentions are produced. For example, my high-level intention i to submit a paper to JELIA before its deadline June 30 is refined into the intention i_1 to register it on Easychair before June 30, the intention i_2 to upload it as a PDF file, etc.

Refinement consists in adding new intentions to the database while staying consistent. Intuitively, to refine an intention i means to add a minimal set of new intentions J to the database which, together with other intentions but i, suffice to entail i. Moreover, the deadlines of the refining intentions should be before that of the refined intention.

Definition 6. *Intention i is refinable to J in Δ, noted $\Delta \models_{\mathcal{T}} i \lhd J$, iff*

1. *there is no $j \in J$ such that $\Delta \models_{\mathcal{T}} j$;*
2. *$\Delta \cup J$ has a $\mathcal{T}$-model;*
3. *$(\Delta \cup J) \setminus \{i\} \models_{\mathcal{T}} i$;*
4. *$(\Delta \cup J') \setminus \{i\} \not\models_{\mathcal{T}} i$ for every $J' \subset J$;*
5. *$\mathsf{end}(j) \leq \mathsf{end}(i)$ for every $j \in J$.*

For our running example we have $\Delta_A \models_{\mathcal{T}} (0, \mathbf{buy}, 3) \lhd \{(0, \mathsf{buyWeb}, 1)\}$.

Proposition 2. *It is decidable whether $\Delta \models_{\mathcal{T}} i \lhd J$.*

5 Refinement and instrumentality

A higher-level intention and the lower-level intentions refining it should stand in a means-end relation: the lower-level means contribute to the higher-level end. This is also called the *instrumentality relation* $[1, 8, 16, 4]$.

Instrumentality cannot be defined from an action theory alone. First, the time point of action execution matters. For example, let us take up our intention of attending JELIA in November. Suppose I also have to go to the conference host city, Larnaca, in May, for some other reason. The postcondition of that action—to be in Larnaca—entails one of the preconditions of the attending JELIA action. However, my May intention does not contribute to my November intention. So the former is not necessarily instrumental for the latter. Second, the preconditions of the means are typically more demanding than the preconditions of the end; similarly, the postconditions of the means are more detailed than those of the end. For example, buying a movie ticket should *a priori* not require an adequate amount of money because there are other ways to buy a ticket, such as online with a credit card.

Formally, the instrumentality relation relates a refined high-level intention to a set of lower-level intentions, given a background database.

Definition 7. *Let Δ be a $\mathcal{T}$-satisfiable database. Let $i \in \Delta$ and let $J \subseteq \Delta$. Then J is* instrumental *for i in Δ, noted $\Delta \models_{\mathcal{T}} J \gtrdot i$, iff*

1. *$\Delta \setminus J \not\models_{\mathcal{T}} i$;*
2. *$(\Delta \setminus J) \cup \{j\} \models_{\mathcal{T}} i$ for every $j \in J$;*
3. *$\mathsf{end}(j) \leq \mathsf{end}(i)$ for every $j \in J$.*

When $\Delta \models_{\mathcal{T}} J \gtrdot i$ then J is a minimal set of intentions satisfying the counterfactual "if J was not in Δ then i would no longer be guaranteed by Δ" and all intentions of J terminate before or together with i. Note that when $\Delta \models_{\mathcal{T}} J \gtrdot i$ then J cannot be empty (because we require $i \in J$).

We now relate intention refinement and instrumentality: when $\Delta \models_{\mathcal{T}} i \lhd J$ then every element of J is instrumental for i in the refined database $\Delta \cup J$.

Theorem 1. *If* $\Delta \models_{\mathcal{T}} i \lhd J$ *then* $\Delta \cup J \models_{\mathcal{T}} \{i, j\} \gg i$ *for every* $j \in J$.

The converse does not hold: instrumentality cannot guarantee that the added intentions are new, so item 1 of Definition 6 does not necessarily hold.

6 Conclusion

We have extended Shoham's database view by temporally extended high-level intentions and STRIPS-like reactive environment events. The successive refinement of high-level intentions into lower-level intentions relies on the consequence relation of our semantics. The refined databases contain high- and low-level intentions that are related by the instrumentality relation. We have shown that satisfiability and consequence checking are still decidable in our extended database perspective.

The closest domain concerning intention refinement would be Hierarchical Task Networks (HTN) [9] which considering refinement of actions in a predefined and primitive way. With an HTN planner, a BDI agent system has been developed [17, 7]. In this paper we focus on refinement which is, in some way, a well-founded belief-intention database expansion. More general expansion may lead to unsatisfiable database and raises issues about withdrawal or revision of intentions. This is further explored in [13].

The next step is to investigate the revision of belief-intention databases. This is typically required when the agent learns a new piece of information about the environment. For example, suppose $(t, e) \in \Delta$ and the agent learns that e will not happen at t. This requires not only to contract other beliefs about facts and events, but also some of the agent's intentions. The instrumentality relation is of fundamental importance here: when $\Delta \models_{\mathcal{T}} J \gg i$ then the end intention i is deeper entrenched in the belief-intention database Δ than the means J to achieve i. So the agent should only abandon i once all possible ways of refining i have turned out to be unavailable. One possible relational postulate for revision is that the end intentions in the revised database should be a subset of the end intentions of the original database. There is currently little work on linking intention revision with instrumentality, with the exception of [11, 19]. However, these contributions are still preliminary as many issues are not yet solved, such as the frame problem or the relation between basic and non-basic actions. We intend to explore in future work the revision of a belief-intention database where rational change relies on instrumentality.

References

1. Audi, R.: A theory of practical reasoning. Journal of American Philosophical Quarterly 19, 25–39 (1982)
2. Baral, C., Gelfond, M.: Reasoning about intended actions. In: Proc. of the 20th National Conference on Artificial Intelligence (AAAI). vol. 20, pp. 689–694. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999 (2005)

3. Bratman, M.: Intention, plans, and practical reason. Cambridge: Harvard University Press (1987), reedited 1999 with CSLI Publications
4. Bratman, M.: Intention, belief, and instrumental rationality. Reasons for action pp. 13–36 (2009)
5. Bratman, M.E., Israel, D.J., Pollack, M.E.: Plans and resource-bounded practical reasoning. Journal of Computational Intelligence 4, 349–355 (1988)
6. Cohen, P.R., Levesque, H.J.: Intention is choice with commitment. Journal of Artificial intelligence 42(2), 213–261 (1990)
7. De Silva, L., Sardina, S., Padgham, L.: First principles planning in BDI systems. In: Proceedings of the 8th International Conf. on Autonomous Agents and Multiagent Systems (AAMAS). vol. 2, pp. 1105–1112 (2009)
8. Dignum, F., Conte, R.: Intentional agents and goal formation. In: Singh, M.P., Rao, A., Wooldridge, M. (eds.) Agent Theories, Architectures, and Languages: Agent Theories, Architectures, and Languages (ATAL-97), pp. 231–243. Springer Verlag, Berlin (1998)
9. Erol, K., Hendler, J., Nau, D.S.: HTN planning: Complexity and expressivity. In: Proc. of the 12th National Conference on Artificial Intelligence (AAAI). vol. 94, pp. 1123–1128 (1994)
10. Fox, M., Long, D.: Modelling mixed discrete-continuous domains for planning. J. Artif. Intell. Res. (JAIR) 27, 235–297 (2006), http://dx.doi.org/10.1613/jair.2044
11. Grant, J., Kraus, S., Perlis, D., Wooldridge, M.: Postulates for revising BDI structures. Synthese 175(1), 39–62 (2010)
12. Herzig, A., Lorini, E., Perrussel, L., Xiao, Z.: BDI logics for BDI architectures: old problems, new perspectives. KI Kństliche Intelligenz (to appear)
13. Herzig, A., Perrussel, L., Xiao, Z.: On hierarchical task networks. In: Michael, L., Kakas, A.C. (eds.) Proc. JELIA 2016. Springer Verlag (2016)
14. Hunsberger, L., Ortiz Jr., C.: Dynamic intention structures I: a theory of intention representation. Proc. of Autonomous Agents and Multi-Agent Systems (AAMAS) 16(3), 298–326 (2008)
15. Icard, T., Pacuit, E., Shoham, Y.: Joint revision of belief and intention. In: Proc. of the 12th International Conference on Principles of Knowledge Representation and Reasoning (KR). pp. 572–574 (2010)
16. Lorini, E., Herzig, A.: A logic of intention and attempt. Synthese 163(1), 45–77 (2008)
17. Sardina, S., de Silva, L., Padgham, L.: Hierarchical planning in BDI agent programming languages: A formal approach. In: Proc. of the 5th International Conference on Autonomous Agents and Multiagent Systems (AAMAS). pp. 1001–1008. ACM (2006)
18. Shoham, Y.: Logical theories of intention and the database perspective. Journal of Philosophical Logic 38(6), 633–647 (2009)
19. Van Zee, M., Doder, D., Dastani, M., Van Der Torre, L.: AGM revision of beliefs about action and time. In: Proc. of the 24th International Joint Conference on Artificial Intelligence (IJCAI). pp. 3250–3256 (2015)