

Joint Revision of Belief and Intention

Thomas Icard
Department of Philosophy
Stanford University

Eric Pacuit
Center for Logic and Philosophy of Science
Tilburg University

Yoav Shoham
Department of Computer Science
Stanford University

Abstract

We present a formal semantical model to capture action, belief and intention, based on the “database perspective” (Shoham 2009). We then provide postulates for belief and intention revision, and state a representation theorem relating our postulates to the formal model. Our belief postulates are in the spirit of the AGM theory; the intention postulates stand in rough correspondence with the belief postulates.

Motivation

While there is an extensive literature developing logical models to reason about changing *informational* attitudes (eg., belief, knowledge, certainty), other mental states have received less attention. However, this is changing with recent articles introducing dynamic logics of intention. These papers take as a starting point logical frameworks derived from Cohen and Levesque’s seminal paper (Cohen and Levesque 1990) aimed at formalizing Bratman’s planning theory of intention (Bratman 1987). In this paper we take a different angle on intentions, focusing on intention revision as it relates to, and is intertwined with, belief revision.

We view the problem of intention revision as a database management problem (see (Shoham 2009) for more on the conceptual underpinnings of this standpoint). At any given moment, an agent must keep track of a number of facts about the current situation. This includes beliefs about the current state, beliefs about possible future states, which actions are available now and in the future, and also what the agent plans to do at future moments. It is important that all of this information be *jointly consistent* at any given moment and furthermore that it can be *modified* as needed while maintaining consistency.

In the following we introduce a simple logic that formally models such a “database”. That is, *consistency* in this logic is meant to represent not only that the agent’s beliefs are consistent and the agent’s future plan is consistent, but also that the agent’s beliefs and intentions together form a *coherent* picture of what may happen, and of how the agent’s own actions will play a role in what happens.

What can cause an agent’s database to change? In this paper, we focus on two main sources:

Copyright © 2010, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

1. The agent makes some observation, e.g. from sensory input. If the new observation is inconsistent with the agent’s beliefs, these beliefs will have to be revised to accommodate it. While we recognize the classical AGM theory (Alchourrón, Gärdenfors, and Makinson 1985) is not without problems, in particular when it comes to iterated revision, our account of belief revision simply adopts this framework. The goal is thus to give general conditions on a single revision with new information *that the agent has already committed to incorporating*.
2. The agent forms a new intention. We focus on *future-directed* intentions, understood as time-labelled actions that might make up a plan. Analogously to belief revision, it is assumed the agent has already committed to a new intention, so it must be accommodated by any means short of revising beliefs. The force of the theory is in restricting how this can be accomplished. To be more precise, we purport to model an intelligent database, which receives instructions from some planner (e.g. a STRIPS-like planner) that is itself engaged in some form of practical reasoning. The job of the database is to maintain consistency and coherence between intentions and beliefs.

Since many intentions are dependent on (lack of) certain beliefs, belief revision will in general trigger intention revision. This is an important part of our model. To be sure, intentions can also give rise to new beliefs. If an agent intends to go to San Francisco, he may proceed on the assumption that he will be in San Francisco. However, this we think of as a different, “optimistic” kind of belief that can be derived from coherent beliefs and intentions. In this paper we model “concrete” beliefs, including, e.g. what (sequences of) actions will be possible in the future, independent of the agent’s plans.

Logical Preliminaries

Entries in the database will be represented by the formal language \mathcal{L} given by the following grammar:

$$\varphi ::= p_t \mid pre(a)_t \mid post(a)_t \mid Do(a)_t \mid \Box\varphi \mid \varphi \wedge \varphi \mid \neg\varphi$$

Intuitively, p_t means that the atomic formula p is true at time t and $Do(a)_t$ means the agent will do (or did) action a at time t . To every action and every time we associate formulas $pre(a)_t$ and $post(a)_{t+1}$, which we treat as distinguished

propositional variables, and are understood as the preconditions and postconditions of a at time t . The modal operator is interpreted as historic necessity. The other boolean connectives and the dual modal operator \diamond are defined as usual.

Definition 1 (Paths). Let P be the set

$$\mathcal{P}(\text{Prop} \cup \{pre(a), post(a) : a \in \text{Act}\}).$$

A path $\pi : \mathbb{Z} \rightarrow (P \times \text{Act})$ assigns to each time t the set of proposition-like formulas true at that time, and the next action a on the path. Let $\pi(t)_1$ denote the left projection and $\pi(t)_2$ denotes the right projection. A path is called *appropriate* if the following obtains:

$$\text{If } \pi(t)_2 = a, \text{ then } post(a) \in \pi(t+1)_1.$$

There is a natural equivalence relation on a set Π of paths: we write $\pi \sim_t \pi'$ if for all $t' \leq t$, $\pi(t') = \pi'(t')$. Intuitively, $\pi \sim_t \pi'$ if π and π' represent the same situation up to time t . We extend the definition of *appropriate* to sets of paths by declaring Π to be appropriate if all paths $\pi \in \Pi$ are appropriate and moreover satisfy the following condition:

$$\text{If } pre(a) \in \pi(t)_1, \text{ then there is some } \pi' \sim \pi \text{ such that } \pi'(t)_2 = a.$$

Definition 2 (Truth Definition). The truth relation \models_{Π} is defined relative to some underlying appropriate set of paths Π . For convenience we leave off the relativizing subscript.

$$\begin{aligned} \pi, t \models \alpha_{t'}, & \text{ iff } \alpha \in \pi(t')_1, \text{ with } \alpha \equiv p, pre(a), \text{ or } post(a). \\ \pi, t \models Do(a)_{t'}, & \text{ iff } \pi(t')_2 = a. \\ \pi, t \models \Box\varphi, & \text{ iff for all } \pi' \in \Pi, \text{ if } \pi \sim_t \pi' \text{ then } \pi', t \models \varphi. \\ \pi, t \models \varphi \wedge \psi, & \text{ iff } \pi, t \models \varphi \text{ and } \pi, t \models \psi. \\ \pi, t \models \neg\varphi, & \text{ iff } \pi, t \not\models \varphi. \end{aligned}$$

The proof of this theorem, giving the theory of our databases, is by standard techniques.

Theorem 1 (The logic L_{Path} of paths). *The following logic is sound and strongly complete with respect to the class of all appropriate sets of paths. We call this logic L_{Path} .*

1. *Propositional Tautologies, Closure under Modus Ponens;*
2. **S5** *axioms and rules for \Box ;*
3. $\bigvee_{a \in \text{Act}} Do(a)_t;$ 5. $Do(a)_t \rightarrow \bigwedge_{b \neq a} \neg Do(b)_t;$
4. $Do(a)_t \rightarrow post(a)_{t+1};$ 6. $pre(a)_t \rightarrow \diamond Do(a)_t;$

Modeling Revision

Beliefs in our framework are represented by sets of L_{Path} -consistent formulas of \mathcal{L} , or equivalently, as (appropriate) sets of paths. Given a set of formulas B , we can consider the set of paths on which all formulas of B hold at time 0, denoted $\rho(B)$. Conversely, given a set of paths Π , we let $\beta(\Pi)$ be defined as the set of formulas valid at 0 in all paths in Π . We will use this correspondence in the representation theorem. For now we restrict our attention to sets of paths, and in particular we will represent beliefs by the minimal set under a total preorder on paths. Intentions in our models will simply be action/time pairs.

Postconditions of actions always hold on a path, but preconditions may not. Even if all of the paths in some (minimal) set include action a being taken at time t , it need not

be that the preconditions also hold along all paths at t . We might therefore think of our belief model as, in some sense, one of “optimistic” or “imaginary” beliefs. On the other hand, we do put a slightly weaker requirement on sets of paths, that the preconditions hold on *some* path in the set. Where again I is a set of pairs (a, t) , we require that the joint preconditions of all intended actions not be *disbelieved* by the agent. This is our notion of coherence.

Definition 3 (Coherence). The pair (Π, I) is said to be *coherent* (at time 0) if there is some path $\pi \in \Pi$, such that $\pi, 0 \models \diamond \bigwedge_{(a,t) \in I} pre(a)_t$.

Intuitively, intentions cohere with beliefs if the agent considers it possible to carry out all of the intended actions. This is a kind of minimal requirement on *rational balance* between the two mental states.

Selection functions are simply the intention revision postulates given in the first section, under a different guise.

Definition 4 (Selection Function). A *selection function* γ is a function that assigns an intention set to a tuple consisting of a set of paths, an intention set and a pair (a, t) satisfying the following conditions. If $\gamma(\Pi, I, (a, t)) = I'$ then,

1. (Π, I') is coherent;
2. If $(\Pi, \{(a, t)\})$ is coherent, $(a, t) \in I'$;
3. If $(\Pi, I \cup \{(a, t)\})$ is coherent, then $I' = I \cup \{(a, t)\}$.
4. $I' \subseteq I \cup \{(a, t)\}$.

In the simple case of the empty intention pair ϵ , this reduces merely to requiring coherence.

Definition 5 (Belief Sets). Suppose Π is an appropriate set of paths. If we define a total preorder \leq on Π , then the *belief set* of (Π, \leq) is the set $\{\pi \in \Pi : \pi \leq \pi' \text{ for all } \pi' \in \Pi\}$. We denote this by $\min_{\leq}(\Pi)$, or just $\min(\Pi)$ when the ordering is understood from context.

Definition 6 (Belief Intention Model). A belief-intention model is a triple (Π, \leq, I, γ) where Π is a set of paths, \leq is a total preorder on Π , I is a finite set of pairs (a, t) with $a \in \text{Act}$ and $t \in \mathbb{Z}^+$, $(\min(\Pi), I)$ is coherent and γ is a selection function.

Definition 7 (Adding an Intention). Let (Π, \leq, I, γ) be a belief-intention model. Adding the intention (a, t) results in the model (Π, \leq, I', γ') where $I' = \gamma(\min(\Pi), I, (a, t))$ and $\gamma' = \gamma$. We denote this model by $(\Pi, \leq, I, \gamma) \bullet (a, t)$.¹

Definition 8 (Adding a Belief). Let (Π, \leq, I, γ) be a belief-intention model. Adding a (consistent) belief φ results in the model $(\Pi, \leq', I', \gamma')$, where $\gamma' = \gamma$, $I' = \gamma(\min_{\leq'}(\Pi), I, \epsilon)$, and \leq' is defined so that $\pi \leq \pi'$, if and only if one of the following holds:

1. $\pi, 0 \models \varphi$ and $\pi', 0 \not\models \varphi$;
2. $\pi, 0 \models \varphi$ and $\pi', 0 \models \varphi$, and $\pi \leq \pi'$;
3. $\pi, 0 \not\models \varphi$ and $\pi', 0 \not\models \varphi$, and $\pi \leq \pi'$.

The new belief-intention model is denoted $(\Pi, \leq, I, \gamma) \star \varphi$.

¹Notice that this setup allows the possibility that $\gamma' \neq \gamma$, so that after revision the selection function itself can change. Of course this would only become interesting in the iterated case

Representation of Revision Postulates

In the following let $Cl(X)$ denote the closure of a set X of \mathcal{L} formulas under consequence in \mathbf{L}_{Path} . And if I is a finite set of pairs (a, t) , with $a \in \mathbf{Act}$ and $t \in \mathbb{Z}^+$, define, $Cohere_I := \diamond \bigwedge_{(a,t) \in I} pre(a)_t$.

Definition 9 (Belief Intention Base). A belief intention base is a pair $\langle B, I \rangle$, where:

- B is a consistent set of formulas such that $Cl(B) = B$.
- I is a finite set of pairs (a, t) .

Definition 10 (Coherence). A belief-intention base $\langle B, I \rangle$ is coherent if $\neg Cohere_I \notin B$.

We then have the following obvious correspondence.

Lemma 1. $\langle B, I \rangle$ is coherent, iff $(\rho(B), I)$ is coherent.

Now having provided all of the necessary formal details, we present our postulates for intention and belief revision.

Definition 11 (Intention Revision). Suppose $\langle B, I \rangle \circ (a, t) = \langle B', I' \rangle$. The operator \circ is called *proper* if the following conditions obtain.

1. $\langle B', I' \rangle$ is coherent;
2. If $\langle B, \{(a, t)\} \rangle$ is coherent, then $(a, t) \in I'$;
3. If $\langle B, I \cup \{(a, t)\} \rangle$ is coherent, then $I \cup \{(a, t)\} \subseteq I'$;
4. $I' \subseteq I \cup \{(a, t)\}$;
5. $B' = B$.

The first postulate simply says that intention revision should restore coherence. The second postulate says that the new intention (a, t) takes precedence over all other currently held intentions; it should be added if it is possible to maintain coherence, even if this means discarding current intentions. The third postulate, taken together with the fourth postulate, says that if it is possible to maintain coherence by simply adding the new intention, then this is the only change that is made. The fourth in addition guarantees that, unlike in the case of belief revision below, no extraneous intentions are ever added. Finally, the fifth postulate says that non-contingent beliefs do not change with intention revision.

We assume every belief revision operator $*$ is given with its own intention revision operator \circ^* , so that a belief revision may trigger an intention revision.

Definition 12 (Belief Revision). Suppose $\langle B, I \rangle * \varphi = \langle B', I' \rangle$. The operator $*$ is called *proper* if the following conditions obtain.

1. $\langle B', I' \rangle = \langle B', I \rangle \circ^* \epsilon$, where \circ^* is proper;
2. φ is consistent, iff $\varphi \in B'$;
3. If $\neg \varphi \notin B$, then $Cl(B \cup \{\varphi\}) = B'$;
4. If $\mathbf{L}_{Path} \vdash \varphi \leftrightarrow \psi$ and $\langle B, I \rangle * \psi = \langle B'', I'' \rangle$, then $B' = B''$;
5. $B' = Cl(B')$;
6. If $\neg \psi \notin B'$ and $\langle B, I \rangle * \psi = \langle B'', I'' \rangle$, then we have $Cl(B' \cup \{\psi\}) \subseteq B''$;
7. If $\langle B, I'' \rangle * \varphi = \langle B'', I''' \rangle$, then $B' = B''$.

Postulate 1 simply says that if intention revision is necessary to retain coherence, this revision is itself proper. Postulate 2 is a slight variation of the AGM success postulate, which we adopt on a par with intention revision postulate 2. In this setting it only makes sense to adopt a new belief if it is non-contradictory. Postulates 3-6 fill out the rest of the AGM theory, and postulate 7 says that the underlying intention set is irrelevant to belief revision (see above).

We can now represent these postulates in terms of the belief intention models of Definition 6.

Theorem 1 (Representation Theorem). For every belief intention base $\langle B, I \rangle$, with proper revision functions $*$ and \circ^* , there is a belief intention model (Π, \leq, I, γ) , such that:

1. $\rho(B) = \min_{\leq}(\Pi)$;
2. I is the same set in the base and in the model;
3. For all $\varphi \in \mathcal{L}$: If $(\Pi, \leq, I, \gamma) * \varphi = (\Pi, \leq', I', \gamma')$ and $\langle B, I \rangle * \varphi = \langle B', I'' \rangle$, then,

$$\rho(B') = \min_{\leq'}(\Pi), \text{ and } I' = I''.$$

The proof of this theorem simply rides on the proof of the representation theorem for AGM in terms of the “system of spheres” interpretation (Grove 1988), with the intention revisions simply going along for the ride.

Conclusions and Future Work

In a sense, one can see the AGM framework for belief revision as identifying what the problem of belief revision is in the first place. The standard postulates can be taken as *constitutive* of a particular kind of doxastic action, according to which the agent has committed to believing some new piece of information and must integrate this new belief with old beliefs. The interesting questions, on this view, arise when we ask how this simple picture can be embellished, to deal with iterated belief revision, interaction with other mental states and actions, and so on. In the same way, one can view our treatment of joint intention and belief revision in this paper as a proposal to define what the problem is about, and to propose a framework in which further questions can be fruitfully asked and explored. Indeed, there are many directions from here that should be explored (see (Shoham 2009) for a list). We leave this for future work.

References

- Alchourrón, C. E.; Gärdenfors, P.; and Makinson, D. 1985. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic* 50(2):510 — 530.
- Bratman, M. 1987. *Intention, Plans and Practical Reason*. Harvard University Press.
- Cohen, P. R., and Levesque, H. 1990. Intention is choice with commitment. *Artificial Intelligence* 42(3):213 — 261.
- Grove, A. 1988. Two modellings for theory change. *Journal of Philosophical Logic* 17.
- Shoham, Y. 2009. Logical theories of intention and the database perspective. *Journal of Philosophical Logic* 38(6).