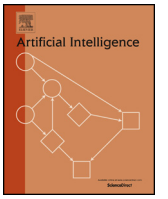
Artificial Intelligence 283 (2020) 103270

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint



Intention as commitment toward time ☆

Marc van Zee [a,*], Dragan Doder [b], Leendert van der Torre [c], Mehdi Dastani [b], Thomas Icard [d], Eric Pacuit [e]

[a] *Google Research, Brain Team, the Netherlands*
[b] *Utrecht University, the Netherlands*
[c] *University of Luxembourg, Luxembourg*
[d] *Stanford University, CA, USA*
[e] *University of Maryland, MD, USA*

ARTICLE INFO

Article history:
Received 11 February 2018
Received in revised form 13 March 2020
Accepted 21 March 2020
Available online 26 March 2020

Keywords:
Intention
BDI logic
Belief revision

ABSTRACT

In this paper we address the interplay among intention, time, and belief in dynamic environments. The first contribution is a logic for reasoning about intention, time and belief, in which assumptions of intentions are represented by preconditions of intended actions. Intentions and beliefs are coherent as long as these assumptions are not violated, i.e. as long as intended actions can be performed such that their preconditions hold as well. The second contribution is the formalization of what-if scenarios: what happens with intentions and beliefs if a new (possibly conflicting) intention is adopted, or a new fact is learned? An agent is committed to its intended actions as long as its belief-intention database is coherent. We conceptualize intention as commitment toward time and we develop AGM-based postulates for the iterated revision of belief-intention databases, and we prove a Katsuno-Mendelzon-style representation theorem.

1. Introduction

Sometime in the near future you will tell one of your household robots: “Bobby, get me some beer from the store.” Bobby confirms your request, but when it is walking to the store it encounters your partner, who says: “Bobby, our house is a mess, go home and clean.” Bobby returns home, takes the mop out of the closet and prepares to start cleaning. Just as it is ready to make its first swipe, one of your friends walks in asking: “Bobby, it has been snowing outside, could you clean my car?” In the meantime, you are getting increasingly frustrated by your lack of beer, and when you see Bobby in the kitchen you shout: “You still didn’t get my beers? Go get them immediately!” After letting Bobby run around for a few days you complain to the manufacturer that your robots do not finish any of the tasks they start with.

After the manufacturer updates the software on your robots, he happily tells you the new version will no longer cause the robots to drop their commitments so quickly. Delighted, you exclaim to your favorite robot: “Bobby, I have some friends coming over tonight, get some ingredients and cook dinner so we can eat at 7pm tonight”. Realizing the shop closes only at 5pm, Bobby 2.0 delays going to the grocery story until the very last moment, hereby keeping its schedule free for other possible tasks. Unexpectedly, on its way to the grocery store Bobby is delayed by an open bridge and arrives at the store

☆ This article is a revised and extended version of the conference papers [25,56,60,58]. More details can be found in the related work section.

* Corresponding author.
E-mail address: marcvanzee@gmail.com (M. van Zee).

https://doi.org/10.1016/j.artint.2020.103270

minutes after closing time. It returns to your home empty-handed, leaving you and your friends hungry. It turns out Bobby 2.0 is delaying every task until just before the deadline. Since tasks often have unexpected delays, this means that most of the tasks are finished too late, or not at all. Frustrated, you call the manufacturer again, complaining that Bobby is procrastinating its commitments.

After uploading yet another version of the software, the manufacturer ensures you that your robots will no longer postpone fulfilling their commitments, nor will they drop them quickly. At this time you are rather skeptical, but still you ask: "Bobby, I'd like to have dinner tonight again. Please buy the ingredients at 2pm, and cook for me at 6pm". Around 12pm, your partner again realizes the house hasn't been cleaned properly for a long time, and therefore tells Bobby to clean the house intensively. At 6pm, you sit at your kitchen table wondering where dinner is, so you call Bobby asking what happened. Bobby explains it had to clean the house at 12pm, which took three hours, so it couldn't fulfill its commitment to go shopping at 2pm.

Disappointed, you return your robots to the manufacturer where they are dismantled.

1.1. Commitment toward time

The story of Bobby the robot is inspired by the example of Willie the robot, due to Cohen and Levesque [13]. In their highly-cited article entitled "Intention = Choice + Commitment," Cohen and Levesque specify the rational balance of autonomous agents, focusing on the role that intention plays in maintaining this balance. Their approach has typified much subsequent research on belief-desire-intention (BDI) logics, namely to understand and study intention as commitment in relation to goals, desires and beliefs. For instance, Rao and Georgeff [39] define various commitment strategies of agents, such as blindly-minded, single-minded, and open-minded commitment strategies. A popular approach to formalize the BDI theory is to specify a temporal logic such as linear-time logic (LTL) or computational-tree logic (CTL*) and use modal operators for mental states and use expressions of the form "some time in the future," or "in the next time moment" to reason about the temporal behavior.

In approaches following the ideas of Cohen and Levesque by, for example, Rao and Georgeff [39] or Meyer et al. [33], intention is typically defined as commitment toward goals. However, being committed toward a goal is only one dimension of a commitment. Another important dimension is *commitment toward time*, i.e., *when* these commitments will be fulfilled. In the example above, Bobby the robot is an *online system*; it is receiving orders, forming plans, scheduling tasks, and executing them, all in parallel and in real-time. Bobby plans its commitments at appropriate moments, making sure the different plans do not overlap or are incompatible, while it at the same time may receive new instructions from users.

Even in a simplified setting where we only consider commitments toward time, already non-trivial complications arise. Commitments can play the role of *assumptions* on which further plans are based. The three versions of Bobby the household robot behave differently concerning their commitment to time. The first version may use a stack-like data structure in order to execute its tasks: it adds each new commitments on the stack, and then executes the tasks on top of its stack. The second version may use a queue, and moreover delays executing these tasks until the very last moment. Finally, the last version is able to schedule its commitments in time, but it cannot reschedule them. None of these versions seem to be able to fulfill commitments in a desirable way. Instead, Bobby should be able to make plans, store the commitments and use them as assumptions in further planning.

1.2. Methodology: the database perspective

Shoham [46] views the problem of intention revision as a database management problem. In particular, he introduced the conceptual underpinnings of the distinction between a reasoner such as a planner, and the involved belief-intention databases. 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, beliefs about which actions are available now and in the future, and also beliefs about plans at future moments. It is important that all of this information be jointly consistent at any given moment and furthermore can be modified as needed while maintaining consistency.

In this article we introduce a logic that formally models such a "database", as visualized in Fig. 1. *Consistency* in this logic is meant to represent not only that the agent's beliefs are consistent and the agent's future plans are 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. Our primary contribution in this article is to focus also on how the database is to be modified, and in the process to provide a clear picture of how intentions and beliefs relate.

In this paper we distinguish two kinds of outputs produced in Fig. 1:

Belief A belief is added to the Belief database. If the new belief is inconsistent with the existing beliefs, then these beliefs will have to be revised to accommodate it. Our account of belief revision follows the classical AGM postulates [1], which we then generalize to iterated revision. The goal is thus to give general conditions on revision with new information that the agent has already committed to incorporating.

Intention An intention is added to the Intention database. Here we focus on future directed intentions, understood as time-labeled actions pairs (a, t) 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.

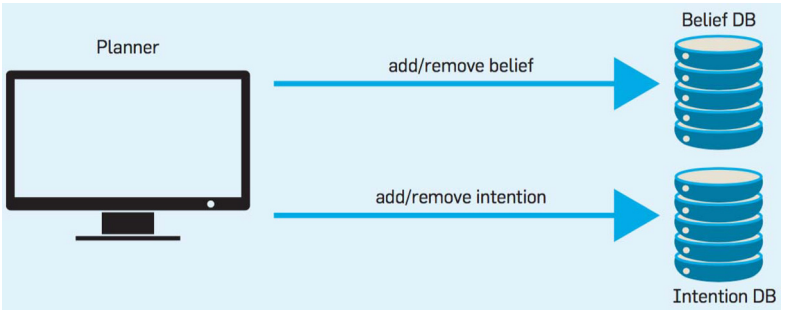


Fig. 1. The database perspective.

> 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 reasoner (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 among intentions and beliefs.

This description, however, obscures some important subtleties in the interaction between beliefs and intentions. The following will serve as a running example that we will use frequently throughout the article.

Example 1 *(Running example).* Bobby the household robot has the goal to buy groceries in the morning and to buy cleaning equipment in the afternoon. However, it only has sufficient budget to do either one of the two, but not both of them. Bobby thus believes it is possible to buy cleaning equipment and to buy food, but it also believes it is impossible to buy both. If Bobby decides to buy food, then it will cook food in the afternoon.

Upon adopting the intention to buy food, Bobby will come to have new beliefs based on the predicted success of this intention, e.g., that he will be able to cook afterwards. These further beliefs are important when planning when or how to cook. The intention is also supported by the absence of certain beliefs. It would be irrational for Bobby to adopt the intention to buy food if it believed it did not have sufficient money. Likewise, even if it originally believed it has sufficient money, upon learning it does not, the intention to cook food should be dropped. Yet, when dropping this intention, other beliefs, such as that he will be able to cook, have to be dropped as well, which may in turn force other intentions and beliefs to be dropped. And so on.

1.3. Weak beliefs

Regular beliefs concern the world as it is, independent of an agent's future plans, but including what sequences of actions are possible. Thus, additionally to atomic facts, an agent may have beliefs about what the preconditions and postconditions of actions are, and about which sequences of actions are jointly possible.

We distinguish *weak* beliefs depending on intentions, from beliefs that do not (*strong* beliefs). In other words, as usual we assume that postconditions lead to weak beliefs. Since we are not considering actions whose effects are uncertain or dependent on the conditions that obtain when the action is taken, if an action is planned the planner believes whatever follows from it.

Postconditions-as-weak-beliefs: *If an agent intends to take an action, it weakly believes that its postconditions will hold.*

A key element in our approach is that we treat preconditions of actions as *assumptions*. This leads to an asymmetry on beliefs about preconditions and postconditions of actions. However, for preconditions we adopt Shoham's weaker requirement, which we call here *preconditions-as-assumptions*:

Preconditions-as-assumptions: *If you intend to take an action you cannot believe that its preconditions do not hold.* [46]

Preconditions of actions are treated as *assumptions*, in the sense that an agent forms intentions under the assumption that these preconditions will be made true somewhere in the future. Treating preconditions as assumptions is a good fit with how real-time planning agents operate, because intended actions may be added as long as they are consistent with beliefs, and once they are accepted they can be used as additional assumptions to further plans [46]. For instance, the household robot Bobby has the intention to cook dinner tonight, which is based on the intention to buy the ingredients, which is in turn based on the assumption that your friends will attend tonight, even if it does not know this for sure yet. It is only when Bobby finds out your friends are not coming, it should drop its intentions. This preconditions-as-assumptions requirement

is sometimes called *strong consistency*, and is weaker than Bratman's *means-end coherence* requirement [8] (see Section 6.1.2 for a more detailed discussion).

The preconditions-as-assumptions requirement has various consequences, both for the logic and for the theory change operators introduced in this paper. For example, suppose Bobby intends to buy food at moment 0, and in addition to buy cleaning equipment at moment 2. This may be coherent despite Bobby's belief that it does not have enough money for both, thanks to an action of robbing a bank that it can perform at 1. If this is the only model where Bobby can buy food at moment 0 and equipment at moment 2, then the resulting belief-intention database may entail $do(rob)_1$, even if Bobby would like to avoid robbing a bank by any means. We will consider these potential complications in Section 4.4 after we have introduced the formal machinery.

1.4. Results

We develop a branching-time temporal logic, called Parameterized-time Action Logic (PAL) in order to formalize beliefs. The language of this logic contains formulas to reason about possibility, preconditions, postconditions, and the execution of actions. The semantics of this logic is close to CTL*, and in this way follows the tradition of BDI logics of Rao and Georgeff [39]. An important difference is that we do not use modal operators to reason about time, but we use explicit time points. We axiomatize this logic and proof that the axiomatization is sound and strongly complete with respect to our semantics.

We separate strong beliefs from weak beliefs as described above. Strong beliefs are beliefs that occur in the belief database, and they are independent of intentions. Weak beliefs are obtained from strong beliefs by adding intentions to the strong beliefs, and everything that follows from that. We then formalize a *coherence condition* on the beliefs and intentions. This condition states that the agent weakly believes it is possible to jointly perform all of its intended actions.

The main technical result of the paper is that we develop a set of postulates for the joint revision of belief and intentions, and that we prove a variation of the Katsuno and Mendelzon [26] representation theorem. To this end, we define a revision operator that revises beliefs up to a specific time point. We show that this leads to models of system behaviors which can be finitely generated, i.e. be characterized by a single formula. We also prove various representation theorems for iterated revision of belief and intention.

1.5. Map of the paper

The structure of this paper is as follows. In Section 2 we introduce and motivate our logic PAL, we axiomatize it and we prove completeness. In Section 3 we separate strong beliefs from weak beliefs, and we formalize a coherence condition on belief and intention. In Section 4 we study single-step revision of beliefs and intentions, and we study iterated revision in Section 5. We discuss related work in both philosophy of mind and AI in Section 6, and we provide directions for future work in Section 7. All the proofs can be found in a separate technical report.[1]

2. Parameterized-time Action Logic (PAL): a logic for belief and intention

Our aim in this section is to develop a logical system that represents an agent's beliefs about the current moment and future moment and actions that may be performed. Table 1 contains all the most important symbols used in this article and their meaning.

2.1. Syntax

Beliefs are represented by the formal language $\mathscr{L}$. The language uses the set Prop containing all atoms which are true or false in a time instance (state). We also consider formulas $do(a)_t$ which will semantically be defined as a transition from t to $t+1$ using action a.

Definition 1 *(Language).* Let

- $\mathsf{Act} = \{a, b, c, \ldots\}$ be a finite set of deterministic primitive actions;
- $\mathsf{Prop} = \{p, q, r, \ldots\} \cup \{pre(\overline{a}), post(a)\}$ be a finite set of propositions where $\overline{a} = (a_1, a_2, \ldots)$ is a non-empty action sequence and $\{a, a_1, a_2, \ldots\} \subseteq Act$ are actions. We denote atomic propositions with χ.

The sets Prop and Act are disjoint. The language $\mathscr{L}$ is inductively defined by the following BNF grammar:

$$\varphi ::= \chi_t \mid do(a)_t \mid \Box_t \varphi \mid \varphi \wedge \varphi \mid \neg\varphi,$$

with $\chi \in \mathsf{Prop}$, $a \in \mathsf{Act}$, and $t \in \mathbb{N}$. Furthermore, we abbreviate $\neg\Box_t\neg$ with $\Diamond_t$, and we define $\bot \equiv p_0 \wedge \neg p_0$ and $\top \equiv \neg\bot$.

[1] See Supplementary material.

Table 1
Symbols used in this paper and their meaning. Note some symbols are used multiple times.

Logic (Section 2, 3)		**Revision** (Section 4, 5)	
Symbol	Meaning	Symbol	Meaning
$\mathscr{L}$	Language of PAL	$\circ$	AGM revision function
Prop	Set of atoms	$\mathbb{X}^{\mid t}$	Some set $\mathbb{X}$ bounded up to t
Act	Set of actions	$\circ_t$	Strong belief revision function
$\overline{a}$	Finite action sequence	$\otimes_t$	Intention revision function
χ	Atomic propositions	$*_t$	Belief-intention revision function
T	Semantic tree (S, R, v, act)	$\pi^{\mid t}$	t-bounded path
S	Set of states in a tree	$m^{\mid t}$	t-bounded model
R	Accessibility relation for tree	γ_I^t	Selection function
v	Valuation function	$\preceq_\psi^t$	Total pre-order over models
act	Action function	ε	Empty intention
π	Path $(s_0, s_1, \ldots)$ in a tree	Ψ	Epistemic state
π_t	The t-th state of the path π	$\circ_t$	Epistemic revision function
$m = (T, \pi)$	Model in PAL	(Ψ, I)	Epistemic belief-intention database
M	Set of models in PAL	$*_t$	Iterated revision function
$\mathbb{M}$	Set of all models in PAL	κ_t	Spohn ranking function
$Mod(\varphi)$	Set of all models of φ	$Bel(\kappa_t)$	Accepted propositions
$\mathbb{SB}$	Set of all strong beliefs	$\bullet_t$	Extended Spohn revision
SB	Set of strong beliefs	$\mathscr{L}^{\mid t}$	Restricted language
$Cn(SB)$	Belief database		
$\mathbb{BD}$	Set of all belief databases		
MSB	Set of models of strong beliefs		
$\mathbb{MSB}$	Set of all MSBs		
(a, t)	Intention		
$\mathbb{I}$	Set of all intentions		
I	Intention database		
$\mathbb{ID}$	Set of all intention databases		
(SB, I)	Belief-intention database		
$\mathbb{BI}$	Set of all BI databases		
$WB(SB, I)$	Weak beliefs		
$Ext(M^{\mid t})$	Set of extensions of $M^{\mid t}$		
$\mathbb{EBI}$	Set of all epistemic databases		
$Cohere(I)$	Coherence formula		

Intuitively, p_t means that the atomic formula p is true at time t, $do(a)_t$ means that action a is executed at time t. To every finite sequence of actions $\overline{a} = (a_1, a_2, \ldots)$ and every time point t we associate a formula $pre(\overline{a})_t$, which is understood as the precondition for subsequently executing actions $a_1, a_2, \ldots$ at time t. Note pre and postconditions are represented as particular propositions, and not mappings abbreviations of other propositional formulas.

We define preconditions for sequences of actions explicitly, because it is difficult to define the precondition for a sequence of actions using only preconditions for individual actions. This can already be witnessed in our running example: Bobby believes the preconditions to buy food and cleaning equipment are true separately, but still does not believe the precondition for performing both actions subsequently is true. These types of formulas will play a crucial role when we formalize the coherence condition in Section 3.

The modal operator $\Box_t$ is interpreted as necessity, indexed with a time point t. Intuitively, a formula of the form $\Box_t p_{t+1}$ means "it is necessary at time t that p is true at time $t+1$. The other boolean connectives are defined as usual.

Example 2 *(Running example (Ctd.)).* Let our language contain:

- $\mathsf{Act} = \{food, equip, cook, nop\}$, where $food$ is the action "buy food", $equip$ is the action "buy cleaning equipment", and $cook$ is the action "cook", and nop is the special "no operation" action,[2]
- Prop consists of pre and $post$ statements with the actions in Act, such as $pre(food)$, $pre(food, equip)$, $pre(food, nop, food)$, $post(food)$, $post(nop)$.

Some examples of formulas in the language generated from Act and Prop are:

- $pre(food)_0 \wedge do(nop)_0 \wedge do(equip)_1$ (the precondition to buy food at time 0 is true, no action is performed at time 0, and Bobby buys cleaning equipment at time 1),
- $\Diamond_0(do(food)_0 \wedge \neg do(cook)_1)$ (it is possible at time 0 to buy food at time 0 and not to cook at time 1),

[2] In many of our examples it is useful include an action that does not do anything. In that case we use the special action nop and the formulas $pre(nop) \equiv \top$ and $post(nop) \equiv \top$.

- $\Diamond_0 do(food)_0 \wedge \Diamond_0 do(equip)_1 \wedge \neg\Diamond_0(do(food)_0 \wedge do(equip)_1)$ (it is possible to buy food at time 0 and it is possible to buy equipment at time 1, but it is not possible to do both),
- $pre(food, cook)_0$ (the precondition to buy food at time 0 and then cook at time 1 is true),
- $do(equip)_1$ (Bobby will buy cleaning equipment at time 1),
- $\Diamond_0\neg\Diamond_1 do(cook)_1$ (it is possible at time 0 that it is not possible at time 1 to cook),
- $\bigvee_{x\in \mathsf{Act}} pre(food, x, equip)_0$ (the precondition to buy food at time 0 and to buy equipment at time 2 is true, if a right action is performed at time 1).

Note that in this article, we use $pre(a)_t$ to denote the proposition that is the precondition of action a, but this is simply a naming convention. For instance, in the example above, we denote the precondition for action $food$ as $pre(food)$, while we explain in natural language that this means the agent has sufficient money to buy food. Thus, we could equally have written $hasEnoughMoneyForFood$ instead of $pre(food)$, but we chose to keep the former notation, to show the interplay between preconditions, actions, and postconditions in our examples more clearly. One may choose to define pre/postconditions as abbreviations of state propositions (possibly with time-indices), but since the internal structure of pre/postconditions is not the focus of our paper, we define them simply as primitive objects (first-class citizens).

The following definition collects all formulas up to some time t in a set $Past(t)$, which will turn out to be convenient when we axiomatize our logic. A formula of the form $do(a)_t$ will be semantically defined as a transition from t to $t+1$. Therefore it does not belong to the formulas true up to time t if it does not fall under the scope of a modality. We will make this more precise when we introduce the semantics in the next subsection.

Definition 2. $Past(t)$ is the set of all formulas from $\mathscr{L}$ generated by boolean combinations of $p_{t'}$, $pre(\overline{a})_{t'}$, $post(a)_{t'}$, $\Box_{t'}\varphi$, and $do(a)_{t'-1}$ where $t' \leq t$ and φ is some formula from $\mathscr{L}$.

Note that φ in the definition above can contain formulas indexed by time points greater than t. For instance, $do(a)_1 \in Past(2)$, $\Box_2\Diamond_5 pre(a)_6 \in Past(2)$, but $do(a)_3 \notin Past(1)$ and $do(a)_1 \vee p_1 \notin Past(1)$.

2.2. Semantics

The semantics of our logic is similar to CTL* [41], namely a tree structure containing nodes and edges connecting the nodes. A tree can equivalently be seen as an unfolded transition system, thereby representing all the possible runs through it. We choose to represent our semantics using trees because it simplifies the completeness proofs. See Reynolds [41] for an overview of different kinds of semantics and conceptual underpinnings.

With each natural number $i \in \mathbb{N}$ we associate a set of states S_i such that all these sets are disjoint. We then define the accessibility relation between states such that it generates an infinite, single tree.

Definition 3 *(Tree).* A tree is quadruple $T = (S, R, v, act)$ where

- $S = \bigcup_{n\in\mathbb{N}} S_n$ is a set of states, such that each S_t is the set of states at time t, $S_i \cap S_j = \emptyset$ for $i \neq j$;
- $R \subseteq \bigcup_{n\in\mathbb{N}} S_n \times S_{n+1}$ is an accessibility relation that is serial, linearly ordered in the past and connected (so S_0 is a singleton);
- $v : S \to 2^{\mathsf{Prop}}$ is a valuation function from states to sets of propositions;
- $act : R \to \mathsf{Act}$ is a function assigning actions to elements of the accessibility relation, such that actions are deterministic, i.e. if $act((s, s')) = act((s, s''))$, then $s' = s''$.

We evaluate formulas on a path in a tree. A path is a sequence of states in a tree, connected by the accessibility relation R.

Definition 4 *(Path).* Given a tree $T = (S, R, v, act)$, a *path* $\pi = (s_0, s_1, \ldots)$ in T is a sequence of states such that $(s_t, s_{t+1}) \in R$. We write π_t to refer to the t-th state of the path π. We use elements of the path as arguments for the valuation function and the action function:

- $v(\pi_t)$ are the propositions true on path π at time t;
- $act((\pi_t, \pi_{t+1}))$ is the next action on path π at time t. We abbreviate $act((\pi_t, \pi_{t+1}))$ with $act(\pi, t)$, since π_{t+1} is uniquely determined by the action.

We identify T with the set of paths in T, and we write $\pi \in T$ to denote that a path π exists in the tree T.

Intuitively, $v(\pi_t)$ are the propositions true at time t on path π, and $act(\pi, t)$ is the next action a on the path. We next define an equivalence relation $\sim_t$ on paths, which is used to give semantics to the modal operator.

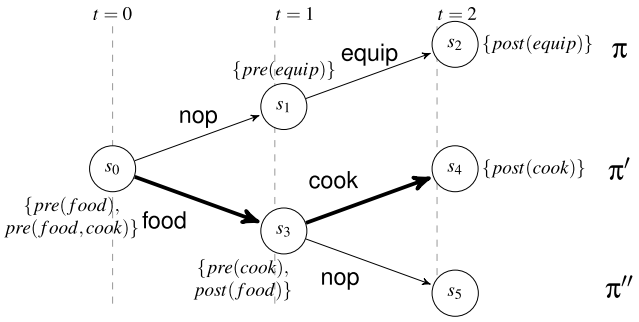


Fig. 2. Example PAL model (T, π') from $t = 0$ to $t = 2$.

Definition 5 *(Path equivalence).* Two paths π and π' are equivalent up to time t, denoted $\pi \sim_t \pi'$, if and only if they contain the same states up to and including time t, i.e.

$$\pi \sim_t \pi' \text{ iff } (\forall t' \leq t).(v(\pi_{t'}) = v(\pi'_{t'})) \text{ and}$$
$$(\forall t' < t).(act(\pi, t') = act(\pi', t')).$$

Formulas in PAL are evaluated on a path. Therefore, a model for a formula is pair consisting of a tree and a path in this tree. This, together with some additional constraints related to the pre- and post-conditions of actions, is our definition of a *model*.

Definition 6 *(Model).* A *model* is a pair (T, π) with $T = (S, R, v, act)$ such that for all $\pi \in T$ the following holds:

1. If $act(\pi, t) = a$, then $post(a) \in v(\pi_{t+1})$;
2. If $pre(a) \in v(\pi_t)$, then there is some π' in T with $\pi \sim_t \pi'$ and $act(\pi', t) = a$;
3. If $pre(a, \overline{b}) \in v(\pi_t)$, then there is some π' in T with $\pi \sim_t \pi'$, $act(\pi', t) = a$, and $pre(\overline{b}) \in v(\pi'_{t+1})$;
4. If $pre(\overline{a}, b) \in v(\pi_t)$, then $pre(\overline{a}) \in v(\pi_t)$.

We refer to models of PAL with $m_1, m_2, \ldots$, we refer to sets of models with $M_1, M_2, \ldots$, and we refer to the set of all models with $\mathbb{M}$.

Remark 1. In our semantics, preconditions are sufficient conditions for actions to be possible, but they are not necessary. Alternatively, one may strengthen this by changing the Condition 2 of Definition 6 from an "if" to an "if and only if". However, our choice is an implementation of Shoham's idea of "opportunistic planning": a planner may form intentions, even though at the moment of planning it may not be clear whether preconditions are true [46].

The conditions on models are there to formalize the consistency conditions from the introduction. Condition 1 is straightforward: we simply expect postconditions to hold in a state after an action has been executed. Condition 2 and 3 put a weaker requirement on the preconditions for actions: If the precondition holds, then there is *some* path in which the action is executed. This is part of the weaker requirement Shoham puts on preconditions. The opposite direction, stating that preconditions are necessary for executing actions, will be formalized with a coherence condition on beliefs and intentions in Section 3.2. Condition 4 of a model simply ensures that if the precondition of a sequence of action is true in a state, then the precondition for any subsequence by removing actions from the end of the sequence is also true in that state.

Example 3 *(Running example (Ctd.)).* Consider the partial PAL model (T, π') of the beliefs of Bobby the household robot from time 0 to time 2 in Fig. 2, where the thick path represents the actual path. We provide some examples of the conditions of our model (Definition 6):

- since $act(\pi, 1) = equip$, $post(equip) \in v(\pi_2)$ holds as well (Condition 1),
- since $pre(cook) \in v(\pi''_1)$, there is some path, namely π' with $\pi' \equiv_t \pi''$ and $act(\pi', 1) = cook$ (Condition 2),
- since $pre(food, cook) \in v(\pi_0)$, there exists some path, namely π' with $\pi \equiv_0 \pi', act(\pi', 0) = food$, and $pre(cook) \in v(\pi'_1)$ (Condition 3),
- since $pre(food, cook) \in v(\pi_0)$, $pre(food) \in v(\pi_0)$ holds as well (Condition 4).

We now provide the truth definitions. Recall that formulas are evaluated on a path as a whole, and not in a state.

Definition 7 *(Truth definitions).* Let $m = (T, \pi)$ be a model with $T = (S, R, v, act)$:
$T, \pi \models \chi_t$ iff $\chi \in v(\pi_t)$ with $\chi \in \mathsf{Prop}$
$T, \pi \models do(a)_t$ iff $act(\pi, t) = a$
$T, \pi \models \neg\varphi$ iff $T, \pi \not\models \varphi$
$T, \pi \models \varphi \wedge \varphi'$ iff $T, \pi \models \varphi$ and $T, \pi \models \varphi'$
$T, \pi \models \Box_t \varphi$ iff for all π' in T: if $\pi' \sim_t \pi$, then $T, \pi' \models \varphi$

The truth definitions state that propositions are simply evaluated using the valuation function v, but *do* statements are different. They are about state transitions, and therefore use the action function act. This is comparable to the distinction between state formulas and path formulas in CTL* (see the related work Section 6.3 for more details). Since we are evaluating formulas from a state, the modal operator $\Box_t$ is indexed with a time point t, and corresponds to the equivalence relations $\sim_t$.

Example 4 *(Running example (Ctd.)).* We provide some example of applications of the truth definition for the model in Fig. 2:
$T, \pi \models pre(food)_0 \wedge do(nop)_0 \wedge do(equip)_1$
$T, \pi \models \Diamond_0(do(food)_0 \wedge \neg do(cook)_1)$
$T, \pi \models \Diamond_0 do(food)_0 \wedge \Diamond_0 do(equip)_1 \wedge \neg\Diamond_0(do(food)_0 \wedge do(equip)_1)$
$T, \pi' \models pre(food, cook)_0$
$T, \pi' \not\models do(equip)_1$
$T, \pi'' \models \Diamond_0 \neg \Diamond_1 do(cook)_1$

Definition 8 *(Model of a formula).* We say that a model m is a model of a formula φ if $m \models \varphi$. We denote the set of all models of a formula φ by $Mod(\varphi)$, i.e.,

$$Mod(\varphi) = \{m \in \mathbb{M} \mid m \models \varphi\}.$$

We define the set of all models of a set of formulas Σ, as $Mod(\Sigma) = \{m \in \mathbb{M} \mid m \models \varphi \text{ for every } \varphi \in \Sigma\} = \bigcap_{\varphi \in \Sigma} Mod(\varphi)$.

Next we turn to the notions of validity, satisfiability, and semantic consequence. Valid formulas hold in every model, and satisfiable formulas hold in some model.

Definition 9 *(Validity, satisfiability, and semantic consequence).*

- φ is *valid*, i.e. $\models \varphi$ iff $Mod(\varphi) = \mathbb{M}$.
- φ is *satisfiable* iff $Mod(\varphi) \neq \emptyset$.
- φ is a *semantic consequence* of a set of formula Σ, i.e. $\Sigma \models \varphi$ iff $Mod(\Sigma) \subseteq Mod(\varphi)$.

2.3. Axiomatization

In this part we present the axiomatization of our logic, and we explain the most important axioms in turn.

Propositional tautologies **(PROP)**
$\Box_t(\varphi \to \varphi') \to (\Box_t \varphi \to \Box_t \varphi')$ **(K)**
$\Box_t \varphi \to \varphi$ **(T)**
$\Diamond_t \varphi \to \Box_t \Diamond_t \varphi$ **(5)**

Axioms PROP, K, T, and 5 together ensure our modal operator is an equivalence relation. This is simply the modal logic system KT5.

$\chi_t \to \Box_t \chi_t$, where $\chi \in \mathsf{Prop}$ **(A1)**
$\Diamond_t \chi_t \to \chi_t$, where $\chi \in \mathsf{Prop}$ **(A2)**

Axioms A1 states that if a proposition is true in a state on a path, then it is necessarily true at that time, i.e., it is true in all equivalent paths. The contraposition of Axiom A2 states the same for negated propositions. These axioms follow from the definition of the equivalence $\sim_t$ between paths: if two paths are equivalent up to time t, then the same propositions are true in time t as well.

$do(a)_t \to \Box_{t+1} do(a)_t$ **(A3)**
$\Diamond_{t+1} do(a)_t \to do(a)_t$ **(A4)**

Axioms A3 and A4 are similar to A1 and A2, but then for the case of actions. Recall that *do* statements are semantically represented as transitions between states (Definition 7). Therefore, the modal operator is indexed with the next time point $t+1$.

$\Box_t \varphi \rightarrow \Box_{t+1} \varphi$ **(A5)**

Axiom A5 is a result of the fact that for some path π the number of paths equivalent with π can only decrease as time moves forward. Therefore, if something is true on all paths equivalent up to time t, then it is necessarily true on all paths equivalent up to the next time moment $t+1$.

$\bigvee_{a \in \mathsf{Act}} do(a)_t$ **(A6)**
$do(a)_t \rightarrow \neg do(b)_t$, where $b \neq a$ **(A7)**

Axioms A6 and A7 together state that exactly one action is executed at every time moment.

$do(a)_t \rightarrow post(a)_{t+1}$ **(A8)**
$pre(a)_t \rightarrow \Diamond_t do(a)_t$ **(A9)**
$(pre(a, \overline{b})_t \wedge do(a)_t) \rightarrow pre(\overline{b})_{t+1}$ **(A10)**
$pre(\overline{a}, b)_t \rightarrow pre(\overline{a})_t$ **(A11)**

Axioms A8-A11 directly correspond to properties 1-4 of a model (Definition 6).

$(do(a)_t \wedge \varphi) \rightarrow \Box_t (do(a)_t \rightarrow \varphi)$ **(A12)**
where $\varphi \in Past(t+1)$

Axiom A12 ensures actions are deterministic. If something holds immediately after performing action a in time t (which is why $\varphi \in Past(t+1)$), then it necessarily holds after performing that action in time t. Note this formula does not hold without the restriction of φ to $Past(t+1)$, because formulas containing time points greater than $t+1$ may depend on actions performed after time t.

In addition to these axioms, PAL has two inference rules, a variant of Necessitation and Modus Ponens:

From φ, infer $\Box_0 \varphi$ **(NEC)**
From $\varphi, \varphi \rightarrow \varphi'$, infer φ' **(MP)**

Note that in NEC we can replace $\Box_0 \varphi$ with $\Box_t \varphi$, for any t, due to Axiom A5. In other words, the following variant of necessitation is a *derivable rule* of the logic:

From φ, infer $\Box_t \varphi$ **(NEC-t)**

Remark 2. Continuing our discussion of Remark 1, one may strengthen Axiom A9 as follows:

$pre(a)_t \leftrightarrow \Diamond_t do(a)_t$ (A9*).

We next formalize the notion of theorems and derivability.

Definition 10 *(Theorems in PAL).* A derivation of φ within PAL is a finite sequence $\varphi_1, \ldots, \varphi_m$ of formulas such that:

1. $\varphi_m = \varphi$;
2. every φ_i in the sequence is either
   (a) (an instance of) one of the axioms,
   (b) the result of the application of Necessitation or Modus Ponens to formulas in the sequence that appear before φ_i.

If there is such a derivation for φ we write $\vdash \varphi$, and we say φ is a *theorem of PAL*.

We define theorems and derivability separately because we restrict the application of the Necessitation rule to theorems only.

Definition 11 *(Derivability in PAL).* A derivation for a formula φ from a set of formulas Σ is a finite sequence $\varphi_1, \ldots, \varphi_m$ of formulas such that:

1. $\varphi_m = \varphi$;
2. every φ_i in the sequence is either a theorem, a member of Σ, or the result of the application of Modus Ponens to formulas in the sequence that appear before φ_i.

If there is such a derivation from Σ for φ we write $\Sigma \vdash \varphi$. We then also say that φ *is derivable from the premises* Σ.

Furthermore, a set of formulas Σ is *consistent* if we cannot derive a contradiction from it, i.e., $\Sigma \nvdash \bot$, and a set of formulas Σ is *maximally consistent* if it is consistent and every superset is inconsistent.

We denote by $Cn(\Sigma)$ the set of consequences of Σ, i.e.

$$Cn(\Sigma) = \{\varphi \mid \Sigma \vdash \varphi\}.$$

2.4. Soundness and completeness

In this section we prove the axiomatization of PAL is sound and strongly complete with respect to its semantics.

Theorem 1 *(Completeness theorem). The logic PAL is sound and strongly complete, i.e.* $\Sigma \vdash \varphi$ *iff* $\Sigma \models \varphi$.

We provide a proof sketch of the theorem. The full proofs of all the results in this article can be found in a separate technical report.[3]

Proof sketch. We prove the following formulation of completeness: each consistent set of formulas Σ has a model. We prove the Lindenbaum lemma, stating that each consistent set can be extended to a maximally consistent set Σ', i.e. Σ' is consistent and each proper superset of Σ' is inconsistent. In the first step we extend Σ to a maximally consistent set Σ^0.

Then for each t we define an equivalence relation $\equiv_t$ on maximally consistent sets in the following way:

$$\Sigma_1^* \equiv_t \Sigma_2^* \text{ iff } \Sigma_1^* \cap Past(t) = \Sigma_2^* \cap Past(t).$$

Let us denote the corresponding equivalence classes by $[\Sigma^*]_t$, which means $\{\overline{\Sigma}^* \mid \Sigma^* \equiv_t \overline{\Sigma}^*\}$.

In the second part of the proof, using the maximally consistent superset Σ^* of Σ (which exists by the Lindenbaum lemma), we define the tree $T_{\Sigma^*} = (S, R, v, act)$:

1. $S = \bigcup_{t\in\mathbb{N}} S_t$ where $S_t = \{[\overline{\Sigma}^*]_t \mid \overline{\Sigma}^* \equiv_t \Sigma^*\}$.
2. sRs' iff $(\exists \overline{\Sigma}^*, t \in \mathbb{N}).(s = [\overline{\Sigma}^*]_t \wedge s' = [\overline{\Sigma}^*]_{t+1})$.
3. $\chi \in v(s)$ iff $(\exists \overline{\Sigma}^*, t \in \mathbb{N}).(s = [\overline{\Sigma}^*]_t \wedge \chi_t \in \overline{\Sigma}^*)$.
4. $a = act((s, s'))$ iff $(\exists \overline{\Sigma}^*).(s = [\overline{\Sigma}^*]_t \wedge s' = [\overline{\Sigma}^*]_{t+1} \wedge do(a)_t \in \overline{\Sigma}^*)$.

Given a maximally consistent set (mcs) Σ^*, we construct a path $\pi_{\Sigma^*} = (s_0, s_1, \ldots)$ from it by letting $s_t = [\Sigma^*]_t$. So $\chi \in v([\Sigma^*]_t)$ iff $\chi_t \in \Sigma^*$ and $a = act(([\Sigma^*]_t, [\Sigma^*]_{t+1}))$ iff $do(a)_t \in T^*$.

If $\pi(\Sigma^*) = (s_0, s_1, \ldots)$, where $s_t = [\Sigma^*]_t$, then one can show that $(T_{\Sigma^*}, \pi(\Sigma^*))$ is a model. Finally, we prove that for each φ, $(T_{\Sigma^*}, \pi(\Sigma^*) \models \varphi$ iff $\varphi \in \Sigma^*)$, using induction on the complexity of φ. Consequently, $(T_{\Sigma^*}, \pi(\Sigma^*)) \models \Sigma$. □

Note that we can check satisfiability of any formula from $\mathscr{L}$ in finite time. Indeed, for every formula $\varphi \in \mathscr{L}$ there is a maximal time index t appearing in φ. By Definition 7, for checking if φ is satisfied in a model m it is enough to check the states and actions in the paths of m up to time $t + 1$. If we restrict the evaluation functions v to the finite set of propositions from Prop relevant for φ,[4] and since we have finitely many deterministic actions, there are only finitely many different ways to build a tree until a fixed time instance. Therefore, the number of those time-restricted trees which satisfy the four conditions of Definition 6 is finite as well. Thus, the satisfiability problem for the logic PAL is decidable.

3. Adding intentions

In the previous section we developed a logic for the belief database of Shoham's database perspective (Fig. 1). We did not yet take intentions into account, which is what we do in this section. Recall intentions are formalized as *discrete atomic action intentions* of the form (a, t). We focus on two main tasks: separating beliefs dependent on intentions (*weak beliefs*) from those that are not (*strong beliefs*), and formalizing a coherence condition on beliefs and intentions. These two tasks correspond to the two subsections of this section.

[3] See supplementary material.

[4] Technical details can be found in the supplementary technical report.

3.1. Separating strong and weak beliefs

The idea behind strong beliefs (the terminology due to Van der Hoek and Wooldridge [54]) is that they represent the agent's ideas about what is inevitable, no matter how it would act in the world. In our setting, a set of strong beliefs is a set of formulas starting either with $\Diamond_0$ or $\Box_0$, and all consequences that follow from it. First, we define a language for strong beliefs.

Definition 12 *(Strong belief).* The *set of all of strong beliefs* $\mathbb{SB}$ for $\mathscr{L}$ are generated by boolean combinations of $\Box_0\psi$, where ψ is a PAL formula. A *strong belief* is an element of $\mathbb{SB}$.

We next provide some examples of strong beliefs for our running example.

Example 5 *(Running example, Ctd.).* Some examples of strong belief formulas are:

- $\Diamond_0(do(food)_0 \wedge \neg do(cook)_1)$
- $\Diamond_0 do(food)_0 \wedge \Diamond_0 do(equip)_1 \wedge \neg\Diamond_0(do(food)_0 \wedge do(equip)_1)$
- $\Diamond_0\neg\Diamond_1 do(cook)_1$
- $\Box_0\Diamond_0 do(cook)_1$

Next we define a set of strong beliefs, which is generated from the set of all strong beliefs, and closed under consequence.

Definition 13 *(Set of strong beliefs).* A *set of strong beliefs* SB is the deductive closure of a subset of formulas from $\mathbb{SB}$, i.e. $SB = Cn(\Sigma)$ where $\Sigma \subseteq \mathbb{SB}$.

The following example shows that a set of strong beliefs may also contain formulas which are not in $\mathbb{SB}$, since they are closed under consequence.

Example 6 *(Set of strong beliefs).* Let $\Sigma = \{\neg\Diamond_0 p_3, \Box_0 q_2\} \subset \mathbb{SB}$, and let the set of strong beliefs $SB = Cn(\Sigma)$. From Axioms A1 and A2 we obtain $\neg p_3 \in SB$, as well as $q_2 \in SB$.

The reader may already have noted that, semantically, strong beliefs are independent of the specific path on which they are true. Indeed, strong beliefs are true in a tree rather than on a single path. Therefore, if a model (consisting of a tree and a path) is a model for a strong belief formula φ, then all possible models with the same tree are models of the strong belief formula φ. We make this idea precise in the following definition.

Definition 14 *(Set of models of strong beliefs (msb set)).* A *set of models of strong beliefs* $MSB \subseteq \mathbb{M}$ (i.e., an *msb set*) is a set of models such that $MSB = \{(T, \pi) : \pi \in T\}$. The set $\mathbb{MSB}$ contains all msb sets.

Definition 14 ensures that if some model (T, π) is in a set of models of a strong belief, then all other models (T, π') are also in this set. Note that it would also be possible to identify strong models with a tree T, but we have chosen not to implement this to keep the presentation concise.

The following proposition shows a direct correspondence between a set of strong beliefs and its models.

Proposition 1. *Given a set of strong beliefs SB, the set of models of SB is an msb set, i.e., $Mod(SB) \in \mathbb{MSB}$.*

We now explain the semantics of strong beliefs models with our running example.

Example 7 *(Running example (Ctd.)).* Consider the tree T of Fig. 2 and let $\overline{\pi} \in \{\pi, \pi', \pi''\}$. The following statements hold:

- $T, \overline{\pi} \models \Diamond_0(do(food)_0 \wedge \neg do(cook)_1)$
- $T, \overline{\pi} \models \Diamond_0 do(food)_0 \wedge \Diamond_0 do(equip)_1 \wedge \neg\Diamond_0(do(food)_0 \wedge do(equip)_1)$
- $T, \overline{\pi} \models \Diamond_0\neg\Diamond_1 do(cook)_1$
- $T, \overline{\pi} \models \Box_0\Diamond_0 do(cook)_1$

We obtain a belief-intention database by adding intentions to the strong beliefs. By intentions we assume action-time pairs, and an intention database is a set of intentions. We also add the constraint that at most one action is intended for a given time moment. We close the set of strong belief under consequence. Alternatively we can also have a (finite) set of strong beliefs, as in Hansson's base revision [22], but we follow the approach of Katsuno and Mendelzon.

Definition 15 *(Belief database, intention database, belief-intention database).* An *intention* (a, t) is a pair consisting of an action $a \in \mathsf{Act}$ and a time point t. $\mathbb{I} = \mathsf{Act} \times \mathbb{N}$ denotes the set of all intentions.

A *belief database* SB is a set of strong beliefs closed under consequence, i.e. $SB = Cn(SB)$. $\mathbb{BD}$ denotes the set of all belief databases.

An *intention database* $I = \{(a_1, t_1), (a_2, t_2), \ldots\}$ is a set of intentions such that no two intentions exist at the same time point, i.e. if $i \neq j$ then $t_i \neq t_j$. $\mathbb{ID} \subseteq 2^{Act \times \mathbb{N}}$ denotes the set of all intention databases.

A *belief-intention database* (SB, I) consists of a belief database and an intention database. $\mathbb{BI} = \mathbb{BD} \times \mathbb{ID}$ denotes the set of all belief-intention databases.

We define weak beliefs by adding intentions to the strong beliefs, and closing the result under consequence.

Definition 16 *(Weak beliefs).* Given a belief-intention database (SB, I), the weak beliefs are defined as follows:

$$WB(SB, I) = Cn(SB \cup \{do(a)_t \mid (a, t) \in I\}).$$

We provide an example for weak beliefs using our running example.

Example 8 *(Running example (Ctd.)).* Suppose the set SB contains strong beliefs describing the tree T of Fig. 2. Some of the formulas in SB are:

- $\Diamond_0(do(food)_0 \wedge do(cook)_1)$,
- $\Diamond_0 do(equip)_1$,
- $\Box_0 pre(food, cook)_0$,
- $\neg\Diamond_0(do(food)_0 \wedge do(equip)_1)$.

Let $I = \{(food, 0), (cook, 1)\}$. Some examples of weak beliefs $WB(SB, I)$ are:

- $do(food)_0 \wedge do(cook)_1$,
- $\neg do(equip)_1$,
- $post(food)_1 \wedge post(cook)_2$.

Note the model (T, π') from Fig. 2 is a model of $WB(SB, I)$.

Note the difference between Example 7 and Example 8. Strong beliefs are true in a tree, while weak beliefs *depend on a path*. In this way, weak beliefs are contingent on the action executed on the actual path. We can thus understand adding intentions to strong beliefs semantically by *choosing a set of paths* in a tree.

Remark 3. Note that since weak beliefs contain strong beliefs with intentions, and everything following from that, they also contain postconditions of actions. For instance, if $I = \{(a, t)\}$ and $SB = \emptyset$, then $post(a)_t \in WB(SB, I)$ (by Axiom A8). However, it does not mean that preconditions of intended actions are believed as well, i.e. $pre(a)_t \notin WB(SB, I)$. So an agent can believe it will execute its intentions, while it doesn't believe the preconditions hold (yet). This is why the implication in Axiom A9 is not a bidirectional implication (see also Remark 2).

3.2. Commitment: the coherence condition on beliefs and intentions

This paper is about commitment. The agent is committed to its intentions as long they are coherent with its beliefs. The coherent condition is that the agent believes it is possible to perform all intended action. We thus require that the joint preconditions of all intended actions not be disbelieved by the agent.

Definition 17 *(Coherence).* Given an intention database $I = \{(b_{t_1}, t_1), \ldots, (b_{t_n}, t_n)\}$[5] with $t_1 < \ldots < t_n$, let

$$Cohere(I) = \Diamond_0 \bigvee_{\substack{a_t \in Act : t \notin \{t_1, \ldots, t_n\} \\ a_t = b_t : t \in \{t_1, \ldots, t_n\}}} pre(a_{t_1}, a_{t_1+1}, \ldots, a_{t_n})_{t_1}. \tag{1}$$

- For a given belief-intention database (SB, I), we say that it is *coherent* iff SB is consistent with $Cohere(I)$, i.e., $SB \nvdash \neg Cohere(I)$.

[5] $t_1, \ldots, t_n$ is not necessarily a sequence of subsequent integers. For instance $t_1 = 2, t_2 = 5$. The disjunction below covers the remaining time indexes with all possible actions.

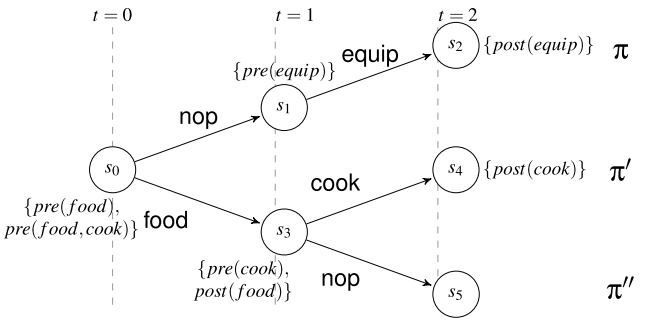


Fig. 3. The tree T of Fig. 2 reprinted.

- A pair (ψ, I) consisting of a strong belief formula $\psi \in \mathbb{SB}$ and an intention database I is coherent iff ψ is consistent with I, i.e. $\psi \nvdash \neg Cohere(I)$.[6]

Note we can define coherence semantically for a given msb set MSB (Definition 14) iff there exists some $m \in MSB$ with $m \models Cohere(I)$. We then obtain the correspondence that (SB, I) is coherent iff $(Mod(SB), I)$ is coherent using completeness trivially.

Let us explain this definition with a simple example.

Example 9. Let $\mathsf{Act} = \{a, b\}$ and $I = \{(a, 1), (b, 3)\}$. Then,[7]

$$Cohere(I) = \Diamond_0 \bigvee_{x \in \mathsf{Act}} pre(a, x, b)_1 = \Diamond_0 (pre(a, a, b)_1 \vee pre(a, b, b)_1).$$

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

In the next section we will consider the revision of belief-intention database. We will require that a belief-intention database is coherent after revision.

Remark 4. Consider our example of Bobby intending to buy food at time 0. As we pointed out, it is not actually necessary that Bobby believes it has sufficient money; only that it does not believe it does not have sufficient money. We can also ask: what can be Bobby's working assumptions about the future, upon adopting this intention? In so far as Bobby is committing himself to this action, we may assume that it will buy food at time 0. If we then consider the paths in our belief models on which this action is taken at time 0, the postconditions will hold along all of them. However, to allow that the preconditions may not yet be believed, we admit paths on which the preconditions do not hold. We only require that they hold on some path in the set, so that Bobby cannot stray too far from reality.

Indeed, this is arguably closer to how we reason about future actions. We often commit to actions without explicitly considering the path that will lead us there. Eventually this decision will have to be made, but there is nothing incoherent about glossing over these details at the current moment. Bobby should assume it will have bought food at time 1 and can continue making plans about what it will do with the food after this. But it should not assume the preconditions will hold until it has made further, specific plans for bringing them about. And at the current time, Bobby may not even bother worrying about it.

We now apply the coherence condition to our running example.

Example 10 *(Running example (Ctd.)).* Let (SB, I) be such that the strong beliefs are represented by the tree in Fig. 3.[8] We consider different choices for I:

- Let $I = \{(food, 0), (cook, 1)\}$. In this case, $(Mod(SB), I)$ is coherent, since there is some model $m \in Mod(SB)$ with $m \models Cohere(I)$, i.e. $m \models \Diamond_0 pre(food, equip)_0$. In fact, from $pre(food, cook) \in v(s_0)$, it follows that $T, \overline{m} \models pre(food, cook)_0$ holds for each model $(T, \overline{m})$. From the completeness theorem, it follows that (SB, I) is coherent as well.

[6] We will use this formulation in the next section when we represent a set of strong beliefs SB by a single formula ψ.

[7] Our construction of preconditions over action sequences may lead to a coherence condition involving a big disjunction. Alternatively, one may explicitly denote the time of each precondition, e.g. $pre(a, b)_{(t_1, t_2)}$. We chose the former since it is closer syntax of the other propositions.

[8] In other words, each tree in each model in the msb set SB is exactly the same up to time $t = 2$ as the tree in Fig. 3.

- Let $I = \{(food, 0), (equip, 1)\}$. In this case $(Mod(SB), I)$ is not coherent, since there is not $m \in Mod(SB)$ with $m \models \Diamond_0 pre(food, cook)_0$. Again by completeness we obtain that (SB, I) is not coherent either.

Next we show that a coherent belief-intention database implies joint consistency of beliefs and intentions.

Proposition 2. *Given some belief-intention database (SB, I), if (SB, I) is coherent, then $WB(SB, I)$ is consistent.*

Proof sketch. Using axioms A9, A10, and A12, for every $a_0, \ldots, a_m \in \mathsf{Act}$ and $\in \mathbb{N}$ one can show that $\{pre(a_0, \ldots, a_m)_t\} \vdash \Diamond_t(do(a_0)_t \wedge \Diamond_{t+1}(do(a_1)_{t+1} \wedge \Diamond_{t+2}(\ldots)))$. By taking the contrapositive of A5, we obtain the theorem of PAL logic

$$\vdash pre(a_0, \ldots, a_m)_t \rightarrow \Diamond_t \bigwedge_{p=0}^{m} do(a_p)_{t+p} \tag{2}$$

For an intention database $I = \{(b_{t_1}, t_1), \ldots, (b_{t_n}, t_n)\}$, with $t_1 < \ldots < t_n$, let us consider its coherence formula $Cohere(I)$ (formula (1) from Definition 17). If we apply the theorem (2) to the formulas under the scope of the disjunction in (1) (i.e., $pre(a_{t_1}, a_{t_1+1}, \ldots, a_{t_n})_{t_1}$), we obtain that $Cohere(I)$ implies

$$\Diamond_0 \bigvee_{\substack{a_t \in Act: t \notin \{t_1, \ldots, t_m\} \\ a_t = b_t: t \in \{t_1, \ldots, t_n\}}} \Diamond_{t_1}(do(a_{t_1})_{t_1} \wedge do(a_{t_1+1})_{t_1+1} \wedge \ldots \wedge do(a_{t_n})_{t_n})$$

Consequently, $Cohere(I)$ implies $\Diamond_0 \Diamond_{t_1} \bigwedge_{k=1}^{n} do(b_{t_k})_{t_k}$, and by A5 this implies $\Diamond_0 \bigwedge_{k=1}^{n} do(b_{t_k})_{t_k}$. Since $I = \{(b_{t_1}, t_1), \ldots, (b_{t_n}, t_n)\}$, the last formula can be rewritten as $\Diamond_0 \bigwedge_{(a,t)\in I} do(a)_t$. Therefore, if (B, I) is coherent, then $B \cup \{\Diamond_0 \bigwedge_{(a,t)\in I} do(a)_t\}$ is a consistent set. By the fact that B is a strong belief set, $B \cup \{\bigwedge_{(a,t)\in I} do(a)_t\}$ is consistent, i.e. $WB(B, I)$ is consistent. □

Note the reverse direction of Proposition 2 does not hold. We demonstrate this in the next example.

Example 11 *(Running example (Ctd.)).* Suppose (SB, I) is such that the strong beliefs are represented by the tree in Fig. 3 and that $I = \{(food, 0), (equip, 1)\}$. In this case the weak beliefs $WB(SB, I)$ are consistent, because the following holds:

$$WB(SB, I) \vdash do(food)_0 \wedge do(equip)_1,$$

no contradiction is derived from this. However, (SB, I) is not coherent, because there is no single path in the model in which all the preconditions of the intended actions hold.

4. Revision of belief and intention

In this section we turn to the *dynamic* part of our belief-intention databases by studying the revision of belief and intention. We provide and motivate a set of revision postulates on a belief-intention database (SB, I) in Section 4.2, and we prove our main representation theorem in Section 4.3. We discuss various examples in Section 4.4.

The challenge of obtaining our result is two-fold:

- When revising a belief database that is bounded up to some time t with a strong belief, we have to ensure that the resulting belief database is also bounded up to t;
- When revising a belief database we also have to ensure the new belief database remains a strong belief.

Our solution to is to bound both the syntax of PAL and the revision operator up to some time t in the first subsection. In the second subsection we do the same for the semantics.

4.1. Preliminaries: belief revision

The AGM postulates [1] formulate properties that should be satisfied by any (rational) revision operators defined on deductively closed sets of propositional formulas. Katsuno and Mendelzon [26] represent a belief set B as a propositional formula ψ such that $B = \{\varphi \mid \psi \vdash \varphi\}$. They define the following six postulates for revision on ψ and show that these are equivalent to the eight AGM postulates:

(R1) $\psi \circ \varphi$ implies φ
(R2) If $\psi \wedge \varphi$ is satisfiable, then $\psi \circ \varphi \equiv \psi \wedge \varphi$
(R3) If φ is satisfiable, then $\psi \circ \varphi$ is also satisfiable
(R4) If $\psi \equiv \psi'$ and $\varphi \equiv \varphi'$, then $\psi \circ \varphi \equiv \psi' \circ \varphi'$
(R5) $(\psi \circ \varphi) \wedge \varphi'$ implies $\psi \circ (\varphi \wedge \varphi')$
(R6) If $(\psi \circ \varphi) \wedge \varphi'$ is satisfiable, then $\psi \circ (\varphi \wedge \varphi')$ implies $(\psi \circ \varphi) \wedge \varphi'$

Given a set $\mathbb{J}$ of all interpretations over some propositional language, they define a faithful assignment as a function that assigns each ψ to a pre-order $\leq_\psi$ on models satisfying the following three conditions:

1. If $J, J' \in Mod(\psi)$, then $J <_\psi J'$ does not hold.
2. If $J \in Mod(\psi)$ and $J' \notin Mod(\psi)$, then $J <_\psi J'$ holds.
3. If $\psi \equiv \phi$, then $\leq_\psi = \leq_\phi$.

They show in a representation theorem that a revision operator $\circ$ satisfies postulates (R1)-(R6) iff there exists a faithful assignment that maps each formula ψ to a total preorder $\leq_\psi$ such that

$$Mod(\psi \circ \varphi) = \min(Mod(\varphi), \leq_\psi).$$

4.2. Revision postulates

Recall from Section 4.1 that we aim to prove a representation theorem comparable to that of Katsuno and Mendelzon [26]. Therefore, we follow their convention to fix a way of representing a belief set SB consisting of strong beliefs by a single strong belief formula ψ such that $SB = \{\varphi \mid \psi \vdash \varphi\}$. One of the main difficulties in this respect is that time in PAL is infinite in the future, so it is generally not possible to represent SB closed under consequence by a single formula ψ, since this may potentially lead to an infinite conjunction. Therefore, we cannot prove the Katsuno and Mendelzon representation theorem directly. In this section, we define a *bounded* revision function and we restrict the syntax of PAL up to a specific time point.

We first define some notation that we use in the rest of this section.

Notation.

- By slight abuse of terminology, a pair $(\psi, I) \in \mathbb{SB} \times \mathbb{ID}$ consisting of a strong belief formula ψ and an intention database I is also called a *belief-intention database*.
- For $\varphi, \psi \in \mathbb{SB}$, we write $\varphi \equiv \psi$ to denote that $\vdash \varphi \leftrightarrow \psi$.
- ε is the special "empty" intention.
- We denote $\mathbb{BI}, \mathbb{SB}, \mathbb{I}$, and $\mathbb{ID}$ bounded up to t with respectively $\mathbb{BI}^{|t}, \mathbb{SB}^{|t}, \mathbb{I}^{|t}$, and $\mathbb{ID}^{|t}$. However, if the restriction is clear from context, we may omit the superscript notation.

Our aim is to define a bounded revision function $*_t$ revising a belief-intention database (ψ, I) with a tuple (φ, i) consisting of a strong belief φ and an intention i, denoted $(\psi, I) *_t (\varphi, i)$. The bounded operator revises the formulas of the restricted language $\mathscr{L}^{|t}$, that represent all the relevant information for planning up to time t.

Definition 18 *(The language $\mathscr{L}^{|t}$).* The language $\mathscr{L}^{|t}$ consists of all formulas $\varphi \in \mathscr{L}$ such that if $p_{t'}, \Box_{t'}, do(a)_{t'}$ or $post(a)_{t'}$ occurs in φ, then $t' \leq t$. Furthermore, if $pre(a_0, \ldots, a_k)_{t'}$ occurs in φ, then $k + t' \leq t$.

For instance, $pre(a)_3, post(a)_3 \in \mathscr{L}^{|3}$, but $pre(a, b, c)_3 \notin \mathscr{L}^{|3}$.

First we define the restricted operator $\circ_t$ which revises strong beliefs up to given time instance t.

Definition 19 *(Strong belief revision function).* A *bounded strong belief revision function* is a function $\circ_t : \mathbb{BD} \times \mathbb{SB} \to \mathbb{BD}$, which maps a strong belief database ψ and a strong belief formula φ—all bounded up to t—to a strong belief database

$$\psi' = \psi \circ_t \varphi,$$

bounded up to t, and which satisfy the postulates of Katsuno and Mendelzon, R1–R6.

The revision operator above captures the intuition that strong beliefs are independent of intentions. This avoids *wishful thinking* meaning that the desire (or intention) for something to be true is used in place of/or as evidence for the truthfulness of the claim.

While revision of beliefs is independent of intentions, the revision of intentions should take beliefs into account as well, in order to ensure coherence. For instance, one can only accommodate a new intention if it is considered possible that these intentions can be achieved. This is formalized in the next revision operator.

Definition 20 *(Intention revision function).* An *intention revision function* $\otimes_t : \mathbb{BI} \times \mathbb{I} \to \mathbb{BI}$ maps a belief-intention database and an intention—all bounded up to t—to a belief-intention database bounded up to t such that

$$(\psi, I) \otimes_t i = (\psi, I'),$$

where the following postulates hold.

(P1) (ψ, I') is coherent.
(P2) If $(\psi, \{i\})$ is coherent, then $i \in I'$.
(P3) If $(\psi, I \cup \{i\})$ is coherent, then $I \cup \{i\} \subseteq I'$.
(P4) $I' \subseteq I \cup \{i\}$.
(P5) For all I'' with $I' \subset I'' \subseteq I \cup \{i\}$:$(\psi, I'')$ is not coherent.

Note that, by the definition, the revision of strong beliefs cannot be triggered by intention revision.

Postulate (P1) states that the outcome of a revision should be coherent. Postulate (P2) states that the new intention i take precedence over all other current intentions; if possible, it should be added, even if all current intentions have to be discarded. We can consider also not prioritized operators, which are not always successful. However, here we follow the standard approach in AGM theory change, which assumes that a check whether the belief base must be updated is done separately from the actual update.[9] Postulate (P3) and (P4) together state that if it is possible to simply add the intention, then this is the only change that is made. These two postulates are comparable to inclusion and vacuity of AGM. Finally, (P5) states that we do not discard intentions unnecessarily. This last postulate is a kind of maximality requirement, and is comparable to the *parsimony requirement* introduced by Grant et al. [21].

Up until now we have defined two revision operators separately: one for revising with a strong belief and one for revising with an intention. Recall that it is our aim to define "revising a belief-intention database with a belief/intention". We want to do this because the separate revision operators we defined up till now do not capture all interactions between beliefs and intentions. In particular, we would like to ensure that we capture the possibility for an agent to revise its intentions after having revised his strong beliefs, in order to restore coherence.

In order to do so we define a single revision function revising a belief-intention database by a pair (φ, i) in terms of the existing operators $\circ_t$ and $\otimes_t$.

Definition 21 *(Belief-intention revision function).* A *belief-intention revision function* is a function $*_t : \mathbb{BI} \times (\mathbb{SB} \times \mathbb{I}) \to \mathbb{BI}$ of the form

$$(\psi, I) *_t (\varphi, i) = (\psi \circ_t \varphi, I) \otimes_t i,$$

where $\circ_t$ is a strong belief revision function and $\otimes_t$ is an intention revision function.

In other words, the procedures runs as follows:

1. Revise strong beliefs using (R1)-(R6),
2. Revise intentions using (P1)-(P5), possibly revising weak beliefs as well.

Therefore, revising strong beliefs does not depend on which intentions an agent had, or which intention it revises with. However, revising intentions *does* have an effect on the weak beliefs (see the last paragraph of Example 13).

We will show how revision works in our logic with various examples at the end of this section, after we have proved the representation theorem.

4.3. Representation theorem

In this subsection we present the main technical result of this paper. We characterize all revision schemes satisfying (R1)-(R6), (P1)-(P5) in terms of minimal change with respect to an ordering among interpretations and a selection function accommodating new intentions while restoring coherence.

In the previous subsection we bounded various sets of formulas up to some time point t. We now do the same for models. We bound models up to t, which means that all the paths in the model are "cut off" at t.

Definition 22 *(t-bounded model).* Suppose some model $m = (T, \pi)$.

- A *t-bounded path* $\pi^{|t}$ is defined from a path π in T as $\pi^{|t} = (\pi'_0, \ldots, \pi'_t)$, where each π'_i contains the restriction of an evaluation of π_i to exactly those χ[10] such that $\chi_i \in \mathscr{L}^{|t}$.
- A *t-bounded model* $m^{|t}$ is the pair $(T^{|t}, \pi^{|t})$ where $T^{|t} = \{\pi_1^{|t} \mid \pi_1 \in T\}$.

We denote the set of all t-bounded models with $\mathbb{M}^{|t}$.

[9] Note that if we would drop postulate P2, then this would correspond to dropping item 2 of Definition 24.
[10] Recall from Definition 1 that we denote atomic propositions with χ.

Recall we defined a set of models of strong beliefs MSB as an *msb set* (Definition 14). A belief database SB consists of a set of strong beliefs, and we showed in Proposition 1 that the set of models of SB is an msb set, i.e. $Mod(SB) \in \mathbb{MSB}$, where $\mathbb{MSB}$ is the set containing all msb sets.

In order to represent revision semantically, we define a t-bounded version of msb sets as well.

Definition 23 *(t-bounded msb set).* Given an msb set MSB (Definition 14), the *t-bounded msb set* contains all t-bounded models of MSB, i.e.

$$MSB^{|t} = \{m^{|t} \mid m \in MSB\}.$$

Given an intention database I, we define a selection function γ_I^t that tries to accommodate a new intention based on strong beliefs. The selection function specifies preferences on which intention an agent would like to keep in the presence of the new beliefs.

Definition 24 *(Selection function).* Given an intention database I, a *selection function* $\gamma_I^t : \mathbb{MSB} \times \mathbb{I} \to \mathbb{ID}$ maps an msb set (Definition 14) and an intention to an updated intention database—all bounded up to t—such that if $\gamma_I^t(MSB^{|t}, i) = I'$, then:

1. $(MSB^{|t}, I')$ is coherent.
2. If $(MSB^{|t}, \{i\})$ is coherent, then $i \in I'$.
3. If $(MSB^{|t}, I \cup \{i\})$ is coherent, then $I \cup \{i\} \subseteq I'$.
4. $I' \subseteq I \cup \{i\}$.
5. For all I'' with $I' \subset I'' \subseteq I \cup \{i\}$:$(MSB^{|t}, I'')$ is not coherent.

The five conditions on the selection function are in direct correspondence with postulates P1–P5 of the intention revision function.

Remark 5. We will show in Corollary 1 below that it is possible to represent each set of strong beliefs SB (Definition 12) by a formula ψ such that $Cn(SB) = Cn(\psi)$. Using this corollary, we adapt the definition of a Katsuno and Mendelzon faithful assignment below.

Katsuno and Mendelzon [26] define a faithful assignment from a belief formula to a pre-order over models. Since we are also considering intentions, we extend this definition such that it also maps intentions databases to selection functions.

Definition 25 *(Faithful assignment).* A *faithful assignment* is a function that assigns to each strong belief formula $\psi \in \mathbb{SB}^{|t}$ a total pre-order $\leq_\psi^t$ over $\mathbb{M}$ and to each intention database $I \in \mathbb{D}^{|t}$ a selection function γ_I^t and satisfies the following conditions:

1. If $m_1, m_2 \in Mod(\psi)$, then $m_1 \leq_\psi^t m_2$ and $m_2 \leq_\psi^t m_1$.
2. If $m_1 \in Mod(\psi)$ and $m_2 \notin Mod(\psi)$, then $m_1 < m_2$.
3. If $\psi \equiv \phi$, then $\leq_\psi^t = \leq_\phi^t$.
4. If $T^{|t} = T_2^{|t}$, then $(T, \pi) \leq_\psi^t (T_2, \pi_2)$ and $(T_2, \pi_2) \leq_\psi^t (T, \pi)$.

Conditions 1 to 3 on the faithful assignment are the same as the conditions that Katsuno and Mendelzon put on a faithful assignment. Condition 4 ensures the two difficulties we pointed out in the beginning of this subsection are handled correctly:

- It ensures we do not distinguish between models in the total pre-order $\leq_\psi^t$ whose trees are the same up to time t. This is essentially what is represented in the revision function by bounding all input of the revision function $*_t$ up to t.
- Moreover, $\leq_\psi^t$ does not distinguish between models obtained by selecting two different paths from the same tree. In other words, it ensures that msb sets (sets of models of a strong belief) remain in the same ordering. This corresponds to the fact that we are using strong belief formulas in the revision, which do not distinguish between different paths in the same tree as well.

We are now ready to state our main theorem.

Theorem 2 *(Representation theorem). The function $*_t : \mathbb{BI} \times (\mathbb{SB} \times \mathbb{I}) \to \mathbb{BI}$ is a belief-intention revision operator iff there exists a faithful assignment that maps each ψ to a total pre-order $\leq_\psi^t$ and each I to a selection function γ_I^t such that if $(\psi, I) *_t (\varphi, i) = (\psi', I')$, then:*

1. $Mod(\psi') = \min(Mod(\varphi), \leq^t_\psi)$
2. $I' = \gamma^t_I(Mod(\psi'), i)$

We will use the remainder of this section to prove some results that we use for the proof of the representation theorem above. We first show the number of t-bounded models is finite.

Lemma 1. *For each $t \in \mathbb{N}$, $\mathbb{M}^{|t}$ is finite.*

Proof. Suppose some $t \in \mathbb{N}$. Since actions are deterministic and there are finitely many actions in our logic, each state has a finite number of successor states. Moreover, since there are finitely many propositions in $\mathscr{L}^{|t}$, the number of possible valuations of the states is finite as well. Therefore, the number of models in $\mathbb{M}^{|t}$ is finite. □

The following lemma obtains a correspondence between semantic consequence of two models equivalent up to t. The proof is by induction on the depth of the formula.

Lemma 2. *For each $\varphi \in \mathscr{L}^{|t}$ and models $m_1, m_2 \in \mathbb{M}$, if $m_1^{|t} = m_2^{|t}$, then $m_1 \models \varphi$ iff $m_2 \models \varphi$.*

Let $Ext(MSB^{|t})$ be the set of all possible extensions of a t-bounded msb set $MSB^{|t}$ to models, i.e.

$$Ext(MSB^{|t}) = \{m \in \mathbb{M} \mid m^{|t} \in MSB^{|t}\}.$$

We next show that we can represent $MSB^{|t}$ by a single strong belief formula using $Ext(MSB^{|t})$.

Lemma 3. *Given a t-bounded msb set $MSB^{|t}$, there exists a strong belief formula $form(MSB^{|t}) \in \mathbb{SB}$ such that $Mod(form(MSB^{|t})) = Ext(MSB^{|t})$.*

The following corollary shows we can represent a belief database consisting of strong beliefs up to some time t with a single formula.

Corollary 1. *Given a t-bounded set of strong beliefs $SB^{|t}$, there exists a strong belief formula $\psi \in \mathbb{SB}$ such that $SB^{|t} = \{\varphi \mid \psi \vdash \varphi\}$.*

4.4. Examples

In this section we discuss various examples of revision of beliefs and intentions in our framework. We start with the example from the introduction, in which Bobby the household robot can only buy food and cleaning equipment if it robs a bank.

Example 12 *(Running example (robbing a bank)).* Consider the following undesired situation. Suppose Bobby has a belief-intention database (ψ, I) and that $I = \{(food, 0)\}$, i.e. Bobby intends to buy food at time 0. Now suppose Bobby would revise by the pair $(\top, (equip, 2))$: it would like to buy cleaning equipment at time 2. This is coherent (despite Bobby's belief that it does not have enough money for both) thanks to an action *rob* of robbing a bank that it can perform at time 1. Suppose moreover that this is the only model where Bobby can perform both $(food, 0)$ and $(equip, 2)$. Then by postulate P2 we obtain that the resulting belief-intention database entails $do(rob)_1$, even if it would like to avoid robbing a bank by any means.

In other words, it seems that the weak beliefs of an agent may entail actions it did not intend to do, if those actions are the only means to carry out the intended actions. However, if Bobby would like to avoid robbing a bank by all means, then this should follow explicitly in its strong belief database, i.e., for all ℓ (up to the considered time t)

$$\psi \vdash \Box_0 \neg do(rob)_\ell.$$

Given that this is a strong belief, there is no possible future in which the agent believes it will rob a bank. A consequence of this is that the new intention $(equip, 2)$ will not be incorporated into the intention database, because the belief-intention database is not coherent after revision. If at some point later in time the agent would consider it possible to rob a bank, then it should revise its strong beliefs accordingly, but until that time the undesired action will not be weakly believed.

The above example shows that the agent can be coherent (it has the preconditions for all intended actions by postulate P2), but unaware of its future action to rob the bank (following from its weak beliefs). As shown in the example, this can be avoided by asserting this explicitly in the logic, however, due to belief revision this can change. If the agent really would never like to do this action, then this action should simply not be part of the available actions of the agent.

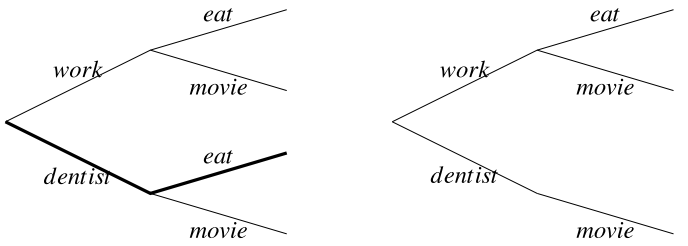


Fig. 4. Left: Partial model of strong beliefs ψ of agent (ψ, I) with $I = \{(dentist, 0), (eat, 1)\}$ (bold lines). Right: Revised strong beliefs of agent after learning it is not possible to eat (eat) after the dentist $(dentist)$.

Example 13 *(Running example (adding an intention)).* Suppose a belief-intention database (ψ, I) such that all models in $Mod(\psi)$ are the same as the partial model in Fig. 3 up to $t = 2$ and suppose that $I = \{(food, 0), (cook, 1)\}$. That is, Bobby has the intention to buy food at time 0 and then to cook at time 1. Suppose now Bobby changes its intention to buy cleaning equipment at time 1. Formally:

$$(\psi, I) *_1 (\top, (equip, 1)) = (\psi, I').$$

First note $(\psi, I \cup \{(equip, 1)\})$ is not coherent because no two intentions can occur at the same time moment. Moreover, since $(\psi, (equip, 1))$ is coherent, from (P2) and (P3) we obtain $(equip, 1) \in I'$. Furthermore, from (P4) we have that $I' \subseteq \{(food, 0), (cook, 1), (equip, 1)\}$. Finally, $(\psi, \{(food, 0), (equip, 1)\})$ is not coherent either, since the agent does not believe the preconditions of buying food and buying equipment are true along a single path. Combining this gives $I' = \{(equip, 1)\}$ as the only coherent outcome. Thus, Bobby no longer intends to buy food and to cook, but to buy cleaning equipment instead.

Note that, although the strong beliefs didn't change after revising with the new intention, the weak beliefs did change. For example, $post(food)_1 \in WB(\psi, I) \setminus WB(\psi, I')$ and $post(equip)_2 \in WB(\psi, I') \setminus WB(\psi, I)$.

The revision function $*_t$ takes a tuple (φ, i) as input, and Definition 21 ensures that revision of strong beliefs occurs prior to the revision of intentions. Therefore, it may seem plausible that revising with (φ, i) is the same as first revising with (φ, ε) and then with $(\top, i)$. In other words, the following postulate seems to follow:

$$\begin{aligned} &\text{If } (\psi, I) *_t (\varphi, i) = (\psi', I') \\ &\text{and } ((\psi, I) *_t (\varphi, \varepsilon)) *_t (\top, i) = (\psi'', I''), \\ &\quad \text{then } \psi' \equiv \psi'' \text{ and } I' = I''. \end{aligned} \tag{P*}$$

However, this property is not sound, and we show in the following example that adding the postulate would in fact conflict with the maximality postulate for intention revision (P5).

Example 14 *(Joint vs separate revision).* The operator $*_t$ can be instantiated to separate revisions of belief-intention databases as follows:

- Revising by $(\top, i)$ mirrors revising by no belief and an intention i, i.e.

$$(\psi, I) *_t (\top, i) = (\psi, I) \otimes_t i.$$

- Revising by (φ, ε) mirrors revising by a strong belief φ and no intention.

In spite of the fact that the operator $*_t$ revises beliefs prior to intentions, it is more expressive than its two instantiated operators combined, and cannot be defined as their composition. Indeed,

$$((\psi, I) *_t (\varphi, \varepsilon)) *_t (\top, i) \neq (\psi, I) *_t (\varphi, i).$$

This follows from the following example: Suppose some belief-intention database (ψ, I) with beliefs up to $t = 2$ corresponding to the model on the left of Fig. 4. It is possible to go to the dentist $(dentist)$ or to stay at work $(work)$, and after that to go eating $(eating)$ or go to the movies $(movies)$.

Before revision, the intentions are $I = \{(dentist, 0), (eat, 1)\}$ (left image of Fig. 4, intentions shown as bold lines).

Suppose now the beliefs are revised with the fact that it is not possible to go eating after going to the dentist (φ), i.e.,

$$\varphi \equiv \Box_0(post(dentist)_t \rightarrow \neg pre(eating)_t),$$

and with the intention to go to the movie at time 1 ($i = (movie, 1)$). The resulting strong beliefs after revising with φ are shown on the right of Fig. 4.

Let us analyze two ways of revising this information:

- Suppose (ψ, I) is revised with both the new belief and the new intention. That is,

  $$(\psi, I) *_2 (\varphi, i) = (\psi', I').$$

  Both $(\psi', \{(dentist, 0), (movie, 1)\})$ and $(\psi', \{(movie, 1)\})$ are coherent, so by the maximality postulate (P5), $I' = \{(dentist, 0), (movie, 1)\}$. Hence, the new intentions are to go to the dentist and then to go to the movie.
- Suppose beliefs are revised prior to intentions. That is,

  $$(\psi, I) *_2 (\varphi, \varepsilon) = (\psi', \overline{I})$$
  $$(\psi', \overline{I}) *_2 (\top, i) = (\psi', \overline{I}').$$

  Now, since $(\psi', \{(dentist, 0)\})$ and $(\psi', \{(eating, 1)\})$ are both coherent, we either have $\overline{I} = \{(dentist, 0)\}$ or $\overline{I} = \{(eating, 1)\}$. Suppose that $\overline{I} = \{(eating, 1)\}$. In that case, since $(\psi', \{(eating, 1), (movie, 1)\}$ is incoherent, we obtain $\overline{I}' = \{(movie, 1)\}$ by the postulates (P2) and (P4).

Thus, we see that revising separately allows a choice between the intention to go eating or to go to the dentist after revising beliefs. When choosing to go eating, the intention again has to be discarded because it is conflicting with the new intention to go to the movie. In joint revision, this is not the case since the choice between eating or the dentist can be made in light of the new incoming intention, and the maximal set can be chosen.

5. Iterated revision

Until now we only considered single-step revision of belief and intention. In this part we develop an account of iterated revision by following the approach of Darwiche and Pearl [15] developed for propositional logic. They observe that the AGM postulates are too permissive to enforce plausible iterated revision. In order to remedy this, they suggest the following changes:

- Instead of performing revision on a propositional formula, perform revision on an abstract object called an *epistemic state* Ψ, which contains the entire information needed for coherent reasoning.
- Postulate (R4) is weakened as follows:
  (R*4) If $\Psi = \Psi'$ and $\varphi \equiv \varphi'$, then $\Psi \circ \varphi \equiv \Psi' \circ \varphi'$.
- The following four desirable postulates are added for iterated revision:
  (C1) If $\varphi \models \varphi'$, then $(\Psi \circ \varphi') \circ \varphi \equiv \Psi \circ \varphi$.
  (C2) If $\varphi \models \neg\varphi'$, then $(\Psi \circ \varphi') \circ \varphi \equiv \Psi \circ \varphi$.
  (C3) If $\Psi \circ \varphi \models \varphi'$, then $(\Psi \circ \varphi') \circ \varphi \models \varphi'$.
  (C4) If $\Psi \circ \varphi \not\models \neg\varphi'$, then $(\Psi \circ \varphi') \circ \varphi \not\models \neg\varphi'$.

Postulate (C1) states that if two beliefs are added, the first is redundant if the second one is more specific. In other words, only revising with the second belief would obtain the same belief set. (C2) states that if the first belief is inconsistent with the second one, then revising with the first belief is unnecessary. (C3) ensures that a belief should be retained when revising with another belief implies it. Finally, (C4) states that if a belief φ is not contradicted after revising with another belief φ', then it should remain uncontradicted when the revising by φ' is preceded by φ.

We now define the revision operator for our logic, assuming that an epistemic state Ψ contains belief, which is represented by a strong belief formula denoted by $Bel(\Psi)$. As usual, we assume that that Ψ means $Bel(\Psi)$ whenever it is embedded in a formula, and that $Bel(\Psi)$ is thus a set of strong beliefs. For example, we say that (Ψ, I) is *coherent* if $(Bel(\Psi), I)$ is coherent. Recall that we omit the temporal restriction superscript from sets if it is clear from context.

Definition 26 *(Bounded epistemic state revision function).* A *bounded epistemic state revision function* is a function $\circ_t$ which maps an epistemic state and a strong belief formula—all bounded up to t—to a strong belief database bounded up to t, and which satisfy the postulates R1–R3, R*4, R5, R6 and C1–C4.

Similar to Section 4.2, we define a bounded revision function up to a fixed time t, assuming that the strong beliefs and intentions in an epistemic belief-intention databases are all bounded up to t.

Definition 27 *(Epistemic belief-intention database).* An *epistemic belief-intention database* (Ψ, I) consists of an epistemic state Ψ, and an intention database I. The set of all epistemic belief-intention databases bounded up to t is denoted by $\mathbb{EBI}^{\lfloor t}$, or simply $\mathbb{EBI}$ if t is clear from context.

Next we define the intention revision operator for an epistemic belief-intention database. We omit the formal definition of the intention revision function on epistemic belief-intention databases, but it is defined analogously to Definition 20. When we revise an epistemic belief-intention database with an intention, we leave the epistemic state Ψ unchanged and we update the intention database by applying Definition 20 to $Bel(\Psi)$.

Definition 28 *(Iterated revision of epistemic belief-intention databases).* An *epistemic belief-intention revision function* is a function $*_t : \mathbb{EBI} \times (\mathbb{SB} \times \mathbb{I}) \to \mathbb{EBI}$ of the form

$$(\Psi, I) *_t (\varphi, i) = (\Psi \circ_t \varphi, I) \otimes_t i,$$

where $\circ_t$ is an epistemic state revision function and $\otimes_t$ is an intention revision function.

When switching from belief revision on a belief state to belief revision on an epistemic state the definition of a faithful assignment should be adopted accordingly. We will do this now for our setting.

Definition 29 *(Faithful assignment for iterated revision).* A *t-bounded faithful assignment for iterated revision* is a function that assigns to each epistemic state Ψ a total pre-order $\leq^t_\Psi$ on all models, and to each intention database $I \in \mathbb{ID}^{|t}$ a selection function γ^t_I (Definition 24), such that the following conditions hold:

1. If $m_1, m_2 \in Mod(\Psi)$, then $m_1 \leq^t_\Psi m_2$ and $m_2 \leq^t_\Psi m_1$.
2. If $m_1 \in Mod(\Psi)$ and $m_2 \notin Mod(\Psi)$, then $m_1 <^t_\Psi m_2$.
3. $\Psi = \Phi$ only if $\leq_\Psi = \leq_\Phi$.
4. If $T^{|t} = T_2^{|t}$, then $(T, \pi) \leq^t_\Psi (T_2, \pi_2)$ and $(T_2, \pi_2) \leq^t_\Psi (T, \pi)$.

If $(\Psi, I) *_t (\varphi, i) = (\Psi', I')$, then:

5. If $m_1 \in Mod(\varphi)$ and $m_2 \in Mod(\varphi)$, then $m_1 \leq^t_\Psi m_2$ iff $m_1 \leq^t_{\Psi'} m_2$.
6. If $m_1 \notin Mod(\varphi)$ and $m_2 \notin Mod(\varphi)$, then $m_1 \leq^t_\Psi m_2$ iff $m_1 \leq^t_{\Psi'} m_2$.
7. If $m_1 \in Mod(\varphi)$, $m_2 \notin Mod(\varphi)$ and $m_1 <^t_\Psi m_2$, then $m_1 <^t_{\Psi'} m_2$.
8. If $m_1 \in Mod(\varphi)$, $m_2 \notin Mod(\varphi)$ and $m_1 \leq^t_\Psi m_2$, then $m_1 \leq^t_{\Psi'} m_2$.

The first four conditions are similar to the conditions on a faithful assignment for single-step revision (Definition 25), with the difference that $\leq^t_\psi$ is replaced with $\leq^t_\Psi$, and that Condition 3 has epistemic states in the antecedent instead of strong belief formulas. Conditions 5-8 are the semantic counterpart of (C1)-(C4).

Theorem 3 *(Representation theorem for iterated revision). A function $*_t : \mathbb{EBI} \times (\mathbb{SB} \times \mathbb{I}) \to \mathbb{EBI}$ is an epistemic belief-intention revision operator iff there exists a faithful assignment for iterated revision that maps each Ψ to a total pre-order $\leq^t_\Psi$ and each I to a selection function γ^t_I such that if $(\Psi, I) *_t (\varphi, i) = (\Psi', I')$, then:*

1. $Mod(\Psi') = \min(Mod(\varphi), \leq^t_\Psi)$
2. $I' = \gamma^t_I(Mod(Bel(\Psi')), i)$

We next provide a concrete epistemic belief-intention revision operator, thus showing *consistency* of all the postulates proposed for the epistemic state and intention revision operators.

Our operator is based on Spohn's *ordinal conditional ranking functions*, which can be seen as representations of epistemic states. For a given time t, our ranking function κ_t deals with epistemic states of our logic in a similar way as the operator based on Spohn's ranking function for propositional epistemic states, introduced in Darwiche and Pearl [15].

Definition 30 *(Spohn ranking function).* A *Spohn ranking function* $\kappa_t : \mathbb{M} \to \mathbb{N}$ assigns a rank to each model such that:

If $m_1 = (T_1, \pi_1)$ and $m_2 = (T_2, \pi_2)$ such that $T_1^{|t} = T_2^{|t}$, then $\kappa_t(m_1) = \kappa_t(m_2)$.

We extend the ranking to propositions as follows:

$$\kappa_t(\varphi) = \min_{m \models \varphi} \kappa_t(m).$$

Definition 31 *(Accepted propositions).* Given a ranking function κ_t, the *accepted propositions* $Bel(\kappa_t)$ are those for which the negation is implausible:

$$Bel(\kappa_t) = \{\varphi \mid \kappa(\neg\varphi) > 0\}.$$

It follows that the models of these propositions are those which have rank 0:

$$Mod(Bel(\kappa_t)) = \{m \mid \kappa_t(m) = 0\}.$$

The fact that $Bel(\kappa_t)$ is a set of strong beliefs follows from the condition we put on the Spohn ranking function.

Now we define our revision operator. For the epistemic part we follow Darwiche and Pearl [15], while for intention revision we give prefer intentions that occur sooner rather than later.

Definition 32 *(Extended Spohn-based revision operator).* The *Extended Spohn-based revision operator* $\bullet_t$ is defined as follows: $(\kappa_t, I) \bullet_t (\varphi, i) = (\kappa_t', I')$, where κ_t' and I' are such that[11]

$$\kappa_t'(m) = \begin{cases} \kappa_t(m) - \kappa_t(\varphi), & \text{if } m \models \varphi; \\ \kappa_t(m) + 1, & \text{if } m \models \neg\varphi, \end{cases}$$

and I' is defined in the following iterative way. If n is the maximal time instance that occurs in I, we define the sets of intentions

1. $I_{-1} = \{i\}$, if $(Bel(\kappa_t'), \{i\})$ is coherent, otherwise $I_{-1} = \emptyset$;
2. for every $\ell \in \{0, 1, \ldots, n\}$
   (a) if there is no a such that $(a, \ell) \in I$, $I_\ell = I_{\ell-1}$;
   (b) otherwise, if a is the unique action such that $(a, \ell) \in I$, then $I_\ell = I_{\ell-1} \cup \{(a, \ell)\}$, if $(Bel(\kappa_t'), I_{\ell-1} \cup \{(a, \ell)\})$ is coherent, otherwise $I_\ell = I_{\ell-1}$.
3. $I' = I_n$.

We now pose our theorem.

Theorem 4. *The function $\bullet_t$ is an epistemic belief-intention revision operator.*

The Extended Spohn-based revision operator is an example of an operator satisfying the epistemic belief-intention revision function for iterated revision (Definition 28) and puts additional constraints on intentions. It prioritizes the new intention i, and after that it iteratively attempts to increase the set of new intentions by adding them in temporal order, starting with the most recent ones. We next give an example of how this may work in practice.

Example 15 *(Running example (extended Spohn-based revision)).* Suppose that Bobby the household robot has an epistemic belief-intention database EBI with intentions $I = \{(food, 0), (equip, 1)\}$, and suppose its epistemic state is formalized as:

- $\kappa_t(m) = 0$, for every $m \in Mod(\varphi_0)$, where

  $$\varphi_0 = \Box_0 pre(food, equip)_0$$

  ("Bobby has precondition to buy both food and cleaning equipment")
- $\kappa_t(m) = 1$, for every $m \in Mod(\varphi_1)$, where

  $$\varphi_1 = \Box_0 \neg pre(food, equip)_0 \wedge \Box_0 pre(food)_0 \wedge \Box_0 pre(equip)_1$$

  ("Bobby can carry out any of two intended actions, but not both of them")
- $\kappa_t(m) = 3$, for every $m \in Mod(\varphi_3)$, where

  $$\varphi_3 = \Box_0 \neg pre(food)_0 \wedge \Box_0 \neg pre(equip)_1$$

  ("Bobby cannot carry out any of two intended actions")
- $\kappa_t(m) = 2$, for all other models, i.e., for $m \in Mod(\varphi_2)$, where $\varphi_2 = \neg\varphi_0 \wedge \neg\varphi_1 \wedge \neg\varphi_3$ (note that we can read φ_2 as "Bobby can buy only the cheaper of the two items")

Thus, Bobby's beliefs are represented by $Bel(\kappa_t)$ (Definition 31), which include that it believes it is possible to carry out both intended actions, so the agent is coherent. In addition, the epistemic state specifies a preference ordering over models as well, which is specified by the ranking κ. It assumes that "Bobby cannot carry out any of two intended actions" is the least plausible proposition.

Suppose next that Bobby revises its epistemic belief-intention database with $(\Box_0 \neg pre(food, equip)_0, \varepsilon)$. The ranking function selects the minimal (i.e., the most plausible) model in which the revised formula is consistent, which has ranking 1.

[11] We show in the technical report that κ_t' is a well defined Spohn ranking function.

After revision, Bobby believes it is able to buy food or buy cleaning equipment, i.e., any of those two, but not both. The new ranking κ_t' is as follows:

- $\kappa_t'(m) = 0$, for every $m \in Mod(\varphi_1)$
- $\kappa_t'(m) = 1$, for every $m \in Mod(\varphi_0 \vee \varphi_2)$
- $\kappa_t'(m) = 2$, for every $m \in Mod(\varphi_3)$

The agent can then use κ_t' for future revision. Using the intention revision operation of Definition 32, it will keep the intention that it intends to carry out first, which means $I' = \{(food, 0)\}$, and the intention to buy cleaning equipment is dropped.

6. Related work

We compare our work to the logic of intention, temporal logic, collective intentions, and our own previous work.

6.1. Logic of intention

We first give an overview of the literature on the logic of intention, before making a comparison with the theory introduced in this paper. In the overview, we discuss the philosophy of mind, the logic of intention of Cohen and Levesque, and other work.

6.1.1. Philosophy of mind

Before the early 1990s most work on intention was done by philosophers. The dominant view prior to the 1980s was that fundamental mental states included belief-like attitudes such as belief and knowledge on the one hand, and desire-like attitudes such as desires and preferences on the other [16]. These capture the "two directions of fit" between world and mind [43]. It was generally believed that intentions could be reduced to these more basic mental states. For instance, an intention can be thought of as a complex kind of belief: an intention to φ at some time t might be a belief that the agent will φ at t, perhaps by virtue of this very belief [24]. This *belief-desire* model of the mind was also common in decision theory in the spirit of Savage [42], as well as in artificial intelligence. The basic idea underlying all this is that rational action amounts to what leads to the most *desirable* outcome, given the agent's *beliefs* about the world and *preferences* about how it ought to be.

In the mid 1980s Michael Bratman wrote the book "Intentions, plans, and practical reason" [8], which turned out to be very influential in both philosophy and in artificial intelligence. In the book, he argued convincingly that intentions cannot be reduced to any "more basic" mental states, but that intentions must be taken as basic themselves. While his argument is quite complex for non-philosophers, the fundamental ideas are very simple. Intentions are subject to their own set of norms, which cannot be reduced to those for beliefs or desires. For instance, there are norms for intention consistency that do not apply to desires: if someone knows that φ and ψ are mutually inconsistent, it is irrational to intend φ and to ψ simultaneously, but it would not be irrational to desire both φ and ψ.

Another example of a norm of intentions, and one which has been very influential in computer science, is the idea that intentions constrain further practical reasoning in a characteristic way. Bratman (and others) have argued that if someone intends to φ, that gives her a *pro tanto* reason to φ, even in the face of evidence that she ought not φ. If we think of intentions as making up a plan, then reconsidering an intention potentially requires revising an entire plan, which can be computationally expensive and problematic in real time. A consequence of this *relative resistance to revision* is that intentions can be very useful for resource-bounded agents who have to select appropriate actions at different times. Both by adopting appropriate *policies* that can apply at different times to similar decision problems, and by devising complex plans which can easily be followed when computation time is limited, an agent with intentions is able to avoid performing complex computations every time a new decision problem arises. Observations such as those above are common knowledge in artificial intelligence by now, but it is important to realize there is a well explored philosophical foundation underlying them.

6.1.2. Cohen and Levesque

Inspired by the philosophical literature, research in artificial intelligence in the 1990s began to explore intentions. One of the first attempts of what was later known as the *belief-desire-intention architecture* was in a paper by Bratman et al. [10], which systematized many aspects of the view laid out in Bratman's book [8], including flowcharts and some suggestions for implementations. Most importantly, it served as a call for other researchers to investigate these matters more systematically and formally.

One of the most well-known and first attempts of a formalization of intention revision was by Cohen and Levesque [13]. They developed a formal logical language, including modal operators for mental state such as "belief" and "having a goal", for temporal expressions, and for descriptions of events such as "A happens" or "x did action a". The main point of the paper is to define a useful notion of intention. To this end they first define a P-GOAL, or *persistent goal*, which is a goal p

that one believes now not to hold, and which will cease to be a goal as soon as either the agent believes p will never hold, or the agent comes to believe p is true. Intending to take an action a is then defined in terms of a P-GOAL for agent x [13, p. 245]:

$$\text{INTEND}(x, a) := (\text{P-GOAL}x[\text{DONE}x(\text{BEL}x(\text{HAPPENS}a))?; a]).$$

Unpacking this, agent x intends to a just in case x has a persistent goal to ensure that x will believe a will be carried out, up until a is in fact carried out. They also define "intending to bring about some state of affairs", and show their approach solves the "Little Nell" problem (see Section 6.2.3) and avoid the "Dentist Problem" [8].

While Cohen and Levesque's approach is much cited, it is fair to say that it is rather complicated. Some early criticisms of technical details can be found in Singh [49]. In the textbook by Shoham and Leyton-Brown [48] the approach is called "the road to hell". Due to the complexity of the logic, mathematical properties such as axiomatizability, decidability and complexity of fragments were never investigated. None of the BDI logics that were introduced subsequently adapted Cohen and Levesque's four steps definition of intention and instead considered intentions to be primitive. Moreover, while Cohen and Levesque's provide some criteria for the abandonment of intentions through the notion of rational balance (forbidding to intend something that is true or believed to be impossible to achieve), it does not further analyze the 'other reasons' for which a persistent goal is abandoned. More details on these critiques can be found in a recent article by Herzig et al. [23].

Some of Cohen and Levesque's shortcomings were corrected to an extent in subsequent work. For instance, Rao and Georgeff [39] provide an approach for an alternative, and arguably simpler, formalism based on CTL.

6.1.3. Other related work on intention revision

There exists previous work on intention dynamics, that is, how intentions change over time. van der Hoek et al. [53] (see also Wooldridge [62] and Wooldridge and Jennings [63]) explore a system similar to Rao and Georgeff's with representations of beliefs, desires, and intentions separately, including a specific module for practical reasoning; they propose how to revise intentions together with beliefs. Their treatment is mainly syntactical, and they do not attempt to obtain a correspondence between postulates for revision and a pre-order over model, which is what we do here. Grant et al. [21] continue this line of work, offering postulates for intention revision (similar to our contribution). They develop AGM-style postulates for belief, intention, and goal revision. They provide a detailed analysis and propose different reconsideration strategies, but restrict themselves to a syntactic analysis as well. Lorini and Herzig [30] introduce a logic of intention with modal operators for attempt, tacking the question of when an agent's intentions translate into an actual attempted action. Finally, Shapiro et al. [45] also explore intention revision, working in a setting with complex hierarchical plans.

6.2. Comparison: the interplay between intention, time, and belief

To compare our theory with the one of Cohen and Levesque, we reconsider the running example. You instruct your future household robot Bobby to buy groceries for you in the afternoon. Bobby therefore believes to be in the supermarket in the afternoon, but this belief is based on the assumption that it has sufficient money to buy a bus ticket. Moreover, you instruct Bobby to clean the windows coming weekend, and Bobby's belief that it will clean the windows is based on the assumption that it has all the necessary cleaning products on time, and that it is not raining at that day. These assumptions are the preconditions of the actions Bobby intends to perform. When Bobby adopts intentions, it may form new beliefs based on these intentions. For instance, after Bobby adopts the intention to buy groceries in the afternoon, it is able to adopt new intentions based on this belief, for instance to cook food in the evening. Such intention-based beliefs are called *weak beliefs* [53,25]. In contrast, unconditional beliefs, beliefs that are no dependent on the predicted success of intentions, are called *strong beliefs*.

Our theory can explain how intentions and weak and strong beliefs change over time. Yesterday Bobby believed that it will rain tomorrow and be sunny the day after, but today it believes that tomorrow the sun may possibly shine and it will be cloudy the day after. Likewise, yesterday Bobby intended to buy groceries tomorrow, but now it intends to do laundry tomorrow. We say there is internal dynamics (rain today and sun tomorrow) and external dynamics (Bobby's beliefs today and Bobby's beliefs tomorrow, or Bobby's intentions today and Bobby's intentions tomorrow).

6.2.1. Assumptions

The first improvement is the formalization of the assumptions of the intentions, i.e. the preconditions of the intended actions. From the intention to take the train to Paris it may seem already quite risky to derive the belief that there still are tickets for the train, and from my intention to present a paper at IJCAI next year, it is surely too strong to infer a belief that I will finish the paper in time and that the paper will be accepted. Such a belief would be based on irrational wishful thinking.

However, we cannot ignore the assumptions either. Suppose I would have believed that there are no more tickets for the train to Paris tomorrow, then it would surely have been irrational to adopt the intention to go to Paris by train tomorrow. On the contrary, suppose I learn now that there are no more tickets for the train tomorrow, then I would drop my intention

to go to Paris. So from the intention to go to Paris we can infer that I do not believe that there are no more tickets. Note the essential difference between believing that there are still tickets, and the absence of a belief that there are no more tickets. This difference is represented very precisely in modal logic by the two formulas $SB_0(tickets)$ and $\neg SB_0\neg(tickets)$. The former implies the latter, but not vice versa.

If we represent a generic intention at time t for α by $I_t\alpha$, and the precondition of α by $pre(\alpha)$, then we may impose the following constraint on our modal logic:

$I_t\alpha$ implies $\neg B_t\neg pre(\alpha)$.

The assumptions may be seen as a kind of presupposition as it is studied in linguistics: the sentence "The king of France is bold" presupposes that France is a kingdom. Likewise, the intention to go to Paris presupposes that there still are such tickets. It is well known that it is challenging to find a suitable definition of presupposition. For example, if we derive the weak intention-based belief that we will be in Paris tomorrow, can we deduce also a weak intention-based belief that there are still tickets? The fact that we will be in Paris implies that we were able to obtain a ticket.

The absence of the belief that there are still tickets may be represented also using negation as failure in logic programming [12], because we cannot derive the belief that there are no more tickets. Moreover, it can also be represented by a justification in default logic [40], or more generally by a consistency check.

Consider the situation where I have the intention to give a seminar in Paris tomorrow, and in addition that I have the intention to give a seminar in London tomorrow. The preconditions of these intended actions are that I am in Paris and I am in London tomorrow, and clearly this is not possible. However, it is possible that I am in Paris tomorrow, and it is possible that I am in London tomorrow. It is just not possible that I am both in Paris and in London.

In other words, the logic has to derive that it is possible for all intended actions to be performed, and thus that for all intended actions, the preconditions hold. We introduce a new name for this happy situation where all intended actions can be performed. We say that a set of intentions and beliefs is *coherent* if it is possible that all actions can be performed and all preconditions hold.

6.2.2. What if statements

One of the contributions in the formalization of the interplay between intention, time, and belief is to model what-if statements: what will happen with my intentions and beliefs if I learn that the tickets for the train tomorrow are sold out? What happens if I learn that my paper for IJCAI is rejected? What happens if I adopt a new intention to go to a birthday party in Amsterdam tomorrow?

Our contribution adopts the AGM Alchourrón et al. [1] framework for theory change. To be able to use this framework, we first need to make some additional restrictions to our language of beliefs and intentions, and we need to define the operators we consider together with their postulates. Then we characterize the set of operators satisfying the postulates using a representation theorem.

Since we do not have a logical language with explicit belief operators, we cannot represent the coherence of a belief-intention base by a formula "it is not believed that ...". Instead, represent the coherence of a belief-intention base in a different way. We use an idea similar to negation as failure: the belief intention base is consistent with the precondition of the (sequence of) intended actions.

6.2.3. The "Little Nell" problem: not giving up too soon

McDermott [31] discusses the following difficulty with a naive planning system:

> Say a problem solver is confronted with the classic situation of a heroine, called Nell, having been tied to the tracks while a train approaches. The problem solver, called Dudley, knows that "If Nell is going to be mashed, I must remove her from the tracks." (He probably knows a more general rule, but let that pass.) When Dudley deduces that he must do something, he looks for, and eventually executes, a plan for doing it. This will involve finding out where Nell is, and making a navigation plan to get to her location. Assume that he knows where she is, and he is not too far away; then the fact that the plan will be carried out will be added to Dudley's world model. Dudley must have some kind of database-consistency maintainer to make sure that the plan is deleted if it is no longer necessary. Unfortunately, as soon as an apparently successful plan is added to the world model, the consistency maintainer will notice that "Nell is going to be mashed" is no longer true. But that removes any justification for the plan, so it goes too. But that means "Nell is going to be mashed" is no longer contradictory, so it comes back in. And so forth.

The agent continuously plans to save Nell, and abandons its plan because it believes it will be successful. If we view this problem from the database perspective, we require some way to separate beliefs about plans from beliefs that are not about plans. Otherwise, the planner may believe certain facts hold and adopt the plans accordingly, while these facts may be dependent on performing the plan.

This problem has been considered extensively in the literature (cf. [13,46,25,54]). Our solution is to separate strong beliefs from weak beliefs. The beliefs in the belief database of the agent (those that will be used by the planner) are strong beliefs, and the weak beliefs are computed by adding intentions to the strong beliefs, and everything following from that.

6.3. Temporal logic

Our theory distinguishes internal and external dynamics. The temporal references of the facts (e.g., $rain_1$, sun_2) represent internal dynamics, and the temporal references on the modal operators (e.g., I_{-1}, WB_1) represent external dynamics. Internal and external dynamics can also be represented by temporal modal operators such as next and until in temporal logics, as discussed in this section. In this article we instead use explicit time indexes, but in the future work in Section 7 we discuss how this can be extended to implicit time in temporal logics. The core ideas and results of this paper do not depend on this choice.

Time in temporal logics can be defined in an *implicit* or *explicit* manner. A time model is implicit when the meaning of formulas depends on the evaluation time, and this is left implicit in the formula. Standard LTL and CTL define time implicitly. For instance, $\Box\Phi$ means that $\forall t \in [T_0, \infty].\Phi(t)$, where T_0 is the evaluation time (the so-called current time instant). A standard way of introducing real time into the syntax of temporal languages constrains the temporal operators with time intervals [19,27,4,3]. In order to model such time intervals, timed automata may be used. These automata model the behavior of time-critical systems. A timed automaton is in fact a program graph that is equipped with a finite set of real-valued clock variables, called *clocks* for short [2]. Timed CTL (TCTL, for short) is a real-time variant of CTL aimed to express properties of timed automata. In TCTL, the until modality is equipped with a time interval such that $\Phi U^J \Psi$ asserts that a Ψ-state is reached within $t \in J$ time units while only visiting Φ-states before reaching the Ψ-state. The formula $\exists\Box^J\Phi$ asserts that there exists a path for which during the interval J, Φ holds; $\forall\Box^J\Phi$ requires this to hold for all paths [6]. While such logics allow one to express timed constraints on the modalities in TCTL, there is no way to refer explicitly to the states at which a certain formula holds.

When time is explicit, the language represents the time through a variable. For example, in the following formula an explicit model of time is used:

$$\forall t.\Box(E \wedge T = t) \rightarrow \Diamond(A \wedge T - t < 10)$$

where E is an event [7]. This is for instance formalized by Ostroff [36] when solving control problems using real-time temporal logic (RTTL).

The logic of strategic abilities ATL* (Alternative-Time Temporal Logic), introduced and studied by Alur et al. [5], is a logical system, suitable for specifying and verifying qualitative objectives of players and coalitions in concurrent game models. Formally, ATL* is a multi-agent extension of the branching time logic CTL* with *strategic path quantifiers* $\langle\langle C\rangle\rangle$ indexed with coalitions C of players. Bulling and Goranko [11] propose a quantitative extension of ATL*, in which it is possible to express temporal constraints as well. For instance, the expression $\phi \wedge x = t$ denotes that ϕ will be true after t transitions, where each transition adds 1 to x.

Many logical systems have been developed for reasoning about the pre and postconditions of actions with explicit time points, such as the Event Calculus [35], Temporal Action Logics [28], extensions to the Fluent Calculus [51], and extensions to the Situation Calculus [37] (see [38, Ch. 2] for an overview). Our logic is considerably more simple, but the reason for this is because of the type of revision we characterized in this article. Although there are a number of correspondences between AGM postulates and some of the approaches above, none of them prove representations theorems linking revision to a total pre-order on models.

As we mentioned, the structure of our models is full branching time structure of CTL*. In addition, we have the actions attached to the elements of accessibility relation R. Since our formulas are different than those of CTL*, it should be noted that the part of PAL built on propositional letters, modalities and Boolean connectives, is embedded in the fragment of CTL* with only $\bigcirc$ (next) and A (universal path quantifier) operators. This is because any formula of the form $\Box_t\phi$ can be written as the formula $\bigcirc^t A\psi$ in CTL*, where $\bigcirc^t$ stands for t occurrences of the $\bigcirc$ operator and ψ is the translation of ϕ, which takes into account that t-th moment becomes the actual moment and in which $\Box_t$ operators are translated likewise. For example, the formula $p_1 \wedge \Box_2((p \vee q)_3 \wedge q_4)$ would be translated to $\bigcirc p \wedge \bigcirc^2 A(\bigcirc(p \vee q) \wedge \bigcirc^2 q)$.[12]

As presented in Meier et al. [32], the satisfiability problem for this fragment of CTL* is known to be PSPACE-complete. It would be interested to see if those results can be modified for PAL and how the addition of actions in our logic influence its complexity.

6.4. Collective intentions

An important issue in the philosophical literature on collective intentionality is the question of whether collective intentions can be reduced to individual intentions (i.e., *reductionist theories*), or whether collective intentions are first-class citizens and cannot be reduced.

Bratman [9] offers a clear example of reductionist theory of collective intention as it decomposes the concept of collective intention in terms of more primitive concepts such as the concepts of individual intention, belief and common belief. He defines "shared cooperative activity" using the following three characteristics:

[12] Note that, some formulas, like $\Box_2 p_1$ cannot be translated directly, but the presence of the K-axiom and A1 and A5 overcome that problem, since they imply that $\Box_t\chi$ is equivalent to χ if $\chi \in Past(t)$. Thus, $\Box_2 p_1$ can be first transformed to an equivalent formula p_1 in our logic, which allows translation to a CTL* formula $\bigcirc p$.

1. Each participant must be mutually responsive to the intentions of others;
2. The participants must each be committed to the joint activity;
3. The participants must each be committed to supporting the effort of the others.

Consider the example of Alice and Bill who intend to paint a house together. Suppose Alice wants to paint the house red and Bill wants to paint it blue. Both are aware that their subplans conflict, and that the other is aware of it as well. Therefore they do not have a shared cooperative activity, even if they end up painting the house together. But subplans do not have to be identical in order to have a shared cooperative activity. For instance, if Bill wants to use cheap paint but Alice want to buy from a specific store, they could buy cheap paint from that specific store. Bratman puts the following constraints on an action J to be a shared cooperative activity:

1. We do J (which can involve cooperation, but doesn't have to);
2. It is common knowledge between us that we are both committed to mesh subplans;
3. (2) leads to (1) by way of mutual responsiveness (in the pursuit of completing our action) of intention and in action.

Examples of non-reductionist theories of collective intention are, for example, Gilbert [20], who simplifies the problem of collective intentionality to something two people walking together. From this, she identifies four necessary conditions on collective intentions. Tuomela and Miller [52] distinguish the concept of joint I-intention from the concept of joint We-Intention. According to their theory, the concept of joint We-intention cannot be reduced to individual mental attitudes of the agents in the group such as beliefs, desires and individual intentions. Imagine for instance that Anne and Bob intend to lift a table together. First, Ann needs to intend to do her part. Next, she should believe that it is possible to lift the table and believe that Bob also intends to do his part. Moreover, Ann needs to believe that Bob also believes that carrying the table is possible.

Searle [44] also argues that collective intentionality cannot be reduced to individual intentionality. According to Searle, coordination and cooperation is crucial in defining we-intentions, and he gives the following counter example to Tuomela and Miller's we-intentions: A group of business school graduates intend to pursue their own selfish interests, but believe that by doing so, they will indirectly serve humanity. They also believe that their fellow graduates will do likewise, but they do not actively cooperate with one another in pursuing their goals. Searle holds that this fulfills the Tuomela and Miller criteria, but collective intentionality does not actually exist in such a situation.

Velleman [61] is concerned with how a group is capable of making a decision using speech acts, which he considers to be intentions in itself. He argues that collective intention is not the summation of multiple individual intentions (as Tuomela and Miller thought), but rather one shared intention. An intention exists outside of the mind of agent, within a verbal statement. The causal power is in the verbal statement, because of the desire not to speak falsely.

The philosophical dispute is reflected in the computer science literature as well. For instance, Cohen and Levesque [14], following their well-known work on individual intentions (discussed above), propose a reductive account of collective intentions by defining them in terms of group goals and mutual beliefs. Dunin-Keplicz and Verbrugge [18], in contrast, regard both collective intentions and individual intentions a "first-class citizens". Frameworks for flexible teamwork also regularly use theories of collective intentions. For instance, Tambe [50] uses the notion of joint intentions by Cohen and Levesque as a basic building block to define teamwork.

6.5. Development of the framework and relation to our previous wok

6.5.1. Development of PAL

This article combines and builds further on a series of previous papers published by us. The initial proposal of the database perspectives is due to Shoham [46]. His proposal is largely informal but the main ideas form the basis of this article and our formalism can be seen as a formalization of his main ideas. The first formalization of the database perspective was in Icard et al. [25], where we presented a formal semantic model to capture action, belief and intention, based on Shoham's database perspective. We provided postulates for belief and intention revision, and stated a representation theorem relating our postulates to the formal model. However, we noted that there were problems with this formalization, and much of our further work consists of developing the right formal framework. First, we showed that the axiomatization of the original logic is incomplete, and we provided a complete axiomatization [56]. We also adapted the coherence condition in various ways (see next subsection). In van Zee et al. [60] we proved representation theorems for the belief database, and in van Zee and Doder [58] we extended this to intentions. We discussed various extensions to our framework, focusing on enterprise-level decision making [55] and a multi-agent perspective [57]. In this article we provide a complete formal model that combines all our previous work. We furthermore provide the full proofs of all the theorems and we develop an account of iterated revision and a multi-agent extension.

6.5.2. Alternative coherence conditions

An obvious coherence condition we can put on beliefs and intentions is the following:

$$B \vdash \bigwedge_{(a,t)\in I} post(a)_t.$$

However, this solution suffers from the well-known "Little Nell" problem, identified by McDermott [31] (Section 6.2.3). A weaker variant is formalized semantically by Icard et al. [25] as follows:

$$\pi, 0 \models \Diamond \bigwedge_{(a,t)\in I} pre(a)_t$$

Although the semantics in the current article is slightly different, the general idea of this formula is clear: There exists a path, equivalent with the current path up to time 0, in which all the preconditions of the intended actions hold.

However, this is clearly too weak. An agent may believe that all the preconditions hold on paths where none of its intended actions are carried out. We discuss various examples for this in other papers [56,58,59].

In order to resolve this, we propose a different condition van Zee et al. [56], requiring that the beliefs of an agent should be consistent with the preconditions of it's intended action. The agent does not have to believe the preconditions of its intended actions, but it should not believe the negation of the precondition of an intended action. Therefore, we introduce precondition formulas that are derived from the contingent beliefs:

$$Pre(B^I) = Cl(B^I \cup \{ \bigwedge_{(a,t)\in I} pre(a)_t \})$$

We can then express the condition as follows: $Pre(B^I)$ is consistent.

In a followup paper [58] we note that this condition is too weak as well (we refer to that paper for examples and additional motivation). The main problem is that it is not possible to define the precondition of a set of actions in terms of preconditions of individual actions, because it cannot be ensured that all the intentions are fulfilled on the same path as well. Therefore, in order to formalize a coherence condition in PAL, we extend the language with preconditions of finite action sequences, which ensures that after executing the first action, the precondition for the remaining actions are still true.

6.5.3. Timeful application

Separately, Shoham further developed his ideas with Jacob Banks, one of his PhD students, and behavioral economist Dan Ariely in the intelligent calendar application Timeful, which attracted over $6.8 million in funding and was acquired by Google in 2015,[13] who aim to integrate it into their Calendar applications. As Shoham [47] says himself: "The point of the story is there is a direct link between the original journal paper and the ultimate success of the company." (p. 47) Thus, it seems clear that his philosophical proposal has led to some success on the practical side. In our research, we aim to show that his proposal can lead to interesting theoretical insights as well.

7. Future work

Shoham [46] suggests a large number of direction of future work for the database perspective. In this article we already studied iterated revision as well as a multi-agent extension. We discuss some remaining directions of research which we deem most promising.

7.1. Quantitative beliefs

Our focus in this paper has been on *qualitative* models of belief and belief revision. However, much of our framework could be naturally extended to capture intention revision in the context of *quantitative* belief revision. Imagine we have a distribution P defined on the (finite) set $\mathbb{M}^{|t}$ of bounded models. This naturally induces a distribution $\mathbb{P}$ on $\mathscr{L}^{|t}$, whereby $\mathbb{P}(\varphi) = P(\{m \in \mathbb{M}^{|t} : m \models \varphi\})$. Whereas the definition of coherence we gave above in Definition 17 is quite minimal, one can imagine replacing this definition with one slightly stronger. A natural strengthening in the probabilistic setting would be to require $\mathbb{P}(Cohere(I)) > \theta_c$ for some threshold $\theta_c > 0$. That is, one might like to be reasonably confident that one's plan will succeed. The question now is how to model qualitative belief in this setting.

Recent work in philosophical logic has suggested various ways of bridging quantitative and qualitative frameworks, using the notion of an *acceptance rule*: a rule for determining a belief set B from a probability measure P. For concreteness, we illustrate how our framework could be extended by drawing on a concrete example, namely the acceptance rule defined by Leitgeb [29]. In short, we would select a second threshold θ_b and, following Leitgeb, propose that the agent believes ψ just in case $\mathbb{P}(\psi \mid \varphi) > \theta_b$ for any φ with $\mathbb{P}(\varphi) > 0$. This notion of "stable belief" fits well with our framework, as it is guaranteed that there is always a single strongest stable proposition ψ. Following our assumption of "opportunistic planning" we would typically have that $\theta_b > \theta_c$, allowing that we do not necessarily outright believe that the preconditions for our intended actions will be satisfied.

[13] http://venturebeat.com/2015/05/04/google-acquires-scheduling-apptimeful-and-plans-to-integrate-it-into-google-apps/.

While we have defined weak beliefs in terms of a pair (SB, I), in this setting it would make sense to define them using the full distribution. Intuitively, we weakly believe φ if its probability is stably high given that all of the intended actions are carried out; that is, we could define

$$WB(\mathbb{P}, I) = \{\varphi : \mathbb{P}\big(\varphi \mid \bigwedge_{(a,t)\in I} do(a,t)\big) > \theta_b\}$$

Because the set of beliefs is stable under conditioning, this set would of course extend the closure of the set of strong beliefs together with $\{do(a,t) : (a,t) \in I\}$.

An interesting question in this setting is whether it would be possible to prove an analogue of our Theorem 2. Leaving the definition of a selection function for intentions in tact (but substituting the quantitative notion of coherence), one question would be what are reasonable postulates characterizing intention revision together with belief revision when the latter amounts to probabilistic conditionalization? Coming from the other direction, is there a natural probabilistic notion of updating that would allow our representation theorem to go through with exactly the same postulates? For instance, Mierzewski [34] has shown that AGM can be recovered in the quantitative setting if updating is by conditionalization, but only after shifting from one's distribution to the maximum entropy measure with the same set of stable (strong) beliefs. That our representation result would go through in our joint revision setting we leave as a worthy conjecture.

7.2. Teleology

One aspect we left out of our study purposely is the idea of teleology, or the purpose for which I form intentions. It is clear that some notion of goal is required to capture teleological aspects of intending, which is why Cohen and Levesque started with goal. Although it was beneficial for our analysis of intention revision to abstract away from goals, they are clearly central to the revision problem. Information about goals allows the agent to, for instance, replace intentions instead of merely discarding them. This paves the road to develop a richer notion of intentions, such as that "intentions normally pose problems for the agents; the agent needs to determine a way of achieving them" [13]. Interestingly, viewed from Shoham's database perspective, adding goals to the formalism blurs the distinction between planner and databases. If the databases take over part of the planning, then well-known problems such as the frame problem become more stringent: Once a fact is established (for example, as a postcondition of an intention), it persists until it explicitly contradicts postconditions established by future intentions. Existing action logics (e.g., the Event Calculus or the Fluent Calculus) and database approaches (e.g., TMMS by Dean and McDermott [17]) have dealt with these problems in detail, so comparing and possibly enriching them with our formalism seems both useful and relevant future work.

8. Conclusion

We study the interplay between intention, time, and belief, and present two main contributions.

The first contribution is to formalize assumptions of intentions. We observe that assumptions cannot be used to derive strong beliefs from weak beliefs, but cannot be ignored either. In order to deal with this, we first develop a branching-time temporal logic, called Parameterized-time Action Logic (PAL) in order to formalize beliefs. The language of this logic contains formulas to reason about possibility, preconditions, postconditions, and the execution of actions. The semantics of this logic is close to CTL*, and in this way follows the tradition of BDI logics of Rao and Georgeff [39]. An important difference is that we do not use modal operators to reason about time, but we use explicit time points. We axiomatize this logic and prove that the axiomatization is sound and strongly complete with respect to our semantics.

We separate strong beliefs from weak beliefs. Strong beliefs are beliefs that occur in the belief database, and they are independent of intentions. Weak beliefs are obtained from strong beliefs by adding intentions to the strong beliefs, and everything that follows from that. We formalize a *coherence condition* on the beliefs and intentions, which is our solution to the first challenge. The condition states that the agent weakly believes it is possible to jointly perform all of its intended actions.

Our second contribution is to model what-if statements. In other words, we study the *dynamics* of the interplay between intentions, time, and beliefs. Our approach is to use the well-known and well-studied AGM theory of belief revision as our starting point. We develop a set of postulates for the joint revision of belief and intentions, and that we prove a variation of the Katsuno and Mendelzon [26] representation theorem. To this end, we define a revision operator that revises beliefs up to a specific time point. We show that this leads to models of system behaviors which can be finitely generated, i.e. be characterized by a single formula. In addition, we study iterated revision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Yoav Shoham for providing useful feedback on the initial ideas of this article. We also would like to thank the anonymous reviewers for their detailed feedback, and in particular for bringing our attention to the example at the beginning of Section 4.4.

Dragan Doder is funded by ANR-11-LABX-0040-CIMI.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.artint.2020.103270.

References

[1] C.E. Alchourrón, P. Gärdenfors, D. Makinson, On the logic of theory change: partial meet contraction and revision functions, J. Symb. Log. 50 (2) (1985) 510–530.
[2] R. Alur, D.L. Dill, A theory of timed automata, Theor. Comput. Sci. 126 (1994) 183–235.
[3] R. Alur, T. Feder, T.A. Henzinger, The benefits of relaxing punctuality, J. ACM 43 (1) (Jan. 1996) 116–146.
[4] R. Alur, T.A. Henzinger, Real-time logics: complexity and expressiveness, Inf. Comput. 104 (1990) 390–401.
[5] R. Alur, T.A. Henzinger, O. Kupferman, Alternating-time temporal logic, in: Revised Lectures from the International Symposium on Compositionality: The Significant Difference, COMPOS'97, Springer-Verlag, London, UK, UK, 1998, pp. 23–60.
[6] C. Baier, J.-P. Katoen, Principles of Model Checking, Representation and Mind Series, The MIT Press, 2008.
[7] P. Bellini, R. Mattolini, P. Nesi, Temporal logics for real-time system specification, ACM Comput. Surv. 32 (1) (Mar. 2000) 12–42.
[8] M.E. Bratman, Intention, Plans, and Practical Reason, Harvard University Press, Cambridge, MA, 1987.
[9] M.E. Bratman, Shared Agency: A Planning Theory of Acting Together, Oxford University Press, 2014.
[10] M.E. Bratman, D.J. Israel, M.E. Pollack, Plans and resource-bounded practical reasoning, Comput. Intell. 4 (3) (1988) 349–355.
[11] N. Bulling, V. Goranko, How to be both rich and happy: combining quantitative and qualitative strategic reasoning about multi-player games (extended abstract), in: F. Mogavero, A. Murano, M.Y. Vardi (Eds.), SR, in: EPTCS, vol. 112, 2013, pp. 33–41.
[12] K.L. Clark, Negation as failure, in: Logic and Data Bases, Springer, 1978, pp. 293–322.
[13] P.R. Cohen, H.J. Levesque, Intention is choice with commitment, Artif. Intell. 42 (2–3) (1990) 213–261.
[14] P.R. Cohen, H.J. Levesque, Teamwork, Noûs 25 (4) (1991) 487–512.
[15] A. Darwiche, J. Pearl, On the logic of iterated belief revision, Artif. Intell. 89 (1–2) (1997) 1–29.
[16] D. Davidson, Actions, reasons, and causes, J. Philos. 60 (23) (1963) 685–700.
[17] T.L. Dean, D.V. McDermott, Temporal data base management, Artif. Intell. 32 (1) (1987) 1–55.
[18] B. Dunin-Keplicz, R. Verbrugge, Collective intentions, Fundam. Inform. 51 (3) (2002) 271–295.
[19] E.A. Emerson, R.J. Trefler, Parametric quantitative temporal reasoning, in: LICS, IEEE Computer Society, 1999, pp. 336–343.
[20] M. Gilbert, Walking together: a paradigmatic social phenomenon, Midwest Stud. Philos. 15 (1) (1990) 1–14.
[21] J. Grant, S. Kraus, D. Perlis, M. Wooldridge, Postulates for revising BDI structures, Synthese 175 (1) (2010) 39–62.
[22] S.O. Hansson, A Textbook of Belief Dynamics. Theory Change and Database Updating, Kluwer Academic, 1999.
[23] A. Herzig, E. Lorini, L. Perrussel, Z. Xiao, BDI Logics for BDI Architectures: Old Problems, New Perspectives, KI - Künstliche Intelligenz, 2016, pp. 1–11.
[24] R. Holton, Partial belief, partial intention, Mind 117 (465) (2008) 27–58.
[25] T. Icard, E. Pacuit, Y. Shoham, Joint revision of belief and intention, in: Proc. of the 12th International Conference on Knowledge Representation, 2010, pp. 572–574.
[26] H. Katsuno, A.O. Mendelzon, Propositional knowledge base revision and minimal change, Artif. Intell. 52 (3) (Dec. 1991) 263–294.
[27] R. Koymans, Specifying real-time properties with metric temporal logic, Real-Time Syst. 2 (4) (Oct. 1990) 255–299.
[28] J. Kvarnström, TALplanner and other extensions to Temporal Action Logic, Ph.D. thesis, Linköpings universitet, 2005.
[29] H. Leitgeb, The stability theory of belief, Philos. Rev. 123 (2) (2014) 131–171.
[30] E. Lorini, A. Herzig, A logic of intention and attempt, Synthese 163 (1) (2008) 45–77.
[31] D. McDermott, A temporal logic for reasoning about processes and plans, Cogn. Sci. 6 (2) (1982) 101–155.
[32] A. Meier, M. Mundhenk, M. Thomas, H. Vollmer, The complexity of satisfiability for fragments of ctl and ctl*, Electron. Notes Theor. Comput. Sci. 223 (2008) 201–213.
[33] J.-J. Meyer, W. van der Hoek, B. van Linder, A logical approach to the dynamics of commitments, Artif. Intell. 113 (1–2) (Sep. 1999) 1–40.
[34] K. Mierzewski, Probabilistic stability: dynamics, nonmonotonic logics, and stable revision, Master's thesis, Universiteit van Amsterdam, 2018.
[35] E.T. Mueller, Commonsense Reasoning: An Event Calculus Based Approach, 2nd edition, Morgan Kaufmann, Amsterdam, 2014.
[36] J. Ostroff, Temporal Logic for Real-Time Systems, Advanced Software Development Series, Research Studies Press, 1989.
[37] N. Papadakis, D. Plexousakis, Actions with duration and constraints: the ramification problem in temporal databases, Int. J. Artif. Intell. Tools 12 (3) (2003) 315–353.
[38] T. Patkos, A formal theory for reasoning about action, knowledge and time, Ph.D. thesis, University of Crete-Heraklion, 2010.
[39] A.S. Rao, M.P. Georgeff, Modeling rational agents within a BDI-architecture, in: Principles of Knowledge Representation and Reasoning, Proceedings of the Second International Conference, Morgan Kaufmann, San Mateo, 1991, pp. 473–484.
[40] R. Reiter, A logic for default reasoning, Artif. Intell. 13 (1–2) (1980) 81–132.
[41] M. Reynolds, An axiomatization of full computation tree logic, J. Symb. Log. 66 (3) (2002) 1011–1057.
[42] L. Savage, The Foundations of Statistics, Wiley, New York, 1954.
[43] J. Searle, Intentionality. An Essay in the Philosophy of Mind, Cambridge University Press, 1983.
[44] J. Searle, The Construction of Social Reality, Free Press, 1995.
[45] S. Shapiro, S. Sardina, J. Thangarajah, L. Cavedon, Revising conflicting intention sets in BDI agents, in: Proceedings of the 11th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2012), Valencia, Spain, June 2012, pp. 1081–1088.
[46] Y. Shoham, Logical theories of intention and the database perspective, J. Philos. Log. 38 (2009) 633–647.
[47] Y. Shoham, Why knowledge representation matters, Commun. ACM 59 (1) (Jan. 2016) 47–49.
[48] Y. Shoham, K. Leyton-Brown, Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations, Cambridge University Press, 2008.
[49] M.P. Singh, A critical examination of the Cohen-Levesque theory of intentions, in: Proceedings of the 10th European Conference on Artificial Intelligence, ECAI '92, John Wiley & Sons, Inc., New York, NY, USA, 1992, pp. 364–368.
[50] M. Tambe, Tracking dynamic team activity, in: AAAI/IAAI, vol. 1, 1996, pp. 80–87.

[51] M. Thielscher, The concurrent, continuous fluent calculus, Stud. Log. 67 (3) (2001) 315–331.
[52] R. Tuomela, K. Miller, We-intentions, Philos. Stud. 53 (3) (1988) 367–389.
[53] W. van der Hoek, W. Jamroga, M. Wooldridge, Towards a theory of intention revision, Synthese 155 (2) (2007) 265–290.
[54] W. Van der Hoek, M. Wooldridge, Towards a logic of rational agency, Log. J. IGPL 11 (2) (2003) 135–159.
[55] M. van Zee, Rational architecture = architecture from a recommender perspective, in: Proceedings of the International Joint Conference on Artificial Intelligence, 2015.
[56] M. van Zee, M. Dastani, D. Doder, L. van der Torre, Consistency conditions for beliefs and intentions, in: Twelfth International Symposium on Logical Formalizations of Commonsense Reasoning, 2015.
[57] M. van Zee, M. Dastani, Y. Shoham, L. van der Torre, Collective intention revision from a database perspective, in: Collective Intentionality Conference, July 2014.
[58] M. van Zee, D. Doder, AGM-style revision of beliefs and intentions, in: Proceedings of the 22nd European Conference on Artificial Intelligence (ECAI'16), September 2016.
[59] M. van Zee, D. Doder, AGM-style revision of beliefs and intentions from a database perspective (preliminary version), in: Proceedings of the 16th International Workshop on Non-monotonic Reasoning (NMR'16), April 2016.
[60] M. van Zee, D. Doder, M. Dastani, L. van der Torre, AGM revision of beliefs about action and time, in: Proceedings of the International Joint Conference on Artificial Intelligence, 2015.
[61] J.D. Velleman, How to share an intention, Philos. Phenomenol. Res. 57 (1) (1997) 29–50.
[62] M. Wooldridge, Reasoning about Rational Agents, MIT Press, 2000.
[63] M. Wooldridge, N.R. Jennings, Agent theories, architectures, and languages: a survey, in: Intelligent Agents, Springer, 1995, pp. 1–39.