Synthese (2010) 175:39–62
Knowledge, Rationality & Action 127–150
DOI 10.1007/s11229-010-9735-1

Postulates for revising BDI structures

John Grant · Sarit Kraus · Donald Perlis · Michael Wooldridge

Received: 28 February 2010 / Accepted: 2 March 2010 / Published online: 21 July 2010

Abstract The process of rationally revising beliefs in the light of new information is a topic of great importance and long-standing interest in artificial intelligence. Moreover, significant progress has been made in understanding the philosophical, logical, and computational foundations of belief revision. However, very little research has been reported with respect to the revision of other mental states, most notably propositional attitudes such as desires and intentions. In this paper, we present a first attempt to formulate a general framework for understanding the revision of mental states. We develop an abstract belief-desire-intention model of agents, and introduce a notion of rationality for this model. We then present a series of formal postulates characterizing the processes of adding beliefs, desires, and intentions, updating costs and values, and removing beliefs, desires, and intentions. We also investigate the computational complexity of several problems involving the abstract model and comment on algorithms for revision.

J. Grant
Department of Mathematics, Towson University, Towson, MD 21252, USA
e-mail: jgrant@towson.edu

S. Kraus
Bar-Ilan University, Ramat-Gan 52900, Israel

S. Kraus
Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA
e-mail: sarit@umiacs.umd.edu

D. Perlis
Department of Computer Science , University of Maryland, College Park, MD 20742, USA
e-mail: perlis@cs.umd.edu

M. Wooldridge (✉)
Department of Computer Science, University of Liverpool, Liverpool L69 3BX, UK
e-mail: mjw@liv.ac.uk

Keywords Revision postulates · Belief revision · Intention revision · BDI agents

1 Introduction

The process of rationally revising beliefs has been an active research area in AI and philosophy for several decades, starting with the seminal work of Gärdenfors (1988). There are several possible approaches to the study of belief revision. One can take a constructive approach, and define algorithms that show exactly how a belief state is to be revised, either to accommodate new information or to remove old information. Alternatively, one can adopt an axiomatic approach, defining the conditions that belief update operations might satisfy; in this respect, the AGM postulates are perhaps the best-known and most successful within the AI community (Alchourron et al. 1985). And finally, of course, one can link the two approaches, defining axioms that characterize the process of rational update, and then giving update procedures that faithfully implement these axioms.

While belief revision and update have been the subject of considerable research, much less attention has been devoted to the revision of other mental states, and in particular, the revision of mental states related to action, such as intentions. There are several reasons for this (van der Hoek et al. 2007): perhaps the most significant is that intentions and related propositional attitudes such as desires are closely linked with attitudes such as belief and knowledge. For example, if an agent intends to bring about a state of affairs ϕ, then this implies that the agent believes that ϕ is possible (Cohen and Levesque 1990). And if an agent intends to achieve a state of affairs ϕ, and later revises its beliefs with the information that ϕ is impossible, then in order to maintain a rational mental state, the agent would also have to update its intentions accordingly; presumably by dropping the intention to achieve ϕ as well as any subordinate intentions. Intentions, beliefs, and other related propositional attitudes are thus intertwined in a complex set of dependencies and other relationships. A theory of mental state revision must show how such complex relationships are rationally maintained and modified by the revision of beliefs and other propositional attitudes.

In this paper, we present a theory of rational mental state revision that deals with the three key propositional attitudes of belief, desire, and intention. We hence refer to our model as a "BDI model", following the usage of this term in the artificial intelligence community (Rao and Georgeff 1991, 1992, 1998; Wooldridge 2000). The paper is structured as follows. In Sect. 2 we define the components of the BDI model we use. Basically, we envision a "rational" agent G with given beliefs and desires as one that selects its set of intentions, I, to perform actions in accordance with recipes linking actions to results, in a manner to most economically meet its desires (goals). This in general allows for more than one rational choice of the set of intentions, since G may have many ways of achieving goals, and two or more of those may well be maximally beneficial. Section 3 contains the proposed postulates for BDI structure revision. We envisage G already being in a rational state, but then circumstances arise that lead to a change in its beliefs or desires, or its valuation or cost functions, so that G must update its intentions I to best fit with these changes. This provides G with another aspect to consider: of the various potential updates I' of I that are maximally beneficial with

regard to the new beliefs, desires, etc., choose one with the least alteration from the old intentions I. There are two reasons for this. First, less effort is involved in less change; but more interestingly, if the new plans implicit in I' are very similar to those in I, then familiarity with the old plans (on the part not only of G but also of other agents that G might be interacting with) will be more likely to be useful rather than disruptive when the updated plans come online (However, the current paper does not consider issues arising from multi-agent settings.). In Sect. 4 we briefly consider some issues that may arise in implementing our postulates. We conclude the paper in Sect. 5.

Before proceeding, we should note that we are aware of only three papers that directly deal with intention revision (Georgeff and Rao 1995; van der Hoek et al. 2007; Shoham 2009). The earliest of these, Georgeff and Rao (1995), considers some problems in the logical formalization of intention revision—it does not discuss postulates for intention revision in general. The approach adopted in van der Hoek et al. (2007) is essentially algorithmic: procedures are given for updating mental states, and a logic is defined to characterize these mental states. Our work is different in many respects. In our framework a BDI-structure contains beliefs, desires, intentions, as well as a valuation function on desires and a cost function on actions. We write postulates for every type of revision including adding or deleting a belief, desire, or intention, as well as different ways of revising the valuation and cost functions. Our postulates are aimed at characterizing rational intention revision as a kind of minimization of revision-effort among possible maximum-benefit revisions. Finally, Shoham (2009) gives a brief discussion of intention revision, emphasizing the preconditions and postconditions of the actions, but does not deal with desires, costs, and values.

2 A model of mental state

In this section we define the model of mental state that we use throughout the remainder of the paper. The model captures three key components of mental states: beliefs, desires, and intentions (Rao and Georgeff 1991), and for this reason we call it a *BDI model*. We begin by introducing BDI structures, and then provide axioms that characterize the *rational balance* of mental states for these structures. We then give several complexity results relating to the maintenance of these structures.

2.1 Components of the model

We assume as given the following:

1. A logical language L_0, used by an agent to represent properties of its environment including its beliefs. We do not place any requirements on this language other than that it contains the usual classical logic connectives, which behave in the classical way, and that there is a proof relation $\vdash$ defined for the language. For example, L_0 might be a first-order logic language; but for the purposes of analysis, we will often assume that L_0 is classical propositional logic. We write F_0 for the set of sentences of this language.

We assume that the belief update operations $\dot{+}$ (revision) and $\dot{-}$ (contraction) have already been defined for L_0 (cf. Alchourron et al. 1985; Gärdenfors 1988). In the analysis that follows, we will usually assume that these operations can be performed in *unit* time, i.e., that we have an "update oracle" for L_0. In fact, update can be a computationally complex process (Baral and Zhang 2005); the point is to factor out this complexity from the analysis of updating an agent's overall mental state, so that we get an understanding of the inherent complexity of *this* process.

2. A set of actions $A = \{\alpha_1, \alpha_2, \ldots\}$. Again, we do not assume any concrete interpretation for actions: they may be actions performed in a physical environment, or in a virtual (software) environment, for example. We do not disallow the possibility that such actions are complex, formed with the program constructs of sequence, selection, and iteration, although we make no use of such constructs in our analysis.
3. A set of "recipes" $R = \{\langle \alpha, \theta \rangle \mid \alpha \in A \text{ and } \theta \in F_0\}$, which represents an agent's knowledge about how to achieve certain states of affairs in the environment. Intuitively, a recipe $\langle \alpha, \theta \rangle$ is used to represent the fact that performing action α will accomplish a state of affairs satisfying θ (see, e.g., discussions in Pollack 1990, 1992). For every recipe $r = \langle \alpha, \theta \rangle \in R$, we assume there is a proposition $r_{\alpha,\theta}$. Intuitively, $r_{\alpha,\theta}$ will be used to mean that: (i) the action α is executable, in that its precondition is currently satisfied, and (ii) the performance of α terminates and makes θ true. We write L_r for the propositional language that contains all the formulas of the form $r_{\alpha,\theta}$ for $\langle \alpha, \theta \rangle \in R$.
   In our presentation we treat the recipes, desires, costs, and values in the agent's knowledge base separate from the agent's beliefs. While it may be natural to regard a recipe as a belief (so that for instance a particular recipe can be rejected if evidence against it appears, or a new one accepted, etc.) we have chosen not to do so here, but rather to postulate a separate recipe set, in order to keep the overall presentation simpler. Only those recipes whose actions the agent believes it can execute are listed as beliefs using the propositions $r_{\alpha,\theta}$. Again, to keep the overall presentation as simple as possible, we do not consider the possibility that an intended action affects the precondition of another action. Similarly, we treat desires, costs, and values as separate from beliefs.
4. We denote by L the logic obtained from L_0 and L_r by closure under the classical logical connectives, and we denote by F the set of formulae of this logic.

We will use the following running example throughout the paper to illustrate various concepts.

Example 2.1 Agent G is in Paris and wishes to visit a colleague C in London. G knows that taking flight F will get him there. G also knows that to take F he must get to Orly airport, and that he can do this by taxi or by bus. Thus he has available two "recipes" to get to London: take a taxi and then flight F, or a bus and then flight F. He has chosen the former since by taking the taxi he gets to the airport early and can eat a good meal there. Now one of the following happens:

A. C calls to say she no longer is in London.
B. G's spouse insists that he take the bus instead.

In each case, G will make some changes in mental state. Here are some reasonable possibilities:

A: G drops the belief that C is in London, the desire to go there, the intention to take flight F and the intention to have lunch at Orly. But he is still hungry and adopts the intention of eating at home.
B: G adopts the new intention of getting to Orly by bus, and gives up the intention of getting there by taxi. He also takes out food so he can eat on the way (a new intention). No beliefs or desires change.

We now need a language in which to formalize the beliefs, desires, and intentions of the agent G in this example. The following presents such a language (we also add some propositions and actions that will be used later as we elaborate on the example).

Example 2.2 The language of the agent in Example 2.1 can be formally defined as given below:

1. L_0 contains the propositions: $at\_airport$, at_london, $hungry$, $early$, $good\_meal$, $present$, and $meeting$.
2. The set of actions: $A = \{bus, taxi, fly, eat\_airport, eat\_home, take\_out, buy, call, fax\}$.
3. The set of recipes: $R = \{\langle bus, at\_airport\rangle, \langle taxi, at\_airport \wedge early\rangle, \langle take\_out, \neg hungry\rangle, \langle eat\_airport, early \rightarrow (\neg hungry \wedge good\_meal)\rangle, \langle eat\_home, \neg hungry\rangle, \langle fly, at\_airport \rightarrow at\_london\rangle, \langle buy, present\rangle, \langle call, meeting\rangle\}$. Thus L_r contains the propositions $r_{bus,at\_airport}$, $r_{taxi,at\_airport\wedge early}$, $r_{take\_out,\neg hungry}$, $r_{eat\_airport,early\rightarrow\neg hungry\wedge good\_meal}$, $r_{eat\_home,\neg hungry}$, $r_{fly,at\_airport\rightarrow at\_london}$, $r_{buy,present}$, $r_{call,meeting}$.

2.2 BDI structures

Next we define the concept of a *BDI structure*. Such structures represent our basic model of the mental state of an agent, used throughout the remainder of the paper. For the moment, however, we do not present the constraints on such structures that correspond to "rational balance".

Definition 2.1 A BDI structure $\mathcal{S}$ is a 5-tuple,

$$\mathcal{S} = \langle B_{\mathcal{S}}, D_{\mathcal{S}}, I_{\mathcal{S}}, v_{\mathcal{S}}, (c_{\mathcal{S}}, C_{\mathcal{S}})\rangle$$

(we will usually omit subscripts) such that

– $B = \{b \in F \mid B_0 \vdash b\}$, where B_0 is a finite set. So B is closed under consequence and has a finite basis. B stands for the *beliefs* of the agent.
– $D \subset F_0$ and D is finite. D stands for the *desires* of the agent. We will use d, d_i, … as meta-variables ranging over D.
– $I \subseteq R$. I stands for the intentions of the agent.
We write $goals(I) = \{\ell \mid \langle \alpha, \ell\rangle \in I\}$ and $actions(I) = \{\alpha \mid \langle \alpha, \ell\rangle \in I\}$.

– $v : \mathcal{P}(D) \rightarrow \mathcal{R}^+ \cup \{0\}$, where $\mathcal{R}^+$ is the set of positive real numbers: v is a valuation function that assigns a nonnegative value to each set of desires of the agent. We extend v to $\bar{v}$ on all subsets X of F_0 as follows: $\bar{v}(X) = v(\{d \mid B \dot{+} X \vdash d\})$. We also require that v satisfy the following "entailment-value" condition:

$$\text{for } T \subseteq D \text{ and } T' \subseteq D, \text{ if } T \vdash T' \text{ then } v(T) \geq v(T').$$

– $c : \mathcal{P}(C) \rightarrow \mathcal{R}^+ \cup \{0\}$, where C is a finite subset of A. (c is a cost function for finite sets of actions). We require $actions(I) \subseteq C$, (i.e., the agent has a nonnegative cost associated with the set of actions it intends to do). Also, we require a condition involving sets of actions:

$$\text{if } K \subseteq K' \subseteq C \text{ then } c(K') \geq c(K).$$

There are several points to make about this definition. First, note that we are explicitly representing an agent's beliefs as a set of logical formulae, closed under deduction. Under this model, an agent is said to believe ϕ if ϕ is present in the agent's belief set. This "sentential" model of belief is widely used within artificial intelligence (Konolige 1986; Genesereth and Nilsson 1987). We use a similar representation for desires. We represent an agent's intentions as a set of recipes that it has selected, and implicit within this set of recipes, the states of affairs that it has committed to bringing about. Thus far, our model closely resembles many other models of BDI agents developed in the literature (Rao and Georgeff 1991, 1992). However, the value (v) and cost (c) functions distinguish it. The function c is restricted to a subset C of A since an agent may not know the cost of all actions. Note that the cost function is not required to be additive, since, for example, the cost of performing two actions may be lower than the sum of performing them separately.

Notice that we have said nothing about the *representation* of such structures, which is a key issue if we try to understand the complexity of revision operations (Baral and Zhang 2005). We must first consider how the value and cost functions are represented, since naive representations of them (listing all input/output pairs that define the function) is not practical. Clearly, it would be desirable to have a representation that was polynomial in the size of the remainder of the structure, and moreover, allows the computation of the corresponding v and c functions in polynomial time. Several possible representations suggest themselves. In this paper, we will assume that, except where otherwise stated, these functions are represented as *straight line programs*, using the approach suggested in Dunne et al. (2005).[1] We say that a *finitely represented* BDI structure is one in which we use these representations. For the most part, these representation issues will have no role to play in our analysis; they become significant only when we start to consider computational issues. For the purposes of the present paper, we do not try to model the time varying properties of structures, and so we do not assume any model or representation of time. Let us return to Example 2.2. We have:

[1] While the details are rather technical, what this basically amounts to is that we assume c and v can be represented as a sequence of simple assignment statements with the length of the sequence bounded by some polynomial in the size of the remainder of the structure.

Table 1 The v value in a given column specifies the value of the set of propositions that are associated with + in that column

at_london	−	+	+	+	+	−	−	−
$\neg hungry$	−	−	+	+	−	+	+	−
$good\_meal$	−	−	−	+	+	−	+	+
v	0	1,000	1,035	1,100	1,060	35	70	45

Example 2.3

1. $B = \{\neg at\_london, \neg at\_airport, hungry, r_{fly,at\_airport \rightarrow at\_london}, r_{bus,at\_airport}, r_{taxi,at\_airport \wedge early}, r_{take\_out,\neg hungry}, r_{eat\_airport,early \rightarrow (\neg hungry \wedge good\_meal)}, r_{eat\_home,\neg hungry}, r_{call,meeting}\}$.
   We do not assume that for the agent $good\_meal \rightarrow \neg hungry$ because of the implicit time factor. A good meal at a later time leaves the agent $hungry$ in the near future. So we explicitly indicate that in the case of getting early to the airport, eating at the airport makes the agent not hungry.
2. $D = \{at\_london, \neg hungry, good\_meal\}$.
3. $I = \{\langle taxi, at\_airport \wedge early\rangle, \langle eat\_airport, early \rightarrow (\neg hungry \wedge good\_meal)\rangle, \langle fly, at\_airport \rightarrow at\_london\rangle\}$.
4. The v function is specified in Table 1. For example, $v(\{at\_london\}) = 1000$.
5. The c function is specified as follows. Going to the airport by bus costs 20 (i.e., $c(\{bus\}) = 20$), and going by taxi costs 50. It is impossible to go both by car and by bus and thus the cost of any set that consists of both of them is ∞ (e.g., $c(\{bus, taxi\}) = \infty$). The cost to fly is 500 and is added to any other costs of other actions, e.g., $c(\{bus, fly\}) = 520$). Eating at home or taking out food costs 20 and eating at the airport costs 50. Again, a combination is impossible, so for example $c(\{eat\_home, take\_out\}) = \infty$. However, eating at the airport after going by taxi is cheaper: $c(\{taxi, eat\_airport\}) = 90$. Buying a present costs 30. The cost of a call is 50. Unless otherwise specified, the costs are summed up, e.g., $c(\{bus, eat\_home\}) = 40$. Note that there is no price associated with sending a fax, i.e., $fax \notin C$.

2.3 Rational BDI structures

The concept of BDI structure that we have defined is very weak: it admits structures that could not be considered "rational", for a number of reasons. In this subsection we give several rationality axioms restricting what is allowed for BDI structures, culminating in the concept of a rational BDI structure.

We start by saying that a BDI-structure is *belief rational* if it satisfies the following requirement:

A1 B is consistent, i.e., $B \nvdash \bot$.

Belief rationality is the weakest rationality constraint that we consider. In particular, belief rationality says nothing about how an agent's intentions relate to its beliefs. It is perhaps simplest to think of belief rationality as an "intermediate" mental state,

which occurs after an agent has updated its beliefs in light of new information about the environment, but before it has adjusted its intentions accordingly. For this purpose we define the concept of when a set of intentions I is feasible in the context of a set of beliefs B irrespective of the rest of the BDI structure. We say that I is *feasible* in the context of B if $\forall\langle\alpha, \theta\rangle \in I, r_{\alpha,\theta} \in B$, (i.e., the action part of every intention is believed to be executable). We capture the relationship between beliefs and intentions in the notion of intention rationality. In addition to the feasibility of I with respect to B we require that there be no conflict between the agent's goals and that the agent not try to make something true if it believes it to be true already. We say that a BDI structure is *intention-rational* if it satisfies the following properties:

A2 I is feasible in the context of B.
A3 $goals(I)$ is consistent.
A4 $\forall\theta \in goals(I), B \not\vdash \theta$.

A BDI structure that is both belief and intention rational is said to be *weakly rational*, and hence we will talk of weakly rational BDI structures ("WRBDI structures"). Weak rationality, as the name suggests, is a "reasonable" mental state: the agent has an internally consistent model of the environment, and has a set of intentions I that is consistent and compatible with this model of the environment. However, this definition does not require an agent to have an *optimal* set of intentions, in the following sense. The problem is that I may be poorly formed with respect to the value and cost functions. The agent may have chosen a set of intentions whose actions are costly or the values of the desires achieved by the actions are small (or both). What we need is the concept of the *benefit* of a BDI structure, defined as the difference between the value of the desires satisfied by the intentions, if they were achieved, and the cost of the actions involved in the intentions. Formally, where $\mathcal{S} = \langle B, D, I, v, (c, C)\rangle$, we define:

$$ben(\mathcal{S}) = \bar{v}(goals(I)) - c(actions(I)).$$

Now, if all other aspects of the BDI structure, (i.e., B, D, v, and (c, C)) are fixed, we would like the agent to choose I to maximize its benefit, i.e., choose intentions so as to maximize the value of the function ben. Sometimes we will just write $ben(I)$ when the other aspects of $\mathcal{S}$ are understood.

We will say that a WRBDI structure $\langle B, D, I, v, (c, C)\rangle$ is *rational* ("RBDI structure") if it satisfies the optimality axiom **A5** given below. We can think of rationality as being an "ideal" mental state for an agent: the agent has a consistent model of the environment, and has selected intentions that are mutually consistent and compatible with this model, and that are in addition optimal.

A5 $\nexists I' \subset R$ such that $\mathcal{S}' = \langle B, D, I', v, (c, C)\rangle$ is a WRBDI and $ben(I') > ben(I)$.

We refer to **A5** as an optimality axiom because it says that within the context of a WRBDI structure, with everything except possibly I fixed, no substitute I' could have been chosen with a higher expected benefit. A set of intentions I that is not optimal in the context of a WRBDI structure is said to be *sub-optimal*.

Example 2.4 The BDI structure of Example 2.3 is an RBDI. In particular, $ben(I) = 1100-590 = 510$. We consider here two alternatives. For $I' = \emptyset$ we have $ben(I') = 0$, and for

$$I'' = \{\langle bus, at\_airport\rangle, \langle take\_out, \neg hungry\rangle, \langle fly, at\_airport \rightarrow at\_london\rangle\}$$

we have $ben(I'') = 1035 - 540 = 495$.

It is easy to see that the following proposition is correct.

Proposition 1 *If $\mathcal{S}_1 = \langle B_1, D_1, I_1, v_1, (c_1, C_1)\rangle$ and $\mathcal{S}_2 = \langle B_2, D_2, I_2, v_2, (c_2, C_2)\rangle$ are two RBDIs such that $B_1 = B_2$, $D_1 = D_2$, $v_1 = v_2$, $c_1 = c_2$, and $C_1 = C_2$, then $ben(\mathcal{S}_1) = ben(\mathcal{S}_2)$.*

2.4 Complexity

We prove several complexity results concerning WRBDI structures.

Proposition 2 *Given a finitely presented WRBDI structure $\mathcal{S}$, in which L_0 is classical propositional logic, checking if the intentions of $\mathcal{S}$ are sub-optimal is NP-complete.*

Proof Membership in NP is by a standard "guess and check" algorithm: guess some $I' \subset R$, verify that $I' \neq I$, that the rationality requirements would hold if the agent had intentions I', and that the benefit with I' is higher than with I. For NP-hardness, we reduce the problem to SAT, the problem of determining whether or not a formula of classical propositional logic is satisfiable. Let ϕ be the given SAT instance over propositional variables $X = \{x_1, \ldots, x_k\}$, which w.l.o.g. we can assume to be in CNF, i.e., $\phi = \wedge_i \psi_i$ where each ψ_i is a set of literals over X. Also, let $C\ell_\phi = \{\psi_i \mid \psi_i \in \phi\}$. For each propositional variable $x_i \in X$, we create in the reduction two actions α_{x_i} and $\alpha_{\neg x_i}$, which will correspond to assignments of truth or falsity to variable x_i, respectively. For each clause $\psi_i \in C\ell_\phi$ we create a propositional variable c_i. We then create a set of recipes $R = \{\langle c_i, \alpha_\ell\rangle \mid c_i$ represents $\psi_i \in C\ell_\phi$ and ℓ is a literal in $\psi_i\}$. Let $B = \emptyset$ and let $D = \{c_1, \ldots, c_k\}$. We define $v(D) = 1$, $v(S) = 0$ for $S \neq D$. Then, for a set of actions T we define $c(T) = 1$ if for some $x_i \in X$, $\{\alpha_{x_i}, \alpha_{\neg x_i}\} \subseteq T$, otherwise $c(T) = 0$. Finally, we define $I = \emptyset$. Now, $ben(I) = 0-0 = 0$, so I will be sub-optimal only if the agent can choose a set of intentions I' such that $ben(I') = 1$, i.e. $v(goals(I')) = 1$ and $c(actions(I')) = 0$: such a set of intentions will define a consistent satisfying valuation for the input formula ϕ because the value function requires that all clauses be satisfied and the cost function ensures that the assignment is consistent. It only remains to note that c and v can be represented as straight line programs. □

The following is now immediate.

Corollary 2.1 *Checking that a WRBDI structure is an RBDI structure is co-NP complete.*

Notice that the proof of this proposition illustrates that there may be many different sets of intentions consistent with a set of beliefs in a WRBDI structure; in the reduction used in the proof, every satisfying assignment for the input formula ϕ will correspond to a different set of intentions. An interesting question, therefore, is whether or not, given a BDI structure $\mathcal{S}$ and a set of intentions I', these intentions are contained in *every* RBDI structure corresponding to $\mathcal{S}$. We say I' is *strongly intended* if it is contained in every optimal set of intentions.

Proposition 3 *Given a finitely presented BDI structure $\mathcal{S}$ in which L_0 is classical propositional logic, and a set of intentions I' over $\mathcal{S}$, checking if I' is strongly intended is co-NP-hard.*

Proof Consider the complement problem. We use a reduction based on that of Proposition 2. Given a SAT instance ϕ, we define a new formula $\phi^* = \phi \wedge z$, where z is a new propositional variable, not occurring in ϕ. We then proceed with the reduction as in Proposition 2, and ask whether $I' = \{\langle z, \alpha_z \rangle\}$ is strongly intended. We claim that I' is strongly intended iff ϕ is satisfiable. To see this, observe that the set of optimal intention sets will contain every satisfying assignment for ϕ^*, and by construction every such assignment must assign z the value true; ϕ^* will have no satisfying assignments iff ϕ is unsatisfiable, in which case $\emptyset$ is the only optimal set of intentions. □

It is worth saying a few words about how our BDI model relates to those developed by others in the theory and practice of practical reasoning systems. In the first and best known implemented BDI system, PRS, an agent was equipped with a library of plans, each with an invocation condition and a context condition (Georgeff and Lansky 1987). When an event occurred within the system which matched the invocation condition, and the context condition was also satisfied, then the desire became active, and was considered as one of the potential desires of the system, i.e., one of the potential plans for execution. Choosing between desires was done by structures called *meta-plans*, which can be understood as plans for managing plans. Thus the PRS programmer had to write a program to choose between potential desires. In AgentSpeak, the logic programming style abstraction of PRS (Rao 1996), choosing between desires/active plans was not considered part of the language, but was done by a programmer-defined *selection function* (Bordini et al. 2007, p. 78). In the framework of the present paper, selecting between alternative sets of possible intentions is achieved by doing a cost/value analysis, via the functions c and v. We thus make a commitment within the model of the present paper to how choices between potential sets of intentions are made. The main advantage of the present approach is that it is largely compatible with existing decision theoretic models; we also note that this model has been implemented in PRS-like systems. For example, Huber's JAM system attaches "utilities" to plans, and selects as an intention the active plan (i.e., desire) with the highest such utility (Huber 1999).

3 Postulates for RBDI structure revision

We are now ready to consider the main purpose of the paper: the proper revision of an RBDI structure. Consider some change to the agent's beliefs, desires, or valuation, or

cost functions. The main source of difficulty is that such a change may require additional changes. For example, a change in beliefs may require a change in the agent's intentions in order for the agent to maintain a rational position. Our most fundamental requirement is that the change must result in an RBDI structure; but that is not enough in itself, as there may be many such intention sets. Accordingly, we introduce the following *parsimony* requirement: when intentions must change, they change *as little as possible*. Thus, we assume that, presented with two alternative intention sets, both yielding equal benefit, an agent will choose the one that minimises the changes required with respect to its set of intentions. There are several reasons for making this assumption. Most obviously, and perhaps most importantly, it seems extremely desirable from the point of view of multi-agent interaction. By changing my intentions as little as possible while remaining optimal I improve my predictability from the point of view of other agents, since they only have to minimally revise their model of my future behaviour. This in turn suggests there will be a reduced need for coordination with other agents following a change in my mental state.

To make this idea formal, we must make precise the notion of "closeness" of sets of intentions. Suppose we start with an RBDI structure $\mathcal{S} = \langle B, D, I, v, (c, C)\rangle$ that is revised to yield an RBDI structure $\mathcal{S}' = \langle B', D', I', v', (c', C')\rangle$. Then we will typically require that $\mathcal{S}'$ represent the "closest" rational update to $\mathcal{S}$, in the sense that it satisfies the following parsimony condition:

(*) for every RBDI structure $\mathcal{S}'' = \langle B', D', I'', v', (c', C')\rangle$, either:
1. $|I'' \cap I| < |I' \cap I|$ (i.e., I'' has less in common with I than I'); or
2. $|I'' \cap I| = |I' \cap I|$ and $|I'| \leq |I''|$ (i.e., I' and I'' have the same number of intentions in common with I, but I' is smaller).

Thus, the meaning of (*) is that for any RBDI structure $\mathcal{S}''$ that differs from $\mathcal{S}'$ at most in its set of intentions, I'' cannot be "nearer" to I than I'. It follows trivially that all parsimonious intention updates have the same benefit.

In this section we provide postulates for the various kinds of revision. We deal with both additions and removals. We will use $\oplus$ and $\ominus$ to denote the RBDI structure revision operations.

3.1 Adding beliefs, desires, and intentions

Suppose an agent acquires some additional information $f \in F$. How should this affect its mental state? Clearly, it should affect the agent's beliefs, and for this we assume an AGM-style belief update action $\dot{+}$. However, we might also expect the operation to have some effect on the other components of an agent's mental state. Typically, we expect an agent to update its beliefs and then check its intentions; these may need updating, in light of the new information. Recall that Axiom **A5** requires a rational BDI agent to have a set of intentions that is optimal with respect to the set of beliefs; and since checking the optimality of a set of intentions is co-NP-complete, this implies that changing beliefs will be computationally hard, also. We will see the implications of this shortly.

First, note that not all revisions are possible. Consider, for instance, the case of adding a desire. At first sight it might seem that we simply need to revise the set D,

but this is not the case. In our framework we also need to expand the v function. Thus, just adding a desire is impossible—we also need to provide additional information, relating to the v function. Suppose now that whenever a desire is added, the proper expansion of the v function is included. Even then, it is possible that the expanded value function no longer satisfies the condition requirement for a valuation function. For instance, if a new desire d' is added, and there is a desire $d \in D$ where $d \vdash d'$ and $v(d') > v(d)$, then the entailment-value condition is violated, so the value of d would have to be increased to fix this. Our revisions must be understood in the sense that the revision axioms apply only in those cases where the revision is possible.

The first revision operation that we consider is that of revising a belief. As we noted earlier, revising a belief may have ramifications for other components of the agent's mental state; but in our simplified treatment (due to our separation of beliefs, recipes, desires, and values), it may result only in a change in intentions. The following postulate says that after the revision, the agent has the "nearest" optimal set of intentions to those that it had before, according to the parsimony condition (*), described above.

add a belief $\langle B, D, I, v, (c, C)\rangle \oplus f(\in F) = \mathcal{S}' = \langle B', D', I', v', (c', C')\rangle$
where
$(B \oplus 1)$ $\mathcal{S}'$ is an RBDI structure.
$(B \oplus 2)$ $B' = B \dot{+} f$
$(B \oplus 3)$ $D' = D$
$(B \oplus 4)$ (*)
$(B \oplus 5)$ $v' = v$.
$(B \oplus 6)$ $(c', C') = (c, C)$

Example 3.1 We return to Example 2.3. Suppose the friend from London calls the agent and tells him that when he arrives in London he will be her guest at a fine restaurant. The agent updates its belief with the proposition $at\_london \rightarrow good\_meal$. As a result it should update its intentions as well since it is no longer necessary to take a taxi and eat at the airport to have a good meal; he will get a good meal in London with no additional cost. There appear to be three possible best sets of intentions. Consider first:

$$I_1 = \{\langle bus, at\_airport\rangle, \langle take\_out, \neg hungry\rangle, \langle fly, at\_airport \rightarrow at\_london\rangle\}.$$

Here, $ben(I_1) = 1100 - 540 = 560$. Next, let:

$$I_2 = \{\langle bus, at\_airport\rangle, \langle fly, at\_airport \rightarrow at\_london\rangle\}.$$

Now, $ben(I_2) = 1060 - 520 = 540$. Finally, let:

$$I_3 = \{\langle bus, at\_airport\rangle, \langle eat\_home, \neg hungry\rangle, \langle fly, at\_airport \rightarrow at\_london\rangle\}.$$

Here, $ben(I_3) = 1100 - 540 = 560$. So I_1 and I_3 have the same highest benefit. Assume the agent chooses I_1.

An interesting question is checking if a set of intentions will necessarily result as a consequence of updating with a belief.

Proposition 4 *Let $\mathcal{S} = \langle B, D, I, v, (c, C)\rangle$ be a finitely presented BDI structure $\mathcal{S}$, in which L_0 is classical propositional logic, let $f \in F - B$, and let I' be a set of intentions over $\mathcal{S}$. Then the problem of checking whether I' will be strongly intended in $\mathcal{S} \oplus f$ is co-NP-hard.*

Proof We can reduce the problem considered in Proposition 3 as follows. Update beliefs with $\top$ (i.e., the logical constant for truth) and then ask whether I' is strongly intended in the resulting update. Now, updating beliefs with $\top$ causes no change to beliefs, and so the upshot is that I' will be strongly intended in the updated (i.e., unchanged) belief set iff I' is strongly intended in the original belief set, which is exactly the problem proved to be co-NP-hard in Proposition 3. (Recall that we are assuming belief update takes unit time.) □

Next, we consider adding a desire. The situation here is immediately more complex than that for adding beliefs, since we cannot simply add a new element to the set D of desires: this is because an agent is required to have a value for all subsets of its desires. Thus, adding a desire involves adding both a new element d, which will be added to the desire set D, and also a functional component (here written as w), which gives the value of each possible subset of desires. We require that this function should agree with the previous value function on all previously defined desires. Note that beliefs remain unchanged by the addition of desires, although the addition of a desire may result in changes to other components of the agent's mental state.

add a desire $\langle B, D, I, v, (c, C)\rangle \oplus (w, d) = \mathcal{S}' = \langle B', D', I', v', (c', C')\rangle$
where $d \in F_0 - D$, $w : T \to \mathcal{R}^+ \cup \{0\}$ for all $T \subseteq D \cup \{d\}$ such that $d \in T$ and v', as given in $(D \oplus 5)$, is an extension of v.
$(D \oplus 1)$ $\mathcal{S}'$ is an RBDI structure.
$(D \oplus 2)$ $B' = B$.
$(D \oplus 3)$ $D' = D \cup \{d\}$
$(D \oplus 4)$ (*)
$(D \oplus 5)$ For all $S \subseteq D \cup \{d\}$

$$v'(S) = \begin{cases} w(S) & \text{if } d \in S \\ v(S) & \text{otherwise.} \end{cases}$$

$(D \oplus 6)$ $(c', C') = (c, C)$

Example 3.2 We return to Example 3.1. Suppose the agent checks his calendar and realizes that his son has a birthday tomorrow; he adopts a new desire to have a present to give him. As a result of this update, D and v change. However, since the agent at this point does not have $r_{buy,present}$ in its belief set and $\langle buy, present\rangle$ is the only recipe leading to $present$, the intention set is not changed.

Adding an intention is the next case we examine. It turns out that in some ways this is the most interesting type of revision, because we must consider several different cases. To see why this is, note that we have already seen that the intentions of an agent may be modified by what we might call *endogenous intention revision*, where the changes in intentions arise as a result of other internal changes in an agent's mental

state. For example, we saw that adding a belief may cause an agent to subsequently modify its intentions in order to stay in a rational state. However, intention revision can also be *exogenous*; that is, directly caused by an external entity. For example, consider a software agent that is associated with an external "owner", who directly instructs the agent to add a new intention.

Within the case of exogenous intention revision, there are several interesting possibilities, which are reflected in the different "add intention" cases we consider. For example, the weakest kind of instruction that might be given by the owner roughly amounts to "add this intention if this can be done at no cost". With this type of revision, captured in our case **add an intention (A)**, below, an agent is presented with a target set of intentions, and is required to consider whether there is an alternative set of intentions that could rationally be adopted, (i.e., with the same benefit as the currently adopted set of intentions), such that the new intentions contain the target intention. However, other types of exogenous intention revisions are also possible. Here, the agent is presented with an intention and instructed to *adjust its mental state so as to make the intention a rational choice*. This may seem odd, but it is not unrealistic. For example, imagine the situation in a company where a worker is given new guidelines about the strategic priorities of the company; this requires the worker to adjust their mental state to accommodate the new position. This required change may lead to belief, desire, and intention revision, but it may also lead to changes in the cost or value functions. Thus, the agent is instructed to change its fundamental values to accommodate the new position. Another way of thinking about this is that the external entity is in possession of additional information that he used when deciding on the intention revision, and the agent tries to adjust its RBDI accordingly.

In what follows, we explore the various possible cases. We emphasise that we consider only the "basic" cases, and explain what must be done in the more complex cases. Notice that we do not define the circumstances under which a particular type of update takes place: we assume the external entity will specify explicitly which type of revision it desires to take place.

In all of these cases, though, there is one modification that we will assume. Consider adding the intention $i = \langle \alpha, \theta \rangle$. Now, the agent's prior intention set I may already contain one or more intentions whose goal is logically implied by θ; the agent need therefore no longer explicitly maintain these intentions, since they will be subsumed by θ. We can therefore delete such intentions without reducing the overall benefit of the intention set. Thus, writing $I^{\theta} = \{\langle \beta, \theta' \rangle \in I \mid \theta \vdash \theta'\}$ we will always start with $I - I^{\theta}$. Also, motivated by the same observation, we modify the parsimony condition (*) as follows:

$(*)^{+i}$ For any RBDI structure $\mathcal{S}'' = \langle B', D', I'', v', (c', C') \rangle$, where $i \in I''$ and $I'' \cap I^{\theta} = \emptyset$, either:
1. $|I'' \cap (I - I^{\theta})| < |I' \cap (I - I^{\theta})|$; or
2. $|I'' \cap (I - I^{\theta})| = |I' \cap (I - I^{\theta})|$ and $|I'| \leq |I''|$.

We now move on to the various **add an intention** cases. The first case, **add an intention (A)**, only requires changing the set of intentions. Intuitively, this case corresponds to the external entity saying "add this intention only if there is a rational set of intentions containing this one, without any changes to beliefs, desires or the valuation

or cost functions". We remark that, as in the case of revisions in general, this particular type of revision is not always possible. In fact, **add an intention (A)** is possible only when the benefit from the deleted intentions is the same as the benefit that is obtained from adding the new intention.

add an intention (A) $\langle B, D, I, v, (c, C)\rangle \oplus^A i (= \langle \alpha, \theta\rangle (\in R)) = \mathcal{S}' = \langle B', D', I', v', (c', C')\rangle$
$(I \oplus^A 1)$ $\mathcal{S}'$ is an RBDI structure.
$(I \oplus^A 2)$ $B' = B$.
$(I \oplus^A 3)$ $D' = D$.
$(I \oplus^A 4)$ $i \in I'$ and $(*)^{+i}$.
$(I \oplus^A 5)$ $v' = v$.
$(I \oplus^A 6)$ $(c', C') = (c, C)$.

Example 3.3 We return to Example 3.1 where the agent chose I_1. Suppose the agent's wife calls and tells him that he should eat at home if it is possible given his current beliefs, desires, valuation and cost functions. Now, based on his wife's request, the external requirement in this case, the agent adds the intention $\langle eat\_home, \neg hungry\rangle$. As explained above, the removal of I^θ requires the removal of the intention $\langle take\_out, \neg hungry\rangle$. Thus the agent obtains I_3 as the result with the same benefit as I_1 and nothing involving beliefs, desires, valuations, or costs is modified.

However, if for example, $r_{eat\_home,\neg hungry}$ did not belong to his beliefs, he would not have changed his intention set, i.e., he would have kept the $\langle take\_out, \neg hungry\rangle$ in his intention set.

Proposition 5 *For all RBDI structures $\mathcal{S}$ and intentions i, if* **add an intention (A)** *is possible, then $ben(\mathcal{S}) = ben(\mathcal{S} \oplus^A i)$.*

Proof If $ben(\mathcal{S}) < ben(\mathcal{S} \oplus^A i)$ then $\mathcal{S}$ would not satisfy the optimality requirement **A5**, while if $ben(\mathcal{S}) > ben(\mathcal{S} \oplus^A i)$ then $\mathcal{S} \oplus^A i$ would not satisfy **A5**. □

While **add an intention (A)** only changes an agent's intentions, the other cases we consider may require changes to other parts of the RBDI structure: **add an intention (B)** changes the beliefs; **add an intention (C)** changes the cost function; **add an intention (D)** changes the desires and hence also the valuation function. We consider these to be the basic cases. Sometimes, several parts of the RBDI structure must be changed: for example, **add an intention (BD)** changes both the beliefs and the desires; we will not deal with these cases in detail as they follow from the basic cases. A rationale for **add an intention (B)** may be that the required intention addition exposes an error in the agent's beliefs, but is a stronger requirement than just changing beliefs.

Next we consider the case where the agent did not include $i = \langle \alpha, \theta\rangle$ as an intention because of its beliefs. There are several reasons why i might not be included because of the agent's beliefs. One is that the agent may believe that θ is already true. Another possibility, however, is that the agent does not believe it can do α. This can be corrected by modifying the set of beliefs. We can think of this as the external entity saying "add this intention if there is a rational set of intentions containing this one after making appropriate changes to your beliefs, without making changes to your desires or the valuation or cost functions".

add an intention (B) $\langle B, D, I, v, (c, C)\rangle \oplus^B i (= \langle \alpha, \theta\rangle (\in R)) = \mathcal{S}' = \langle B', D', I', v', (c', C')\rangle$
$(I \oplus^B 1)$ $\mathcal{S}'$ is an RBDI structure.
$(I \oplus^B 2)$ $B' = ((B \dot{-} r_{\alpha,\theta} \rightarrow \theta) \dot{-} \theta) \dot{+} r_{\alpha,\theta}$
$(I \oplus^B 3)$ $D' = D$.
$(I \oplus^B 4)$ $i \in I'$ and $(*)^{+i}$.
$(I \oplus^B 5)$ $v' = v$.
$(I \oplus^B 6)$ $(c', C') = (c, C)$

Notice that $(I \oplus^B 2)$ causes the agent to delete both θ and $r_{\alpha,\theta} \rightarrow \theta$. Otherwise, adding $r_{\alpha,\theta}$ may immediately cause the agent to believe θ, making the intention immediately redundant. Of course, we are here only removing explicit implications $r_{\alpha,\theta} \rightarrow \theta$, and there may well of course be more complex subsets B^* of B such that $B^* \vdash \theta$. For the purposes of the present paper, we will not be concerned with such more complex structures, though of course a future treatment might consider such revisions. As indicated above, we can think of **add an intention (B)** as a situation in which the agent assumes that the external entity is more knowledgeable than itself, and that it erroneously believed the deleted beliefs or missed the added beliefs.

Proposition 6 *Let $\mathcal{S} = \langle B, D, I, v, (c, C)\rangle$ be an RBDI structure and $i = \langle \alpha, \theta\rangle$ an intention such that $B \nvdash \theta$ and $B \vdash r_{\alpha,\theta}$. If* **add an intention (A)** *of i is possible, then also* **add an intention (B)** *of i is possible. Also, $\mathcal{S} \oplus^A i$ and $\mathcal{S} \oplus^B i$ are either the same or differ only in their intention set.*

Example 3.4 Now we return to Example 3.2 with intention set I_1. Suppose the agent's wife calls and asks him to buy a present at the airport for their son's birthday. The agent does not have a recipe associating buying a present with having a present, but following the **add an intention (B)** postulates, it will update its beliefs with $r_{buy,present}$ and its new intention set will be $I_1 \cup \{\langle buy, present\rangle\}$ with benefit $1150 - 570 = 580$. It is worth noting that there is an additional intention set with the same benefit: just replace the intention $\langle eat\_home, \neg hungry\rangle$ by $\langle take\_out, \neg hungry\rangle$. However, the parsimony requirement, $(*)^{+i}$ of $I \oplus^B 4$, led to our choice in order to minimize the changes in the intention set.

Suppose that the agent already believes he has a present for his son, i.e., $present \in B$ and he got a similar call as above. In addition to the above changes to its RBDI he will also remove this belief ($present$) trusting his wife's judgment and assuming that he erroneously believed he had a present.

Next, suppose that the agent did not include $i = \langle \alpha, \theta\rangle$ as an intention because no cost was known for α or the cost was known but was too high. The postulate **add an intention (C)** deals with such a situation; we chose this postulate for uniformity. The way we deal with this situation is to redefine the cost function c so that the cost of α is lowered to 0. Of course, this is, in a sense, arbitrary; in fact, any value could be chosen as long as $\mathcal{S}'$ satisfies the requirements. The point is to adjust the mental state of the agent, and in particular its cost and value functions, to admit the intention i as being rational, for which the following serves.

add an intention (C) $\langle B, D, I, v, (c, C)\rangle \oplus^C i (= \langle \alpha, \theta\rangle (\in R)) = \mathcal{S}' = \langle B', D', I', v', (c', C')\rangle$

$(I \oplus^C 1)$ $\mathcal{S}'$ is an RBDI structure.
$(I \oplus^C 2)$ $B' = B$.
$(I \oplus^C 3)$ $D' = D$.
$(I \oplus^C 4)$ $i \in I'$ and $(*)^{+i}$.
$(I \oplus^C 5)$ $v' = v$.
$(I \oplus^C 6)$ $C' = C \cup \{\alpha\}$
$c'(R) = c(R - \{\alpha\})$.

Example 3.5 Again, going back to Example 3.1, suppose the agent is told by his boss to take a taxi in order to get to the airport. Here, the **add an intention (C)** case should be applied and the cost of traveling by taxi will be reduced to 0 (e.g., the cost will be paid by his company). The first step is to use the rule about I^θ to remove the intention $\langle bus, at\_airport\rangle$. We obtain $I'_1 = (I_1 - \{\langle bus, at\_airport\rangle\}) \cup \{\langle taxi, at\_airport\rangle\}$. Now, $ben(I'_1) = 1100 - 520 = 580$.

Proposition 7 *For all RBDI structures $\mathcal{S} = \langle B, D, I, v, (c, C)\rangle$ and intentions $i = \langle \alpha, \theta\rangle$ such that $B \not\vdash \theta$ and $B \vdash r_{\alpha,\theta}$ it is always possible to* **add an intention (C)** *of i.*

For the last case we present here, we suppose that the agent does not include $i = \langle \alpha, \theta\rangle$ as an intention because achieving θ does not help in attaining a desire. The solution we adopt in **add an intention (D)** is to add an appropriate desire and adjust the value function v so that this desire has a high value. The net result is to make i a rational choice of intention. As with **add an intention (C)**, the actual choice of value we choose for the added desire to ensure that i is a rational intention is arguably arbitrary: any value could be used for θ as long as $\mathcal{S}'$ satisfies the requirements, making i a rational intention.

add an intention (D) $\langle B, D, I, v, (c, C)\rangle \oplus^D i(= \langle \alpha, \theta\rangle (\in R)) = \mathcal{S}' = \langle B', D', I', v', (c', C')\rangle$
$(I \oplus^D 1)$ $\mathcal{S}'$ is an RBDI structure.
$(I \oplus^D 2)$ $B' = B$.
$(I \oplus^D 3)$ $D' = D \cup \{\theta\}$.
$(I \oplus^D 4)$ $i \in I'$ and $(*)^{+i}$.
$(I \oplus^D 5)$ For all $S \subseteq D \cup \{\theta\}$

$$v'(S) = \begin{cases} v(S) & \text{if } \theta \notin S \\ v(S - \theta) + v(D) & \text{if } \theta \in S \end{cases}$$

$(I \oplus^D 6)$ $(c', C') = (c, C)$.

Example 3.6 Again, we go back to Example 3.1 starting with intention set I_1. Suppose the agent's boss tells him to call a customer in order to set up a meeting. This requires adding an intention $\langle call, meeting\rangle$ which is an **add an intention (D)** case. The agent will add to its desires list the desire $meeting$ resulting in $D' = \{at\_london, \neg hungry, good\_meal, meeting\}$. He will also update the v function: $v(\{meeting\}) = 1100$ which is the value of $v(\{at\_london, \neg hungry, good\_meal\})$.

Furthermore the value of all subsets of D with $meeting$ will be added. For example, $v(\{meeting, \neg hungry\}) = 1100 + 35 = 1135$. So the new intention set will be $I_1 \cup \{\langle call, meeting\rangle\}$ with *ben* value $2200 - 590 = 1610$.

Proposition 8 *For all RBDI structures $\mathcal{S}$, $\mathcal{S} = \langle B, D, I, v, (c, C)\rangle$, and intentions $i = \langle \alpha, \theta\rangle$ such that $\alpha \in C$, $B \not\vdash \theta$ and $B \vdash r_{\alpha,\theta}$ it is always possible to* **add an intention (D)** *of i.*

3.2 Updating cost and valuation functions

Updating the cost and the valuation functions is relatively straightforward. However, it is important to note that not all updates are possible; the changed functions must satisfy all the constraints of RBDI structures, such as the entailment-value condition for the valuation function. We start by considering three operations on the cost function: enlarging the domain, reducing the domain, and changing the value.

The enlarging the domain operation obtains an action a that is not in C and a function that associates a cost with any subset of C together with a. It updates the cost function accordingly, keeping the other parts of the cost function unchanged. This may lead to a change of the intention set: it may be beneficial to perform the new action.

enlarge the domain of the c function $\langle B, D, I, v, (c, C)\rangle \oplus (e, a) = \mathcal{S}' = \langle B', D', I', v', (c', C')\rangle$ where $a \in A - C$, and $e : K \to \mathcal{R}^+ \cup \{0\}$ for all $K \subseteq C \cup \{a\}$ such that $a \in K$.
$(C \oplus 1)$ $\mathcal{S}'$ is an RBDI structure.
$(C \oplus 2)$ $B' = B$.
$(C \oplus 3)$ $D' = D$
$(C \oplus 4)$ (*)
$(C \oplus 5)$ $v' = v$
$(C \oplus 6)$ $C' = C \cup \{a\}$ and

$$c'(K) = \begin{cases} e(K) & \text{if } a \in K \\ c(K) & \text{if } a \notin K \end{cases}$$

The postulate for reducing the domain operation is the reverse of the previous one. Given an action in C it deletes this action and restrict the cost function accordingly. This may lead to a change in the intention set in case the deleted action was in the original intention set.

reduce the domain of the c function $\langle B, D, I, v, (c, C)\rangle \ominus b = \mathcal{S}' = \langle B', D', I', v', (c', C')\rangle$ where $b \in A - actions(I)$
$(C \ominus 1)$ $\mathcal{S}'$ is an RBDI structure.
$(C \ominus 2)$ $B' = B$.
$(C \ominus 3)$ $D' = D$
$(C \ominus 4)$ (*)
$(C \ominus 5)$ $v' = v$.
$(C \ominus 6)$ $c' = c$ restricted to $C' = C - \{b\}$.

Example 3.7 We return to Example 2.3. Suppose the agent gets an e-mail specifying that sending a fax in the airport costs 20 and both calling and sending a fax costs 60. The costs of the other sets including fax is the sums of the costs. That is, C is enlarged with fax and $e(\{fax\}) = 20$, $e(\{fax, call\}) = 60$ and, for example $e(\{fax, bus\}) = 40$.

Suppose further that he is told that there is no bus going to the airport. That is, *bus* is removed from A and hence from C and the cost function is updated accordingly, e.g., $c(\{fax, bus\}) = 40$ is removed.

The changing of the value of the cost function operator gets a partial cost function. It changes the value associated with the subsets of C that are specified in the new function and keeps the rest of the function unchanged. This may lead to a change in the intention set since a set of actions may become cheaper to perform.

change the value of the c function $\langle B, D, I, v, (c, C)\rangle \circ b = \mathcal{S}' = \langle B', D', I', v', (c', C')\rangle$ where $b : A' \to \mathcal{R}^+ \cup \{0\}$, $A' \subseteq \mathcal{P}(C)$.
$(C \circ 1)$ $\mathcal{S}'$ is an RBDI structure.
$(C \circ 2)$ $B' = B$.
$(C \circ 3)$ $D' = D$
$(C \circ 4)$ (*)
$(C \circ 5)$ $v' = v$.
$(C \circ 6)$ $C' = C$ and

$$c'(E) = \begin{cases} b(E) & \text{if } E \in A' \\ c(E) & \text{if } E \notin A' \end{cases}$$

Updating the cost function may be more reasonable than adding an intention when **add an intention (A)** is not possible. Consider the following example.

Example 3.8 We revise Example 3.5 and consider the case that the agent's boss tells him that taking a taxi costs only 10 (due to a deal reached with the taxi company). Then, the intention set will be updated resulting with the same intention set as in that example and I_1' is obtained. But now $ben(I_1') = 1100 - 530 = 570$, reflecting the cost of the taxi rather than assuming that this cost is zero.

In the case of the v function recall that the domain is $\mathcal{P}(D)$. Hence, without changing D we can neither enlarge nor reduce the domain. The only update is to change the value of the v function in some way.

change the value of the v function $\langle B, D, I, v, (c, C)\rangle \circ w = \mathcal{S}' = \langle B', D', I', v', (c', C')\rangle$ where $w : F' \to \mathcal{R}^+ \cup \{0\}$, $F' \subseteq \mathcal{P}(D)$, and v' as given below in $(V \circ 5)$ satisfies (*).
$(V \circ 1)$ $\mathcal{S}'$ is an RBDI structure.
$(V \circ 2)$ $B' = B$.
$(V \circ 3)$ $D' = D$
$(V \circ 4)$ (*)
$(V \circ 5)$

$$v'(G) = \begin{cases} w(G) & \text{if } G \in F' \\ v(G) & \text{if } G \notin F' \end{cases}$$

$(V \circ 6)$ $(c', C') = (c, C)$.

3.3 Removing beliefs, desires, and intentions

Removing beliefs and intentions will also lead to changes in the intention set. For removing beliefs the agent will follow a predefined AGM-style $\dot{-}$ operation.

remove a belief $\langle B, D, I, v, (c, C)\rangle \ominus f = \mathcal{S}' = \langle B', D', I', v', (c', C')\rangle$
where
$(B \ominus 1)$ $\mathcal{S}'$ is an RBDI structure.
$(B \ominus 2)$ $B' = B \dot{-} f$
$(B \ominus 3)$ $D' = D$
$(B \ominus 4)$ (*)
$(B \ominus 5)$ $v' = v$
$(B \ominus 6)$ $(c', C') = (c, C)$

Example 3.9 Let us return to Example 2.4. Suppose the agent is told that due to a weather condition it is not possible to fly to London. So it should remove $r_{fly,at\_airport \rightarrow at\_london}$ from its belief set. There are now two good possible intention sets: $I'_1 = \{\langle eat\_home, \neg hungry\rangle\}$ and $I'_2 = \{\langle take\_out, \neg hungry\rangle\}$ with $ben(I'_1) = ben(I'_2) = 35 - 20 = 15$. Assume the agent chooses I'_1.

Removing a desire will require an update of the value function.

remove a desire $\langle B, D, I, v, (c, C)\rangle \ominus d = \mathcal{S}' = \langle B', D', I', v', (c', C')\rangle$
where
$(D \ominus 1)$ $\mathcal{S}'$ is an RBDI structure.
$(D \ominus 2)$ $B' = B$.
$(D \ominus 3)$ $D' = D - \{d\}$
$(D \ominus 4)$ (*)
$(D \ominus 5)$ v' is the restriction of v to $\mathcal{P}(D')$.
$(D \ominus 6)$ $(c', C') = (c, C)$

Example 3.10 Again we return to Example 2.4. Suppose that C calls to say she no longer is in London. The agent would like to remove his desire to be at London. The result is the same as the one obtained in Example 3.9 but for a different reason.

Proposition 9 *For all RBDI structures $\mathcal{S}$ and $d \in F_0$, $ben(\mathcal{S}) = ben(\mathcal{S} \oplus \langle w, d\rangle \ominus d)$.*

Note that it is not always the case that $\mathcal{S} = \mathcal{S} \oplus \langle w, d\rangle \ominus d$ because there may be multiple intention sets with the same benefit as exemplified in Example 3.9.

Just as we can add an intention by changing various components of the BDI structure, we can delete an intention in various ways. The appropriate modification of the parsimony condition $(*)$ in this case is:

$(*)^{-i}$ For any RBDI structure $\mathcal{S}'' = \langle B', D', I'', v', (c', C')\rangle$, such that $i \notin I''$
1. $|I'' \cap I| < |I' \cap I|$; or
2. $|I'' \cap I| = |I' \cap I|$ and $|I'| \leq |I''|$.

Here we give the details only for the first two: the one where the simple removal of the intention is possible and the case where we remove $r_{\alpha,\theta}$ from the beliefs; so i is no longer feasible for the agent.

remove an intention (A) $\langle B, D, I, v, (c, C)\rangle \ominus^A i (= \langle \alpha, \theta\rangle (\in R)) = \mathcal{S}' = \langle B', D', I', v', (c', C')\rangle$
$(I \ominus^A 1)$ $\mathcal{S}'$ is an RBDI structure.
$(I \ominus^A 2)$ $B' = B$
$(I \ominus^A 3)$ $D' = D$.
$(I \ominus^A 4)$ $i \notin I'$ and $(*)^{-i}$
$(I \ominus^A 5)$ $v' = v$.
$(I \ominus^A 6)$ $(c', C') = (c, C)$.

Example 3.11 We return to Example 3.9 where the agent has an intention to eat at home. Suppose his wife tells him that he cannot eat at home. Again, because of an external constraint, the agent modifies his intentions. In this case he removes the intention $\langle eat\_home, \neg hungry\rangle$ from I_1'. The result will be the I_2' from that example with the same benefit as before.

Proposition 10 *For all RBDI structures $\mathcal{S}$ and intentions i, if* **add an intention (A)** *is possible, then it is possible to* **remove an intention (A)** *i from $\mathcal{S} \oplus^A i$ and $ben(\mathcal{S}) = ben(\mathcal{S} \oplus^A i) = ben(\mathcal{S} \oplus^A i \ominus^A i)$.*

Note however that the intention sets of $\mathcal{S}$ and $\mathcal{S} \oplus^A i \ominus^A i$ may be different because of the possibility of multiple intention sets with the same benefit. The second way to remove an intention, by removing a belief, is always possible.

remove an intention (B) $\langle B, D, I, v, (c, C)\rangle \ominus^B i = \langle \alpha, \theta\rangle (\in R) = \mathcal{S}' = \langle B', D', I', v', (c', C')\rangle$
$(I \ominus^B 1)$ $\mathcal{S}'$ is an RBDI structure.
$(I \ominus^B 2)$ $B' = B \dot{-} r_{\alpha,\theta}$
$(I \ominus^B 3)$ $D' = D$.
$(I \ominus^B 4)$ $i \notin I'$ and $(*)^{-i}$
$(I \ominus^B 5)$ $v' = v$
$(I \ominus^B 6)$ $(c', C') = (c, C)$

Example 3.12 Let us consider a different version of Example 3.9. Instead of being told that there is a weather problem, the agent is just told not to fly to London. This is a **remove an intention (B)** case. The agent will remove from its belief set $r_{fly,at\_airport \rightarrow at\_london}$ resulting in the same intention set as in Example 3.9.

While it is always possible to remove an intention using definition **(B)**, **add an intention (B)** and **remove an intention (B)** are not opposite actions in the sense that $\mathcal{S} \oplus^B i \ominus^B i \neq \mathcal{S}$ is possible. In the next proposition, we provide a case where the equality holds even for $\oplus^B$.

Proposition 11 *For all RBDI structures $\mathcal{S} = \langle B, D, I, v, (c, C)\rangle$ and intentions $i = \langle \alpha, \theta\rangle$ such that $B \not\vdash \theta$ and $B \dot{+} r_{\alpha,\theta} \dot{-} r_{\alpha,\theta} = B$, $ben(\mathcal{S}) = ben(\mathcal{S} \oplus^B i \ominus^B i)$.*

4 Implementation issues

The previous section presented postulates for RBDI structure revision. However, we did not present any algorithms to accomplish these revisions. Observe first that for all

the revisions only the set of intentions, I', has to be constructed as all the other parts of the structure are given explicitly in the postulates. Actually, there is a rather obvious algorithm to construct such an I' for each revision. Let $K = \{I' \subset R \mid \text{ for each } i = \langle \alpha, \theta \rangle \in I', r_{\alpha,\theta} \in B'\}$. For each such $I' \in K$ check that $\mathcal{S}'$ is an RBDI structure. Let $\mathcal{T}$ be the set of all such RBDI structures. Finally, check condition (*) or its appropriate modification as given in the postulate for each element of $\mathcal{T}$ and pick as the answer one such RBDI structure.

It would of course be desirable to have an algorithm that is more efficient than such a brute force method. A natural approach is to consider possible constraints on the language L_0 of beliefs, so that the search space of possible updates is as small as possible. However, the following example demonstrates that even if we restrict ourselves only to considering the comparatively simply operation of belief revision, the search space can potentially include all possible sets of intentions. Intuitively this is because the calculation of the benefit of a BDI structure is highly sensitive to the value and cost functions. So even a high value for the desires achieved by a set of intentions may not be enough to circumvent a corresponding high cost of the actions, and conversely, a low cost of actions may be balanced by a low value for the achieved desires.

Example 4.1 Let $F_0 = \{p, q, p \rightarrow q, q \rightarrow p\}$, $A = \{\alpha_1, \alpha_2\}$ and $R = \{\langle \alpha_1, p \rangle, \langle \alpha_2, q \rangle\}$. Consider the RBDI structure $\mathcal{S} = \langle B, D, I, v, (c, C) \rangle$ where $B = \{q \rightarrow p, r_{\alpha_2,q}, r_{\alpha_1,p}\}$, $D = \{p, q\}$ $C = A$ and the value of a desire and the cost of an action is two and one, respectively, i.e., for any $S \subseteq D$, $v(S) = 2 * |S|$ and for any $A \subseteq C$, $c(A) = |A|$. Thus $I = \{\langle \alpha_2, q \rangle\}$. Suppose $\mathcal{S}$ is updated and the new RBDI structure is $\mathcal{S}' = \langle B', D', I', v', (c', C') \rangle$. The following 4 updates lead to all possible sets of intentions.

1. In $\mathcal{S} \oplus q$, $I' = \emptyset$.
2. In $\mathcal{S} \ominus q \rightarrow p$, $I' = \{\langle \alpha_2, q \rangle, \langle \alpha_1, p \rangle\}$.
3. In $\mathcal{S} \oplus p$, $I' = I$.
4. In $\mathcal{S} \ominus r_{\alpha_2,q}$, $I' = \{\langle \alpha_1, p \rangle\}$.

5 Summary and discussion

We introduce the concept of a BDI structure for capturing the beliefs, desires, and intentions of agents as well as the values the agent gives to desires and the costs the agent has for doing actions. The initial concept of BDI structure turns out to be very weak and we impose rationality axioms on them. In particular, we define a rational BDI structure to have an optimality condition involving values and costs. Then we present our postulates for many types of revisions to such structures. We end by commenting on algorithms for implementing revisions.

There are a number of obvious issues for further work. First, an important issue, not properly resolved in our or, to the best of our knowledge, any other work, is the relationship between "practical reasoning" models (such as the BDI model we discuss in this paper) and other models of rational decision making—in particular, *decision theory*. Decision theory is a mathematical theory of rational decision making. Decision theory defines a rational agent as one that *maximizes expected utility*. The most

common model of decision theory is broadly as follows. Assume A is the set of all possible actions available to an agent. The performance of an action by the agent may result in a number of possible outcomes, where the set of all outcomes is $\Omega = \{\omega, \omega', \ldots\}$. Let the probability of outcome $\omega \in \Omega$ given that the agent performs action $\alpha \in A$ be denoted by $P(\omega \mid \alpha)$, and finally, let the *utility* of an outcome $\omega \in \Omega$ for the agent be given by a function $U : \Omega \to \mathcal{R}^+$. If $U(\omega) > U(\omega')$, then the agent prefers outcome ω over outcome ω'.

The *expected utility* of an action $\alpha \in A$ is denoted $EU(\alpha)$. Thus $EU(\alpha)$ represents the utility that an agent could expect to obtain by performing action α.

$$EU(\alpha) = \sum_{\omega \in \Omega} U(\omega) P(\omega \mid \alpha) \tag{1}$$

According to this model of decision theory, a rational agent is one that chooses to perform an action α that maximizes $EU(\ldots)$.

Decision theory, expressed in this simple way, is a normative theory of rational action: it attempts to define, mathematically, the "best" action to perform, assuming the model given by P and U. Of course, this model does not claim to explain how people actually make decisions, and indeed it is not useful for this. Nor is the theory a computational one.

The relationship between BDI-like models and decision theory was discussed by Pollack in (1992). Pollack suggested that the basic difference between the two models is that decision theory seems to imply "continual optimisation": before every action, we compute expected utilities, and then perform the action that maximises this value. In contrast, Pollack suggested that intentions are not "continually optimised" in this way: we choose intentions and commit to them, implicitly assuming that they are optimal. We typically only infrequently reconsider our intentions, perhaps when we become aware that they are doomed, or that there is obviously a better alternative. Intentions in our present work play a role somewhat between the model of intentions as discussed by Pollack, and the model of continual optimisation proposed by decision theory. Clearly, there are intuitive links between the two types of models, and our model of BDI agents has quite a strong flavour of decision theory to it. Can our model (or some variant of it) serve as a bridge between the two frameworks? That is, can we produce a model that allows us to explain or reconcile decision theory and practical reasoning BDI-type models?

Another interesting question is the relationship of our work and our model to *implemented* BDI models (Georgeff and Lansky 1987; Rao and Georgeff 1991; Bordini et al. 2007). Typically, the procedures for updating intentions and desires in such implemented models are (necessarily!) quite simple. Can the (cost, value) models used here be of value within such systems, to give a deeper basis for revising intentions and desires? Can we reconcile the formal semantics of such systems (Bordini et al. 2007) with our framework?

Finally, another interesting question is to consider what happens when we put multiple agents together, and to extend the model to account for such interactions. Here, it becomes interesting to consider situations such as what happens when one agent drops an intention that another agent is relying upon; this might cause the

second agent to adjust its intentions in turn. We can envisage the ramifications of such a change rippling out throughout the agent society, with intentions being adjusted until a kind of equilibrium state is reached, in which no agent has any incentive to adjust its mental state further. Formulating such equilibria is an interesting issue.

Acknowledgements We thank the referees for their valuable comments, which led to substantial improvements to the paper. This research was supported by the NSF, the Air Force Office of Scientific Research, and the Office of Naval Research.

References

Alchourron, C. E., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic, 50*, 510–530.

Baral, C., & Zhang, Y. (2005). Knowledge updates: Semantics and complexity issues. *Artificial Intelligence, 164*(1–2), 209–243.

Bordini, R., Hübner, J. F., & Wooldridge, M. (2007). *Programming multi-agent systems in AgentSpeak using Jason*. Chichester: Wiley.

Cohen, P. R., & Levesque, H. J. (1990). Intention is choice with commitment. *Artificial Intelligence, 42*, 213–261.

Dunne, P. E., Wooldridge, M., & Laurence, M. (2005). The complexity of contract negotiation. *Artificial Intelligence, 164*(1–2), 23–46.

Gärdenfors, P. (1988). *Knowledge in flux*. Cambridge, MA: The MIT Press.

Genesereth, M. R., & Nilsson, N. (1987). *Logical foundations of artificial intelligence*. San Mateo, CA: Morgan Kaufmann Publishers.

Georgeff, M. P., & Lansky, A. L. (1987). Reactive reasoning and planning. In *Proceedings of the sixth national conference on artificial intelligence (AAAI-87)* (pp. 677–682). Seattle, WA.

Georgeff, M. P., & Rao, A. S. (1995). The semantics of intention maintenance for rational agents. In *Proceedings of the fourteenth international joint conference on artificial intelligence (IJCAI-95)* (pp. 704–710). Montrèal, Quèbec, Canada.

Huber, M. (1999). JAM: A BDI-theoretic mobile agent architecture. In *Proceedings of the third international conference on autonomous agents (Agents 99)* (pp. 236–243). Seattle, WA.

Konolige, K. (1986). *A deduction model of belief*. San Mateo, CA: Pitman Publishing: London and Morgan Kaufmann.

Pollack, M. E. (1990). Plans as complex mental attitudes. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 77–104). Cambridge, MA: The MIT Press.

Pollack, M. E. (1992). The uses of plans. *Artificial Intelligence, 57*(1), 43–68.

Rao, A. S. (1996). AgentSpeak(L): BDI agents speak out in a logical computable language. In W. Van de Velde & J. W. Perram (Eds.), *Agents breaking away: Proceedings of the seventh European workshop on modelling autonomous agents in a multi-agent world, (LNAI Volume 1038)* (pp. 42–55). Berlin, Germany: Springer-Verlag.

Rao, A. S. & Georgeff, M. P. (1991). Modeling rational agents within a BDI-architecture. In R. Fikes & E. Sandewall (Eds.), *Proceedings of knowledge representation and reasoning (KR&R-91)* (pp. 473–484). San Mateo, CA: Morgan Kaufmann Publishers.

Rao, A. S. & Georgeff, M. P. (1992) An abstract architecture for rational agents. In C. Rich, W. Swartout, & B. Nebel (Eds.), *Proceedings of knowledge representation and reasoning (KR &R-92)* (pp. 439–449).

Rao, A. S., & Georgeff, M. P. (1998). Decision procedures for BDI logics. *Journal of Logic and Computation, 8*(3), 293–344.

Shoham, Y. (2009). Logical theories of intention and the database perspective. *Journal of Philosophical Logic, 38*(6), 633–648.

van der Hoek, W., Jamroga, W., & Wooldridge, M. (2007). Towards a theory of intention revision. *Synthese, 155*(2), 265–290.

Wooldridge, M. (2000). *Reasoning about rational agents*. Cambridge, MA: The MIT Press.