Heterogeneous BDI Agents

Maria Fasli

University of Essex, Department of Computer Science, Wivenhoe Park, Colchester
CO4 3SQ, UK
mfasli@essex.ac.uk

Abstract. The study of formal theories of agents has intensified over the last decade since such formalisms can be viewed as providing the specifications for building agent-based systems. One such theory views agents as having beliefs, desires and intentions (BDI). The BDI paradigm provides us with the means of describing different types of agents; a desirable quality, since agent-based systems are employed in various domains with diverse characteristics and therefore different requirements. This is accomplished by adopting a set of constraints that describe how the three attitudes are related to each other, called a notion of realism. Although three such notions have been explored in the literature, namely *strong realism* , *realism* and *weak realism*, no systematic attempt has been undertaken to study other available options. In this paper we explore the dynamics and possible interrelations between the three attitudes and we propose notions of realism for heterogeneous BDI agents. We explore a more wide range of possibilities by considering a combination of the types of relations between accessible worlds. Moreover, we distinguish between two broad categories of agents, bold and circumspect, according to the relation between beliefs and intentions. We explore several interesting notions of realism for such agents and we argue that these come close to the desiderata for rational BDI agents.

Keywords: agent modeling, dynamics of mental attitudes, BDI logics

1 Introduction

As increasingly sophisticated systems are built based on the notions of an agent and a multi-agent system, the need for adequate theories that will be able to describe, predict, and explain the behaviour of such systems is increasing accordingly. These theories can then serve as specification and validation tools to the designers of multi-agent systems. Such theories are based on a mentalistic view of the agents, that is artificial agents are intentional systems [7] that have information about the world as well as objectives which they attempt to accomplish by performing actions. Although the issue of ascribing human attitudes to computational systems is debatable [22], the intentional stance does seem to provide us with a powerful abstraction tool for explaining the behaviour of such systems.

Naturally, philosophical theories of the mind and practical reasoning have been the major source of insight into the issue of formalising the properties of rational agents. Among the most influential works in the area of practical reasoning is that of Bratman [3]. Bratman argued that intentions play a prominent role in an agent's decision making and based on his work one of the most well known frameworks for agents has been developed: the Belief-Desire-Intention (BDI) paradigm [25, 27]. In accordance, agents are ascribed beliefs, desires and intentions. The BDI paradigm provides us with the means of describing different types of agents. This is accomplished by adopting a set of constraints that describe how the three attitudes are related to each other. This set of constraints is called a notion of realism. Three such notions have been explored in the literature: *strong realism*, *realism* and *weak realism* characterising a cautious, an enthusiastic and a balanced agent respectively [27]. This diversity in the type of agents that can be described is actually essential, considering the fact that agent-based systems have been used in implementing a variety of applications ranging from e-commerce to space mission control for which the domain characteristics can be different. As Rao and Georgeff [27] note, there may not be a unique BDI system suitable for all applications, since different domains have different characteristics and thus different requirements regarding rational behaviour.

Apart from the original notions of realism considered by Rao and Georgeff, there has been no systematic study of other available options. More recently, Wooldridge [33] discussed and clarified a lot of the issues surrounding the BDI paradigm such as for instance the reading of the BDI modalities. However, in his work he only considers the original notions of realism and no analysis or consideration of further options is provided. This paper builds on the work of Rao and Georgeff [27] and addresses the issue of the dynamics between the three attitudes and the different types of agents that can be described. Their work is extended further in two ways. Firstly, we consider a combination of relations between the three attitudes. Rao and Georgeff only considered uniform relations such as the three attitudes being related via the subset or the intersection relation between the respective sets of accessible worlds. In particular, we categorise notions of realism according to set relations between accessible worlds that can be adopted, and catalogue their properties. Secondly we distinguish between two broad categories of agents: circumspect and bold. This distinction is according to the relation between the agent's beliefs and intentions. The aim is to engineer suitable specifications for agents that come close to the desiderata for rational reasoning agents [3, 27].

Although the paper uses the BDI paradigm, it does not focus on the merits of the logical framework itself. This is used as a means to an end. The main focus is on exploring the dynamics between the attitudes and the modeling of heterogeneous agents in this framework.

2 From Cognitive Ingredients to Rational Agents

Although increasingly sophisticated systems are built based on the notion of an agent, this very same term meets no uniform definition. The most commonly understood and acceptable description is that of an agent being a computational system capable of exhibiting autonomous, reactive, proactive and social behaviour [32]. Formalising theories that will explain the behaviour of such a complex system in a natural, intuitive and efficient way is a non-trivial task. The intentional notions or otherwise known as propositional attitudes are not problem-free [12, 23, 24, 28, 30]. Nevertheless, the use of the intentional stance for explaining the behaviour of artificial agents seems to be the only method that works. In developing formalisms for representing the properties of agents, agent theorists are faced with two key issues: which are the attitudes that constitute an agent's cognitive state, and what are the relations and dynamics between these attitudes.

The first fundamental problem is to decide which combination of attitudes is appropriate for characterising an agent's cognitive state. Information attitudes express an agent's information about the world, while pro-attitudes guide an agent's actions. The nature of information attitudes such as knowledge and beliefs and pro-attitudes such as desires and wishes is distinctively different. The former have a mind-to-world direction of fit, while the latter a world-to-mind one [29]. Assuming that the agent's cognitive state consists of the information, motivation and deliberation states, which are the correct attitudes to represent them? There is no definitive answer to this question.

Undoubtedly, at least one attitude expressing the agent's information about the world is required. This is usually that of belief or knowledge [5, 16], although which one is the most appropriate is far from clear. Even though one information attitude is considered in general to be adequate, a lot of effort has been put into investigating theories that include both knowledge and belief [15, 19, 31].

The motivation element of an agent is usually described by pro-attitudes such as desires, goals and preferences [5, 21, 25, 27]. Desires are regarded as expressing states of affairs that the agent would prefer to be in, or perhaps the ideal states for the agent, or the agent's options. They represent a tendency or an impulse of the agent towards a state. Goals are often taken to be consistent desires [5]. Wishes have also been considered as an agent's primary motivation attitude since they are considered to express how the agent would like the world to be [16].

The deliberation state of the agent is often represented by intentions which are pro-attitudes as well [25, 27]. Intentions describe states of affairs that the agent is actually committed to bringing about. An intention constitutes reason for action, it is a conscious wish to carry out an act, and a philosophical theory of actions must include an account of what is for an agent to do something intentionally [1, 3, 13]. Searle [29] considers the content of an intention to be a causally self-referential representation of its conditions of satisfaction. Undoubtedly, intentions are very closely related with beliefs and desires [1, 6]. However, some philosophers support a stronger reductive approach according to which intentions are not primitive propositional attitudes but they can be reduced to their constituents beliefs and desires [2]. Whether or not such a reduction of intentions to beliefs and desires can be vindicated is the subject of debate. Searle [29] argues that this may not be possible due to the special causal self-referentiality of intentions. In his philosophical investigation Bratman [3] claims that the notion of intention is distinct and cannot be reduced to those of beliefs and desires. Challenging the belief-desire model for explaining rational behaviour, he maintains that intentions play a major role in an agent's practical reasoning by being conduct-controlling and not simply potential influencers of behaviour as desires usually are. Bratman's philosophical work has been extremely influential. Perhaps the most well known framework for agents, the BDI paradigm [27, 25], is based on his work and reflects his ideas on practical reasoning. In this approach, an agent's cognitive state is described in terms of beliefs, desires and intentions representing the information, motivation and deliberation aspects respectively.

The second fundamental problem in developing a theory of rational reasoning agents is to give an account of the relationships and dynamics between an agent's cognitive ingredients. Deciding on a set of attitudes to represent an agent's cognitive state is not enough. In particular, a complete agent theory would have to explain how an agent's cognitive ingredients lead it to select sequences of actions (plans) and act upon them. Thus, rational behaviour should be the result of the interaction of an agent's cognitive ingredients. However, we do not have a universally accepted theory to draw upon; the dynamics and interrelationships between the various attitudes are far from clear. With notions such as belief, desires, intentions or knowledge we mostly rely on intuitions. Furthermore, the greater the number of attitudes one considers, the more complicated their interrelations. In essence, providing rules that define the dynamics of attitudes that result in rational behaviour is extremely difficult. As a consequence, the theories that we build are very difficult to validate. This task becomes more complicated by the fact that the kind of behaviour that we would expect from a rational agent depends very much upon the particular application and the domain characteristics. As pointed out by Rao and Georgeff [27], there may not be a unique type of agent suitable for all applications, since different domains have different characteristics and thus different requirements regarding rational behaviour.

3 The Logical Framework

The following sections present the logical framework which is based on the BDI paradigm, albeit with a few minor modifications. The interested reader is referred to [25, 27] for the fully-fledged details of the original BDI framework.

3.1 Syntax

The logical language $\mathcal{L}$ is a many-sorted first order language which enables quantification over two sorts of individuals, namely *Agents* and *Other*. The former denotes the set of individual agents, while the latter indicates all the other individuals/objects in the universe of discourse. The main constructs of the language include the standard connectives for negation ($\neg$), disjunction ($\vee$), equality ($=$) and the first-order quantifier ($\forall$). *true* is taken to be an abbreviation for some fixed propositional tautology. $\mathcal{L}$ is augmented with modal operators that enable

Table 1. The syntax of $\mathcal{L}$

< agent-term >	::=	<agent-var > | < agent-const >
<other-term>	::=	<other-var> | < other-const >
<term>	::=	<agent-term> | <other-term>
<pred-symbol>	::=	An element of the set of Pred symbols
<var>	::=	<agent-var>| <other-var>
<state-wff>	::=	< pred-symbol > (<term>,...,<term>)| *Bel*(< agent-term >, < state-wff >)| *Des*(<agent-term>,<state-wff>)| *Intend*(<agent-term>, <state-wff>)| (<term>=<term>)| ¬ <state-wff> | <state-wff>∨<state-wff> | ∀<var><state-wff> | *A* <path-wff>
<path-wff>	::=	<state-wff> |¬ < path-wff> | <path-wff>∨<path-wff>| ∀ <var><state-wff > |*X*<path-wff> | <path-wff> *U* <path-wff>

the expression of an agent's cognitive state and the dynamics of the environment. The operators Bel, Des and $Intend$ express the agents' beliefs, desires and intentions respectively. The temporal operators A (universal path quantifier, or inevitable), X (next), and U (until) express properties over time. This branching temporal component is based on CTL logic [8]. The detailed syntax of the language is provided in Table 1.

The standard abbreviations from propositional logic for the connectives $\wedge$, $\Rightarrow$ and $\Leftrightarrow$ are adopted and $false$ is taken to be an abbreviation for $\neg true$. Furthermore, the existential quantifier $\exists$ is defined in terms of the universal quantifier as usual, $\exists x\phi \equiv \neg\forall x\neg\phi$. Finally, the operators E (existential path quantifier, or optional), F (sometimes), and G (always) are defined in terms of the primitive temporal operators [8] as follows:
$E(\phi) \equiv \neg A(\neg\phi)$
$F(\phi) \equiv trueU\phi$
$G(\phi) \equiv \neg F(\neg\phi)$

3.2 Semantics

Semantics to the language is given in terms of possible worlds [14] and Kripke structures [20]. However, the possible worlds for these models are not flat structures but have a branching time nature. A model for $\mathcal{L}$ is a structure $M =< W, T, \prec, \mathcal{U}, \mathcal{B}, \mathcal{D}, \mathcal{I}, \pi >$ where W is a set of worlds, T is a set of time points, $\prec$ is a total, backwards-linear branching time relation on time points, $\mathcal{U}$ is the universe of discourse which is a tuple itself $\mathcal{U} =< \mathcal{U}_{Agents}, \mathcal{U}_{Other} >$, $\mathcal{B}$ is the belief accessibility relation $\mathcal{B} : \mathcal{U}_{Agents} \rightarrow \wp(W \times T \times W)$. The accessibility relation $\mathcal{B}$ defines which worlds are possible for each agent $i \in \mathcal{U}_{Agents}$, thus $\mathcal{B}_i(w, t, w')$ denotes that if an agent i is in world w at time point t, then from this world w'

is accessible, or possible according to the agent's beliefs. Similarly, $\mathcal{D}$ and $\mathcal{I}$ are the desire and intention accessibility relations. Finally π interprets the atomic formulas of the language.

The belief-, intention-, and desire-accessible worlds are themselves branching time structures. Time has a single linear past while the future is branching and this reflects the uncertainty in the agent's choices. Thus, a path is a possible future. For instance, if the starting point is t_0, then there may be two futures possible: $t_0, t_1, t_2, t_4, t_5, ...$ and $t_0, t_1, t_3, t_6, t_7, ...$ Due to the branching-time nature of possible worlds, there are two types of formulas in $\mathcal{L}$: state and path formulas. The former are evaluated in a particular world in a particular point in time, while the latter are evaluated in a particular world along a certain path. A path $t_0, t_1, ...$ in a world w will be denoted $w, t_0, t_1, ...$ A fullpath in a world w is an infinite sequence of time points $(w, t_0, t_1, ...)$ [25].

The semantics for state formulas is provided below:
$M(v, w, t) \models P(\tau_1, ..., \tau_k)$ iff $< v(\tau_1), ..., v(\tau_k) > \in \pi(P^k, w, t)$
$M(v, w, t) \models \neg\phi$ iff $M(v, w, t) \not\models \phi$
$M(v, w, t) \models \phi \vee \psi$ iff $M(v, w, t) \models \phi$ or $M(v, w, t) \models \psi$
$M(v, w, t) \models (\tau_1 = \tau_2)$ iff $\| \tau_1 \| = \| \tau_2 \|$
$M(v, w, t) \models \forall x(\phi)$ iff for $d \in \mathcal{U}$ such that x and d are of the same sort, we have $M(v[d/x], w, t) \models \phi$
$M(v, w, t) \models Bel(i, \phi)$ iff $\forall w, w', t$ s.t. $\mathcal{B}_i(w, t, w')$, we have $M(v, w', t) \models \phi$
$M(v, w, t) \models Des(i, \phi)$ iff $\forall w, w', t$ s.t. $\mathcal{D}_i(w, t, w')$, we have $M(v, w', t) \models \phi$
$M(v, w, t) \models Intend(i, \phi)$ iff $\forall w, w', t$ s.t. $\mathcal{I}_i(w, t, w')$, we have $M(v, w', t) \models \phi$
$M(v, w, t_0) \models A(\phi)$ iff for all fullpaths $(w, t_0, t_1, ...)$ s.t. $M(v, w, t_0, t_1, ...) \models \phi$
$M(v, w, t_0) \models E(\phi)$ iff there exists a fullpath $(w, t_0, t_1, ...)$ s.t. $M(v, w, t_0, t_1, ...)$ $\models \phi$

The clause for belief states that an agent i believes ϕ, iff ϕ is true in all its belief-accessible worlds at time point t. The operator A is said to be true of a path formula ϕ at a particular point in a time-tree if ϕ is true of all paths emanating from that point. Since the existential path quantifier E is defined as the dual of the universal path quantifier A, a formula of the form $E(\phi)$ is interpreted as "on some path, ϕ is true", that is the formula will be true in some time point t if there is at least one path emanating from t such that the path formula ϕ is true on this path.

The semantics for path formulas is as follows:
$M(v, w, t_0, t_1, ...) \models \phi$ iff $M(v, w, t_0) \models \phi$
$M(v, w, t_0, t_1, ...) \models \neg\phi$ iff $M(v, w, t_0, t_1, ...) \not\models \phi$
$M(v, w, t_0, t_1, ...) \models \phi \vee \psi$ iff $M(v, w, t_0, t_1, ...) \models \phi$ or $M(v, w, t_0, t_1, ...) \models \psi$
$M(v, w, t_0, t_1, ...) \models \forall x(\phi)$ iff for $d \in \mathcal{U}$ such that x and d are of the same sort, we have $M(v[d/x], w, t_0, t_1, ...) \models \phi$
$M(v, w, t_0, t_1, ...) \models \phi U \psi$ iff $\exists k$, $k \geqslant 0$ such that $M(v, w, t_k, ...) \models \psi$ and $\forall j$, $0 \leqslant j \leqslant k$ then $M(v, w, t_j, ...) \models \phi$
$M(v, w, t_0, t_1, ...) \models X(\phi)$ iff $M(v, w, t_1, t_2, ...) \models \phi$

For instance, the semantics of the formula $X\phi$ states that a formula ϕ is true at the next point in time along a path $w, t_0, t_1, ...$ if ϕ iff it is true along the path $w, t_1, t_2, ...$

From now on, wffs that contain no positive occurrences of A outside the scope of the modal operators Bel, Des and $Intend$ will be called I-formulas (inevitable formulas), while wffs that contain no positive occurrences of E outside the scope of these operators will be called O-formulas (optional formulas) [27].

3.3 Basic BDI Axiom System

For the Bel operator we adopt the standard KD45 (weak S5) modal system:

B-K. $Bel(i, \phi) \wedge Bel(i, \phi \Rightarrow \psi) \Rightarrow Bel(i, \psi)$

B-D. $Bel(i, \phi) \Rightarrow \neg Bel(i, \neg\phi)$

B-S4. $Bel(i, \phi) \Rightarrow Bel(i, Bel(i, \phi))$

B-S5. $\neg Bel(i, \phi) \Rightarrow Bel(i, \neg Bel(i, \phi))$

B-Nec. if $\vdash \phi$ then $\vdash Bel(i, \phi)$

Thus, we require the accessibility relation for belief $\mathcal{B}$ to be serial, transitive and Euclidean [4, 9]. The K axiom and the Necessitation rule are inherent of the possible worlds approach and they hold in normal modal logics regardless of any restrictions that we may impose on the accessibility relations. Hence, agents are logically omniscient with respect to their attitudes [9]. The D axiom expresses the consistency of beliefs, that is not both ϕ and $\neg\phi$ are true at the same time. The S4 and S5 axioms express the agent's positive and negative introspective capabilities regarding its beliefs. Formal proof of the correspondence of the properties of the accessibility relation with the respective axioms can be found in [9, 4, 17]. For both desires and intentions we adopt the D system:

Desires

D-K. $Des(i, \phi) \wedge Des(i, \phi \Rightarrow \psi) \Rightarrow Des(i, \psi)$

D-D. $Des(i, \phi) \Rightarrow \neg Des(i, \neg\phi)$

D-Nec. if $\vdash \phi$ then $\vdash Des(i, \phi)$

Intentions

I-K. $Intend(i, \phi) \wedge Intend(i, \phi \Rightarrow \psi) \Rightarrow Intend(i, \psi)$

I-D. $Intend(i, \phi) \Rightarrow \neg Intend(i, \neg\phi)$

I-Nec. if $\vdash \phi$ then $\vdash Intend(i, \phi)$

Hence, the accessibility relations $\mathcal{D}$ and $\mathcal{I}$ respectively are required to be serial. The D axiom expresses the consistency of desires and intentions. Although in the philosophical literature desires are allowed to be inconsistent and therefore they seem to tag along the agent towards different paths of action, here we will assume that desires are consistent. The CTL axiomatisation can be found in [8, 27]. The axioms for the three attitudes along with the CTL axiomatisation constitute the basic BDI system.

4 Relations between Modalities

Following [27] we can define relations between the three attitudes along two dimensions in the BDI paradigm: by imposing restrictions on the relationships between sets of accessible worlds, and by imposing restrictions on the structure of the worlds.

4.1 Set Relations

Since the modalities are underpinned by their respective sets of accessible worlds, it seems reasonable to consider relations between the three sets in order to establish relations between the modalities. Thus, if X and Y are the sets of accessible worlds characterising any two modalities Xm and Ym, the following relations can hold:

- X is a subset of Y, $X \subseteq Y$
- Y is a subset of X, $Y \subseteq X$
- The intersection of X and Y is not the empty set, $X \cap Y \neq \emptyset$
- The intersection of X and Y is the empty set, $X \cap Y = \emptyset$

For the first type of relation if X_i and Y_i are the corresponding accessibility relations, formally we have the following semantic condition:
$X_i(w,t) \subseteq Y_i(w,t)$: $\forall\ w, w', t$ if $X_i(w,t,w')$ then $Y_i(w,t,w')$
This condition yields an axiom schema of the form $Ym(i,\phi) \Rightarrow Xm(i,\phi)$ between the modalities.

Proposition 1. *Assume that Xm and Ym are the two modalities defined by the two accessibility relations X_i and Y_i respectively. Then if $X_i(w,t) \subseteq Y_i(w,t)$ we have that $Ym(i,\phi) \Rightarrow Xm(i,\ \phi)$ is valid.*

Proof. Assume that $M(v,w,t) \models Ym(i,\phi)$ for an arbitrary $M(v,w,t)$. According to semantics we have $M(v,w',t) \models \phi$ for all w' such that $Y_i(w,t,w')$. Since $X_i(w,t) \subseteq Y_i(w,t)$ we have $M(v,w'',t) \models \phi$ for all w'' such that $X_i(w,t,w'')$. It now follows that $M(v,w,t) \models Xm(i,\phi)$. □

The second type of relation is semantically captured as follows:
$Y_i(w,t) \subseteq X_i(w,t)$: $\forall\ w, w', t$ if $Y_i(w,t,w')$ then $X_i(w,t,w')$
As a result the axiom schema $Xm(i,\phi) \Rightarrow Ym(i,\phi)$ relates the two modalities.

Proposition 2. *Assume that Xm and Ym are the two modalities defined by the two accessibility relations X_i and Y_i respectively. Then if $Y_i(w,t) \subseteq X_i(w,t)$ we have that $Xm(i,\phi) \Rightarrow Ym(i,\ \phi)$ is valid.*

Proof. Similar to that of Proposition 1. □

The third type of relation semantically requires the following:
$X_i(w,t) \cap Y_i(w,t) \neq \emptyset$: $\forall\ w, t\ \exists w'\ Y_i(w,t,w')$ and $X_i(w,t,w')$
This condition corresponds to the axiom schema $Xm(i,\phi) \Rightarrow \neg Ym(i,\neg\phi)$.

Proposition 3. *Assume that Xm and Ym are the two modalities defined by the two accessibility relations X_i and Y_i respectively. Then if $X_i(w,t)\ \cap Y_i(w,t) \neq \emptyset$ we have that $Xm(i,\phi) \Rightarrow \neg Ym(i,\neg\phi)$ is valid.*

Proof. Assume $M(v, w, t) \models Xm(i, \phi)$. Then for all w' such that $X_i(w, t, w')$ we have $M(v, w', t) \models \phi$ and thus $M(v, w', t) \not\models \neg\phi$. Since $X_i(w, t) \cap Y_i(w, t) \neq \emptyset$, there is at least one world w' such that $Y_i(w, t, w')$. It now follows that $M(v, w, t) \not\models Ym(i, \neg\phi)$ and hence $M(v, w, t) \models \neg Ym(i, \neg\phi)$ as required. □

The fourth relation intuitively means that the two sets of accessible worlds are completely decoupled and therefore there is no relation between the modalities. However, we do not consider this last type of relation to yield any interesting properties for our purposes, and for that reason it will not be given any further consideration.

The above generic propositions can be used in order to reproduce the results of the following sections.

4.2 Pairwise BDI Properties

Following the results from the previous section, here we present the pairwise properties between the three attitudes starting with those that ensue from considering subset relations. Beliefs and desires are considered first. The first case is that the set of desire-accessible worlds is a subset of the belief-accessible worlds, that is $\mathcal{D}_i(w, t) \subseteq \mathcal{B}_i(w, t)$. By substituting $\mathcal{D}_i$ and $\mathcal{B}_i$ for X_i and Y_i in Proposition 1, we obtain the following property:

$Bel(i, \phi) \Rightarrow Des(i, \phi)$

Semantically, if an agent believes ϕ, then ϕ is true in all the belief-accessible worlds. Since the desire-accessible worlds are a subset of the belief-accessible worlds, this means that ϕ will also be true in the desire-accessible worlds. Therefore, if an agent believes ϕ, then it desires it as well.

Conversely, the set of belief-accessible worlds can be a subset of the desire-accessible worlds, that is $\mathcal{B}_i(w, t) \subseteq \mathcal{D}_i(w, t)$. By using the result of Proposition 1, the ensuing property is:

$Des(i, \phi) \Rightarrow Bel(i, \phi)$

Hence, if an agent desires ϕ, then semantically ϕ is true in all the desire-accessible worlds. Since the belief-accessible worlds are a subset of these worlds, then it follows that ϕ is true in the belief-accessible worlds as well. Therefore, if an agent desires ϕ, then it believes it.

Similarly, there are two cases for intentions and beliefs: (i) the set of intention-accessible worlds is a subset of the belief-accessible worlds, $\mathcal{I}_i(w, t) \subseteq \mathcal{B}_i(w, t)$, and (ii) the set of belief-accessible worlds is a subset of the intention-accessible worlds, $\mathcal{B}_i(w, t) \subseteq \mathcal{I}_i(w, t)$. Intentions are related to beliefs via the following properties respectively:

$Bel(i, \phi) \Rightarrow Intend(i, \phi)$

$Intend(i, \phi) \Rightarrow Bel(i, \phi)$

The former states that if an agent believes ϕ, then it intends it, while the latter states that if an agent intends ϕ, then it believes it. Following the same pattern, there are two axiom schemas that describe the possible relations between desires and intentions alike:

$Des(i, \phi) \Rightarrow Intend(i, \phi)$

$Intend(i, \phi) \Rightarrow Des(i, \phi)$

The first asserts that if an agent desires ϕ, then it intends it and the latter that if an agent intends ϕ, then it also desires it.

The various forms of consistency relations between the attitudes are considered next. Again starting with beliefs and desires, the intersection of the desire-accessible and belief-accessible worlds is not the empty set. Semantically, if an agent believes ϕ, then ϕ is true in all the belief-accessible worlds. Since the intersection of this set and the set of desire-accessible worlds is not the empty set, this means that in those worlds that lie in the intersection, ϕ is also true. Therefore, an agent does not have a desire $\neg\phi$. Conversely, if an agent desires ϕ, then ϕ is true in all its desire-accessible worlds. Since the intersection of this set and the set of belief-accessible worlds is not the empty set, this means that in those worlds that lie in the intersection, ϕ is also true. Therefore, an agent does not have a belief $\neg\phi$. The above take the form of the following schemas (Proposition 3):

$Bel(i, \phi) \Rightarrow \neg Des(i, \neg\phi)$

$Des(i, \phi) \Rightarrow \neg Bel(i, \neg\phi)$

The first property expresses that an agent's beliefs are consistent with its desires and conversely, the agent's desires are consistent with its beliefs. Similarly, if the intersection of the belief- and intention-accessible worlds is not the empty set then the following properties relate intentions with beliefs:

$Bel(i, \phi) \Rightarrow \neg Intend(i, \neg\phi)$

$Intend(i, \phi) \Rightarrow \neg Bel(i, \neg\phi)$

The first schema says that an agent does not believe a proposition the negation of which is intended, and conversely an agent does not intend a proposition the negation of which is believed. In the same way, if the intersection of the desire- and intention-accessible worlds is not the empty set, then the following schemas express that intentions are consistent with desires and conversely desires are consistent with intentions:

$Des(i, \phi) \Rightarrow \neg Intend(i, \neg\phi)$

$Intend(i, \phi) \Rightarrow \neg Des(i, \neg\phi)$

All the above presented axiom schemas apply to general formulas of the language. For instance, if $Intend(i, A(G(\phi)))$ is the case, then according to the last schema it is also the case that $\neg Des(i, \neg A(G(\phi)))$.

4.3 Structural Relations

Given that the possible worlds are not flat structures but have a branching time nature, this provides us with a way of refining the various axiom schemas that were introduced in the previous subsection. By imposing structural relations between worlds the application of the axioms can be restricted to subsets of the wffs of the language.

The basic structural relationship is that of a world w' being a subworld of another world w denoted $w' \subseteq w$. A world w' is a subworld of w, if w' is a subtree of w $((w', t_0, t_1, ...) \subseteq (w, t_0, t_1, ...))$, but they are otherwise identical to each other. The first such relation is called the structural subworld relation:

$X_i(w,t) \subseteq_{sub} Y_i(w,t)$:
$\forall\ w, w', t$ if $X_i(w,t,w')$ then $\exists\ w''$ s.t. $Y_i(w,t,w'')$ and $w' \subseteq w''$

If Xm and Ym are two modalities underpinned by the accessibility relations X_i and Y_i respectively, the above constraint corresponds to the axiom schema $Ym(i, A(\phi)) \Rightarrow Xm(i, A(\phi))$.

Lemma 1. *If $M(v,w,t) \models A\phi$ and $w' \subseteq w$, then $M(v,w',t) \models A\phi$.*

Proof. Assume $M(v,w,t_0) \models A\phi$. From the semantics of A we have that $M(v,w,t_0,t_1,...) \models \phi$ for all paths $(w,t_0,t_1,...)$. Since $w' \subseteq w$ we have that $(w',t_0,t_1,...) \subseteq (w,t_0,t_1,...)$. It now follows that $M(v,w',t_0,t_1,...) \models \phi$ and thus $M(v,w',t_0) \models A\phi$. □

Proposition 4. *Assume that Xm and Ym are two modalities defined by the two accessibility relations X_i and Y_i respectively. Then if $X_i(w,t) \subseteq_{sub} Y_i(w,t)$ we have that $Ym(i, A(\phi)) \Rightarrow Xm(i,\ A(\phi))$ is valid.*

Proof. If $X_i(w,t) \subseteq_{sub} Y_i(w,t)$ then $\forall\ w, w', t$ if $X_i(w,t,w')$ then $\exists\ w''$ such that $Y_i(w,t,w'')$ and $w' \subseteq w''$. Suppose $M(v,w,t) \models Ym(i, A(\phi))$ and $M(v,w,t) \models \neg Xm(i, A(\phi))$. Then there must be some w' $X_i(w,t,w')$ such that $M(v,w',t) \models \neg A(\phi)$. Since $X_i(w,t) \subseteq_{sub} Y_i(w,t)$ then there must be some w'' $Y_i(w,t,w'')$ such that $w' \subseteq w''$. Since $M(v,w,t) \models Ym(i, A(\phi))$ from the semantics of Ym it follows that $M(v,w'',t) \models A(\phi)$ and thus by Lemma 1 $M(v,w',t) \models A(\phi)$. However, this contradicts the original assumption, and hence the assumption must be false. □

In this way the application of the property is restricted to I-formulas of the language and as a result the axiom expresses attitudes towards inevitable states of affairs, or inevitable futures. This gives us a more fine-grained analysis of the relations between attitudes. For instance consider the axiom connecting intentions and desires $Intend(i, A(\phi)) \Rightarrow Des(i, A(\phi))$. This now says that if an agent intends that ϕ is inevitably true, then it desires it is inevitably true.

The second type of relation is called structural superworld relation:
$X_i(w,t) \subseteq_{sup} Y_i(w,t)$:
$\forall\ w, w', t$ if $X_i(w,t,w')$ then $\exists\ w''$ s.t. $Y_i(w,t,w'')$ and $w'' \subseteq w'$

The structural superset relation restricts the application of the axiom to O-formulas: $Ym(i, E(\phi)) \Rightarrow Xm(i, E(\phi))$. Thus, attitudes are expressed towards optional states of affairs, or options. For instance, consider the axiom $Intend(i, E(\phi)) \Rightarrow Bel(i, E(\phi))$. This says that if an agent intends that ϕ is optionally true, then it believes it is optionally true.

Lemma 2. *If $M(v,w',t) \models E\phi$ and $w' \subseteq w$, then $M(v,w,t) \models E\phi$.*

Proof. Assume $M(v,w',t_0) \models E\phi$. From the semantics of E we have that $M(v,w',t_0,t_1,...) \models \phi$ for some path $(w'$,t_0,t_1,...). Since $w' \subseteq w$ we have that $(w',t_0,t_1,...) \subseteq (w,t_0,t_1,...)$. It now follows that $M(v,w,t_0,t_1,...) \models \phi$ and thus $M(v,w,t_0) \models E\phi$. □

Proposition 5. *Assume that Xm and Ym are two modalities defined by the two accessibility relations X_i and Y_i respectively. Then if $X_i(w,t) \subseteq_{sup} Y_i(w,t)$ we have that $Ym(i, E(\phi)) \Rightarrow Xm(i,\ E(\phi))$ is valid.*

Proof. If $X_i(w,t) \subseteq_{sup} Y_i(w,t)$ then $\forall\ w, w', t$ if $X_i(w,t,w')$ then $\exists\ w''$ such that $Y_i(w,t,w'')$ and $w'' \subseteq w'$. Suppose $M(v,w,t) \models Ym(i, E(\phi))$ and $M(v,w,t) \models \neg Xm(i, E(\phi))$. Then there must be some w' $X_i(w,t,w')$ such that $M(v,w',t) \models \neg E(\phi)$. Since $X_i(w,t) \subseteq_{sup} Y_i(w,t)$ then there must be some w'' $Y_i(w,t,w'')$ such that $w'' \subseteq w'$. From the semantics of Ym it follows that $M(v,w'',t) \models E(\phi)$ and thus by Lemma 2 $M(v,w',t) \models E(\phi)$ which contradicts the original assumption. Hence the assumption must be false. □

The third relation is called structural consistency superworld relation:
$X_i(w,t) \cap_{sup} Y_i(w,t) \neq \emptyset$:
$\forall w, t\ \exists w'\ Y_i(w,t,w')$ s.t. $\exists\ w''\ X_i(w,t,w'')$ and $w' \subseteq w''$
As a result the relation between the two modalities Xm and Ym is described by the axiom schema $Xm(i, A(\phi)) \Rightarrow \neg Ym(i,\ \neg A(\phi))$. Such an axiom schema expresses consistency properties towards inevitable state of affairs.

Proposition 6. *Assume that Xm and Ym are two modalities defined by the two accessibility relations X_i and Y_i respectively. Then if $X_i(w,t) \cap_{sup} Y_i(w,t) \neq \emptyset$ we have that $Xm(i, A(\phi)) \Rightarrow \neg Ym(i,\ \neg A(\phi))$ is valid.*

Proof. Assume that $M(v,w,t) \models Xm(i, A(\phi))$ and $M(v,w,t) \models Ym(i,\ \neg A(\phi))$. Since $X_i(w,t) \cap_{sup} Y_i(w,t) \neq \emptyset$, there exists a w' $Y_i(w,t,w')$ and a w'' $X_i(w,t,w'')$ such that $w' \subseteq w''$. From the semantics of Xm and Ym it follows that $M(v,w'',t) \models A(\phi)$ and $M(v,w',t) \models \neg A(\phi)$. However, since $w' \subseteq w''$ it follows from Lemma 1 that $M(v,w',t) \models A(\phi)$. However, this contradicts the original assumption, and hence the assumption is false. □

Finally, the last type of relation is called the structural consistency subworld relation:
$X_i(w,t) \cap_{sub} Y_i(w,t) \neq \emptyset$:
$\forall w, t\ \exists w'\ Y_i(w,t,w')$ s.t. $\exists\ w''\ X_i(w,t,w'')$ and $w'' \subseteq w'$
This now corresponds to the axiom schema $Xm(i, E(\phi)) \Rightarrow \neg Ym(i,\ \neg E(\phi))$. This in turn means that the application of the consistency axioms is restricted to O-formulas.

Proposition 7. *Assume that Xm and Ym are two modalities defined by the two accessibility relations X_i and Y_i respectively. Then if $X_i(w,t) \cap_{sub} Y_i(w,t) \neq \emptyset$ we have that $Xm(i, E(\phi)) \Rightarrow \neg Ym(i,\ \neg E(\phi))$ is valid.*

Proof. The proof follows in a similar way to that of Proposition 6. □

Table 2 summarises the generic structural relations that can be adopted and the corresponding axioms.

Table 2. Generic structural relations between modalities

Relation	Axiom Schema
$X_i(w,t) \subseteq_{sub} Y_i(w,t)$	$Ym(i, A(\phi)) \Rightarrow Xm(i, A(\phi))$
$X_i(w,t) \subseteq_{sup} Y_i(w,t)$	$Ym(i, E(\phi)) \Rightarrow Xm(i, E(\phi))$
$X_i(w,t) \cap_{sup} Y_i(w,t) \neq \emptyset$	$Xm(i, A(\phi)) \Rightarrow \neg Ym(i, \neg A(\phi))$
$X_i(w,t) \cap_{sub} Y_i(w,t) \neq \emptyset$	$Xm(i, E(\phi)) \Rightarrow \neg Ym(i, \neg E(\phi))$

4.4 Further Relations

Rao and Georgeff in [25] provide further semantic conditions between accessible worlds in order to capture additional interesting relationships between beliefs, desires and intentions, apart from those already captured by the strong realism constraints. In particular, they want to capture that:
a) if an agent intends ϕ then it is quite reasonable to assume it believes that it intends it,
b) if an agent desires ϕ then it believes that it desires it, and finally, and
c) if an agent intends ϕ then it is quite reasonable to assume it desires to intend ϕ.
The authors therefore provide the following axioms for capturing these properties along with the corresponding supporting semantic conditions.

Belief of Intentions
$Intend(i, \phi) \Rightarrow Bel(i, Intend(i, \phi))$
(a) $\forall w, w', w'', t \; \mathcal{B}_i(w, t, w') \wedge \mathcal{I}_i(w, t, w'') \Rightarrow \mathcal{B}_i(w', t, w'')$

Belief of Desires
$Des(i, \phi) \Rightarrow Bel(i, Des(i, \phi))$
(b) $\forall w, w', w'', t \; \mathcal{B}_i(w, t, w') \wedge \mathcal{D}_i(w, t, w'') \Rightarrow \mathcal{B}_i(w', t, w'')$

Desires about Intentions
$Intend(i, \phi) \Rightarrow Des(i, Intend(i, \phi))$
(c) $\forall w, w', w'', t \; \mathcal{D}_i(w, t, w') \wedge \mathcal{I}_i(w, t, w'') \Rightarrow \mathcal{D}_i(w', t, w'')$

However, a closer examination of the semantic conditions reveals that they do not correspond to the proposed axioms. In particular, consider the first semantic condition. According to the proposed relation between accessible worlds, if a world w' is belief-accessible from w and another world w'' is intention-accessible from w then w'' should be belief-accessible from w'. In other words, the accessibility relation for belief $\mathcal{B}_i$ is Euclidean over $\mathcal{B}_i$ and $\mathcal{I}_i$. This however, does not capture the axiom of beliefs about intentions. This semantic condition supports the following schema:
$\neg Bel(i, \phi) \Rightarrow Bel(i, \neg Intend(i, \phi))$
Similar remarks can be made for the other two semantic conditions and their relation to the axioms.

Different semantic conditions need to be adopted in order to incorporate the properties above in the BDI logics:

Lemma 3. *The axiom schema:*
$Intend(i, \phi) \Rightarrow Bel(i, Intend(i, \phi))$

is sound in all models that satisfy the semantic condition:
BI. $\forall w, w', w'', t\ \mathcal{B}_i(w,t,w') \wedge \mathcal{I}_i(w',t,w'') \Rightarrow \mathcal{I}_i(w,t,w'')$

Proof. Assume $Intend(i,\phi)$ at w at time point t. Then for all w' such that $\mathcal{I}_i(w,t,w')$, $M(v,w',t) \models \phi$ (i). Let $\mathcal{B}_i(w,t,w')$ and $\mathcal{I}_i(w',t,w'')$, then by the semantic condition of BI it is the case that $\mathcal{I}_i(w,t,w'')$, and so by (i) we obtain $M(v,w'',t) \models \phi$ as required. □

Lemma 4. *The axiom schema:*
$Des(i,\phi) \Rightarrow Bel(i, Des(i,\phi))$
is sound in all models that satisfy the semantic condition:
BD. $\forall w, w', w'', t\ \mathcal{B}_i(w,t,w') \wedge \mathcal{D}_i(w',t,w'') \Rightarrow \mathcal{D}_i(w,t,w'')$

Proof. Assume $Des(i,\phi)$ at w at time point t. Then for all w' such that $\mathcal{D}_i(w,t,w')$, $M(v,w',t) \models \phi$ (i). Let $\mathcal{B}_i(w,t,w')$ and $\mathcal{D}_i(w',t,w'')$, then by the semantic condition of BD $\mathcal{D}_i(w,t,w'')$ is obtained, and so by (i) it is the case that $M(v,w'',t) \models \phi$ as required. □

Lemma 5. *The axiom schema:*
$Intend(i,\phi) \Rightarrow Des(i, Intend(i,\phi))$
is sound in all models that satisfy the semantic condition:
DI. $\forall w, w', w'', t\ \mathcal{D}_i(w,t,w') \wedge \mathcal{I}_i(w',t,w'') \Rightarrow \mathcal{I}_i(w,t,w'')$

Proof. Assume $Intend(i,\phi)$ at w at time point t. Then for all w' such that $\mathcal{I}_i(w,t,w')$, $M(v,w',t) \models \phi$ (i). Let $\mathcal{D}_i(w,t,w')$ and $\mathcal{I}_i(w',t,w'')$, then by the semantic condition of DI we obtain $\mathcal{I}_i(w,t,w'')$ and so by (i) it is the case that $M(v,w'',t) \models \phi$ as required. □

5 Notions of Realism

By combining binary relations between the modalities of belief, desires and intentions, we can construct notions of realism. A notion of realism describes the dynamics between the three attitudes and as a consequence different types of realism may characterise different types of agents. Three such notions of realism have been considered in the literature, namely *strong realism*, *realism* and *weak realism* [5, 25–27].

5.1 Strong Realism

In *strong realism*, the set of belief-accessible worlds is a subset of the desire-accessible worlds, which in turn is a subset of the intention-accessible worlds, Figure 1(i). According to *strong realism*, if an agent intends a state of affairs ϕ, then it desires it, and moreover believes it:

$Intend(i,\phi) \Rightarrow Des(i,\phi) \Rightarrow Bel(i,\phi)$

Table 3 contains the set relations along with the corresponding axiom schemas for *strong realism*. The respective system is called S-BDI.

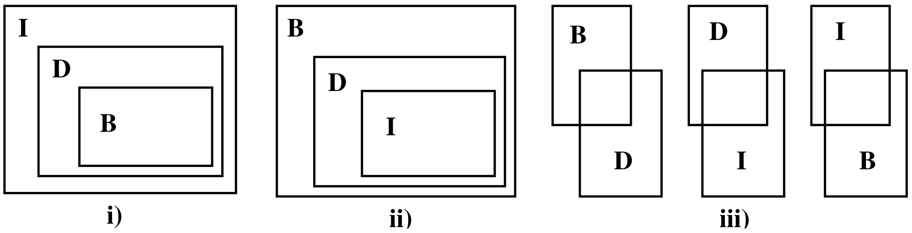

Fig. 1. Notions of realism: i) *strong realism*, ii) *realism*, iii) *weak realism*.

Proposition 8. *The following are theorems in S-BDI:*
1) $Intend(i, \phi) \Rightarrow Bel(i, Bel(i, \phi))$
2) $Des(i, \phi) \Rightarrow Bel(i, Bel(i, \phi))$

Proof. 1. Suppose $Intend(i, \phi)$. Then, according to the strong realism axiomatisation $Intend(i, \phi) \Rightarrow Bel(i, \phi)$, and by modus ponens $Bel(i, \phi)$. From the S4 axiom for belief $Bel(i, \phi) \Rightarrow Bel(i, Bel(i, \phi))$, is obtained and by applying modus ponens $Bel(i, Bel(i, \phi))$.
2. Suppose $Des(i, \phi)$. Then, according to the strong realism axiomatisation $Des(i, \phi) \Rightarrow Bel(i, \phi)$, and by modus ponens $Bel(i, \phi)$. From the S4 axiom for belief it is the case that $Bel(i, \phi) \Rightarrow Bel(i, Bel(i, \phi))$, and by applying modus ponens $Bel(i, Bel(i, \phi))$. □

The basic axiom schemas for *strong realism* can be further refined by considering structural relations between worlds. Thus if the application of the axioms is restricted to I-formulas and O-formulas we have two additional systems S-BDI$_A$ and S-BDI$_E$, the subscript indicating that the axioms apply to inevitable (A) and optional (E) formulas respectively. The agent described by *strong realism* is a "cautious" agent [27]. An agent intends states of affairs that are part of its desires and which are also believed. This characterisation seems more intuitive when the *strong realism* properties are considered in the context of optional formulas. Hence, an agent intends (optional) states of affairs that are part of its desires as long as it believes them to be options. Thus, an agent that intends that ϕ is optionally true, believes it to be optionally true.

Moreover, additional restrictions between worlds can be added in order to capture the properties presented in Section 4.4. Thus:

Proposition 9. *The following are theorems in the S-BDI system if the three semantic conditions BI, BD and DI are satisfied:*
1) $Intend(i, \phi) \Rightarrow Bel(i, Des(i, \phi))$
2) $Intend(i, \phi) \Rightarrow Bel(i, Bel(i, \phi))$
3) $Des(i, \phi) \Rightarrow Bel(i, Bel(i, \phi))$
4) $Intend(i, \phi) \Rightarrow Des(i, Des(i, \phi))$
5) $Intend(i, \phi) \Rightarrow Des(i, Bel(i, \phi))$

Proof. 1) Assume $Intend(i,\phi)$. By the belief of intentions axiom $Intend(i,\phi) \Rightarrow Bel(i, Intend(i,\phi))$ is obtained, and by modus ponens $Bel(i, Intend(i,\phi))$ (*). From the strong realism axioms $Intend(i,\phi) \Rightarrow Des(i,\phi)$ and by necessitation for belief $Bel(i, Intend(i,\phi) \Rightarrow Des(i,\phi))$. Now by distribution for belief $Bel(i, Intend(i,\phi)) \Rightarrow Bel(i, Des(i,\phi))$ and from (*) and modus ponens $Bel(i, Des(i,\phi))$ is obtained.

2) Assume $Intend(i,\phi)$. By the belief of intentions axiom $Intend(i,\phi) \Rightarrow Bel(i, Intend(i,\phi))$, and by modus ponens $Bel(i, Intend(i,\phi))$ (*). By strong realism it is the case that $Intend(i,\phi) \Rightarrow Bel(i,\phi)$, and by necessitation for belief $Bel(i, Intend(i,\phi) \Rightarrow Bel(i,\phi))$. By distribution for belief $Bel(i, Intend(i,\phi)) \Rightarrow Bel(i, Bel(i,\phi))$ is obtained, and from (*) and modus ponens $Bel(i, Bel(i,\phi))$.

3) Assume $Des(i,\phi)$. Then from the belief of desires axiom $Des(i,\phi) \Rightarrow Bel(i, Des(i,\phi))$, and by modus ponens $Bel(i, Des(i,\phi))$ (*). By strong realism we obtain $Des(i,\phi) \Rightarrow Bel(i,\phi)$ and by necessitation for belief $Bel(i, Des(i,\phi) \Rightarrow Bel(i,\phi))$. Now by distribution for belief $Bel(i, Des(i,\phi)) \Rightarrow Bel(i, Bel(i,\phi))$, and from (*) and modus ponens $Bel(i, Bel(i,\phi))$ is obtained.

4) Assume $Intend(i,\phi)$. By the desires of intentions axiom $Intend(i,\phi) \Rightarrow Des(i, Intend(i,\phi))$, and by modus ponens $Des(i, Intend(i,\phi))$ (*). By strong realism we get $Intend(i,\phi) \Rightarrow Des(i,\phi)$ and by necessitation for desires we obtain $Des(i, Intend(i,\phi) \Rightarrow Des(i,\phi))$. Now by distribution for desires we have $Des(i, Intend(i,\phi)) \Rightarrow Des(i, Des(i,\phi))$, and from (*) and modus ponens $Des(i, Des(i,\phi))$ as required.

5) Assume $Intend(i,\phi)$. Then, from the desires of intentions axiom we obtain $Intend(i,\phi) \Rightarrow Des(i, Intend(i,\phi))$, and by modus ponens $Des(i, Intend(i,\phi))$ (*). By strong realism it is the case that $Intend(i,\phi) \Rightarrow Bel(i,\phi)$ and by necessitation for desires $Des(i, Intend(i,\phi) \Rightarrow Bel(i,\phi))$. By distribution of desires we obtain $Des(i, Intend(i,\phi)) \Rightarrow Des(i, Bel(i,\phi))$, and from (*) and modus ponens $Des(i, Bel(i,\phi))$. □

The strong realism constraints do not seem to be appropriate for modal systems in which the information state of the agent is represented in terms of knowledge instead of belief. This is hardly surprising when one considers the knowledge axiomatisation with the T-axiom in conjunction with the strong realism axioms. In such a system the following would be theorems:
$Intend(i,\phi) \Rightarrow \phi$
$Des(i,\phi) \Rightarrow \phi$
which however, seem to be quite strong assertions for intentions and desires.

5.2 Realism

The second notion of realism is called simply *realism* and was first considered by Cohen and Levesque [5] in their theory of intentions. In terms of set relations, the set of intention-accessible worlds is a subset of the desire-accessible worlds, and the set of desire-accessible worlds is a subset of the belief-accessible worlds, Figure 1(ii). In *realism*, if an agent believes ϕ then it desires it, and if it desires it, then it intends it as well. Formally:

$Bel(i,\phi) \Rightarrow Des(i,\phi) \Rightarrow Intend(i,\phi)$

An agent based on *realism* is an "enthusiastic" agent [27]. This characterisation can be understood better in the context of optional formulas. According to the *realism* axioms, an agent intends all the options that it believes it has available. Table 3 details the basic axioms that can be imposed as part of the notion of *realism*. These can be further refined by restricting their application to I- and O-formulas.

Further interrelationships can be captured by imposing the semantic conditions of Section 4.4.

Proposition 10. *The following are theorems in the R-BDI system if the three semantic conditions BI, BD and DI are satisfied:*
1) $Intend(i,\phi) \Rightarrow Bel(i,\neg Des(i,\neg\phi))$
2) $Intend(i,\phi) \Rightarrow Bel(i,\neg Bel(i,\neg\phi))$
3) $Des(i,\phi) \Rightarrow Bel(i, Intend(i,\phi))$
4) $Des(i,\phi) \Rightarrow Bel(i,\neg Bel(i,\neg\phi))$
5) $Intend(i,\phi) \Rightarrow Des(i,\neg Des(i,\neg\phi))$
6) $Intend(i,\phi) \Rightarrow Des(i,\neg Bel(i,\neg\phi))$

Proof. 1) Assume $Intend(i,\phi)$. Then, from the belief of intentions axiom we have $Intend(i,\phi) \Rightarrow Bel(i, Intend(i,\phi))$, and by modus ponens $Bel(i, Intend(i,\phi))$ (*). By realism we have $Intend(i,\phi) \Rightarrow \neg Des(i,\neg\phi)$, and by necessitation for belief $Bel(i, Intend(i,\phi) \Rightarrow \neg Des(i,\neg\phi))$. By distribution of belief we obtain $Bel(i, Intend(i,\phi)) \Rightarrow Bel(i,\neg Des(i,\neg\phi))$, and from (*) and modus ponens $Bel(i,\neg Des(i,\neg\phi))$ as required.

2) Assume $Intend(i,\phi)$. By the belief of intentions axiom $Intend(i,\phi) \Rightarrow Bel(i, Intend(i,\phi))$, and by modus ponens $Bel(i, Intend(i,\phi))$ (*). By realism we obtain $Intend(i,\phi) \Rightarrow \neg Bel(i,\neg\phi)$, and by necessitation for belief we have $Bel(i, Intend(i,\phi) \Rightarrow \neg Bel(i,\neg\phi))$. By the B-K axiom $Bel(i, Intend(i,\phi)) \Rightarrow Bel(i,\neg Bel(i,\neg\phi))$. From (*) and modus ponens $Bel(i,\neg Bel(i,\neg\phi))$ is obtained.

3) Assume $Des(i,\phi)$, then by belief of desires $Des(i,\phi) \Rightarrow Bel(i, Des(i,\phi))$, and by modus ponens $Bel(i, Des(i,\phi))$ (*). From realism it is known that $Des(i,\phi) \Rightarrow Intend(i,\phi)$ and by necessitation for belief $Bel(i, Des(i,\phi) \Rightarrow Intend(i,\phi))$. By the distribution of belief $Bel(i, Des(i,\phi)) \Rightarrow Bel(i, Intend(i,\phi))$, and by (*) and modus ponens $Bel(i, Intend(i,\phi))$.

4) Assume $Des(i,\phi)$. By the belief of desires axiom we get $Des(i,\phi) \Rightarrow Bel(i, Des(i,\phi))$, and by modus ponens $Bel(i, Des(i,\phi))$ (*). By realism it is known that $Des(i,\phi) \Rightarrow Intend(i,\phi)$ and by necessitation for belief we obtain $Bel(i, Des(i,\phi) \Rightarrow Intend(i,\phi))$. By distribution of belief $Bel(i, Des(i,\phi)) \Rightarrow Bel(i, Intend(i,\phi))$, and from (*) and modus ponens $Bel(i, Intend(i,\phi))$ is obtained.

5) Assume $Intend(i,\phi)$. By the desires of intentions axiom $Intend(i,\phi) \Rightarrow Des(i, Intend(i,\phi))$, and by modus ponens $Des(i, Intend(i,\phi))$ (*). By realism we obtain $Intend(i,\phi) \Rightarrow \neg Des(i,\neg\phi)$ and by necessitation of desires we obtain $Des(i, Intend(i,\phi) \Rightarrow \neg Des(i,\neg\phi))$. By distribution of desires we obtain $Des(i, Intend(i,\phi)) \Rightarrow Des(i,\neg Des(i,\neg\phi))$, and from (*) and modus ponens $Des(i,\neg Des(i,\neg\phi))$ is obtained.

6) Assume $Intend(i,\phi)$. By the desires of intentions axiom $Intend(i,\phi) \Rightarrow Des(i, Intend(i,\phi))$, and by modus ponens $Des(i, Intend(i,\phi))$ (*). By realism we obtain $Intend(i,\phi) \Rightarrow \neg Bel(i,\neg\phi)$, and by necessitation for desires we have $Des(i, Intend(i,\phi) \Rightarrow \neg Bel(i,\neg\phi))$. By the distribution of the desires axiom $Des(i, Intend(i,\phi)) \Rightarrow Des(i,\neg Bel(i,\neg\phi))$, and thus from (*) and modus ponens $Des(i,\neg Bel(i,\neg\phi))$. □

5.3 Weak Realism

The third notion of realism is called *weak realism*. Set theoretically, the intersection of the intention- and desire-, intention- and belief-, and belief- and desire-accessible worlds is not the empty set as is shown in Figure 1(iii). Hence, if an agent believes ϕ, then it does not desire its negation, if it desires ϕ, it does not intend its negation, and if it intends ϕ, then it does not believe its negation. The agent characterised by *weak realism* is a more "balanced" agent than the other two types [27]. Structural relations between the accessible worlds can also be imposed in a similar way as in the first two notions of realism. The relations that characterise *weak realism* and the corresponding axioms are provided in Table 3.

If the semantic conditions of Section 4.4 are imposed then we have the following:

Proposition 11. *The following are theorems in W-BDI if the semantic conditions BI, BD and DI are satisfied:*
1) $Intend(i,\phi) \Rightarrow Bel(i,\neg Des(i,\neg\phi))$
2) $Intend(i,\phi) \Rightarrow Bel(i,\neg Bel(i,\neg\phi))$
3) $Des(i,\phi) \Rightarrow Bel(i,\neg Intend(i,\neg\phi))$
4) $Des(i,\phi) \Rightarrow Bel(i,\neg Bel(i,\neg\phi))$
5) $Intend(i,\phi) \Rightarrow Des(i,\neg Des(i,\neg\phi))$
6) $Intend(i,\phi) \Rightarrow Des(i,\neg Bel(i,\neg\phi))$

Proof. 1) Assume $Intend(i,\phi)$. Then from the belief of intentions axiom we have $Intend(i,\phi) \Rightarrow Bel(i, Intend(i,\phi))$, and by modus ponens $Bel(i, Intend(i,\phi))$ (*). By weak realism it is known that $Intend(i,\phi) \Rightarrow \neg Des(i,\neg\phi)$ and by necessitation for belief $Bel(i, Intend(i,\phi) \Rightarrow \neg Des(i,\neg\phi))$. Now by distribution for belief $Bel(i, Intend(i,\phi)) \Rightarrow Bel(i,\neg Des(i,\neg\phi))$, and from (*) and modus ponens $Bel(i,\neg Des(i,\neg\phi))$.

2) Assume $Intend(i,\phi)$. By the belief of intentions axiom $Intend(i,\phi) \Rightarrow Bel(i, Intend(i,\phi))$, and by modus ponens $Bel(i, Intend(i,\phi))$ (*). By weak realism we have $Intend(i,\phi) \Rightarrow \neg Bel(i,\neg\phi)$, and by necessitation for belief we obtain $Bel(i, Intend(i,\phi) \Rightarrow \neg Bel(i,\neg\phi))$. By distribution for belief we obtain $Bel(i, Intend(i,\phi)) \Rightarrow Bel(i,\neg Bel(i,\neg\phi))$, and from (*) and modus ponens $Bel(i,\neg Bel(i,\neg\phi))$ is obtained.

3) Assume $Intend(i,\phi)$. By the desires of intentions axiom $Intend(i,\phi) \Rightarrow Des(i, Intend(i,\phi))$, and by modus ponens $Des(i, Intend(i,\phi))$ (*). By weak realism it is known that $Intend(i,\phi) \Rightarrow \neg Bel(i,\neg\phi)$, and by necessitation for

Table 3. Relations and axioms in the three basic systems of realism

	Relation	Axiom Schema
S-BDI	$B_i(w,t) \subseteq D_i(w,t) \subseteq I_i(w,t)$	$Intend(i,\phi) \Rightarrow Des(i,\phi) \Rightarrow Bel(i,\phi)$
R-BDI	$I_i(w,t) \subseteq D_i(w,t) \subseteq B_i(w,t)$	$Bel(i,\phi) \Rightarrow Des(i,\phi) \Rightarrow Intend(i,\phi)$
W-BDI	$B_i(w,t) \cap D_i(w,t) \neq \emptyset$	$Bel(i,\phi) \Rightarrow \neg Des(i,\neg\phi)$
	$B_i(w,t) \cap I_i(w,t) \neq \emptyset$	$Bel(i,\phi) \Rightarrow \neg Intend(i,\neg\phi)$
	$I_i(w,t) \cap D_i(w,t) \neq \emptyset$	$Intend(i,\phi) \Rightarrow \neg Des(i,\neg\phi)$

desires $Des(i, Intend(i,\phi) \Rightarrow \neg Bel(i,\neg\phi))$. By distribution for desires we have $Des(i, Intend(i,\phi)) \Rightarrow Des(i, \neg Bel(i,\neg\phi))$, and from (*) and modus ponens we obtain $Des(i, \neg Bel(i,\neg\phi))$.

4) Assume $Intend(i,\phi)$. Then from the desires of intentions axiom it is the case that $Intend(i,\phi) \Rightarrow Des(i, Intend(i,\phi))$, and by modus ponens we gat $Des(i, Intend(i,\phi))$ (*). By weak realism we have $Intend(i,\phi) \Rightarrow \neg Des(i,\neg\phi)$ and by necessitation for desires $Des(i, Intend(i,\phi) \Rightarrow \neg Des(i,\neg\phi))$. By distribution for desires we obtain $Des(i, Intend(i,\phi)) \Rightarrow Des(i, \neg Des(i,\neg\phi))$, and from (*) and modus ponens $Des(i, \neg Des(i,\neg\phi))$.

5) Assume $Intend(i,\phi)$. By the desires of intentions axiom $Intend(i,\phi) \Rightarrow Des(i, Intend(i,\phi))$, and by modus ponens $Des(i, Intend(i,\phi))$ (*). By the axioms of weak realism it is known that $Intend(i,\phi) \Rightarrow \neg Des(i,\neg\phi)$, and by necessitation for desires $Des(i, Intend(i,\phi) \Rightarrow \neg Des(i,\neg\phi))$. By distribution for desires we obtain $Des(i, Intend(i,\phi)) \Rightarrow Des(i, \neg Des(i,\neg\phi))$ and from (*) and modus ponens $Des(i, \neg Des(i,\neg\phi))$.

6) Assume $Intend(i,\phi)$. Then from the desires of intentions axiom it is the case that $Intend(i,\phi) \Rightarrow Des(i, Intend(i,\phi))$, and by modus ponens we have $Des(i, Intend(i,\phi))$ (*). By weak realism it is known that $Intend(i,\phi) \Rightarrow \neg Bel(i,\neg\phi)$, and by necessitation for desires $Des(i, Intend(i,\phi) \Rightarrow \neg Bel(i,\neg\phi))$. By distribution for desires $Des(i, Intend(i,\phi)) \Rightarrow Des(i, \neg Bel(i,\neg\phi))$, and now from (*) and modus ponens $Des(i, \neg Bel(i,\neg\phi))$ is obtained. □

6 Asymmetry Thesis and the Side-effect Problem

As we saw from the previous discussion, various forms of realism can be captured, each dictating different relations between the attitudes and consequently characterising a different type of agent. But how do we evaluate the properties of a BDI agent?

Bratman [3] and Rao and Georgeff [27] argued that certain principles should be taken into account if we are to accept that a BDI system captures the properties of a rational agent. These properties are known as the asymmetry thesis (AT) or the incompleteness and the inconsistency principles, and they hold pairwise between desires, beliefs, and intentions. They are listed in Table 4, albeit the naming scheme that we use is a bit different.

Bratman [3] argues that Intention-Belief Inconsistency should not be allowed. It is irrational for an agent to intend to bring about a state of affairs and believe

Table 4. Asymmetry thesis principles

Principle	Formula
A1 I-B Inconsistency	$\vdash Intend(i,\phi) \Rightarrow \neg Bel(i,\neg\phi)$
A2 I-B Incompleteness	$\not\vdash Intend(i,\phi) \Rightarrow Bel(i,\ \phi)$
A3 I-D Incompleteness	$\not\vdash Intend(i,\phi) \Rightarrow Des(i,\ \phi)$
A4 I-D Inconsistency	$\vdash Intend(i,\phi) \Rightarrow \neg Des(i,\neg\phi)$
A5 B-D Incompleteness	$\not\vdash Bel(i,\phi) \Rightarrow Des(i,\ \phi)$
A6 B-I Incompleteness	$\not\vdash Bel(i,\phi) \Rightarrow Intend(i,\ \phi)$
A7 D-B Inconsistency	$\vdash Des(i,\phi) \Rightarrow \neg Bel(i,\neg\phi)$
A8 D-I Incompleteness	$\not\vdash Des(i,\phi) \Rightarrow Intend(i,\phi)$
A9 D-B Incompleteness	$\not\vdash Des(i,\phi) \Rightarrow Bel(i,\ \phi)$

that it does not do it. This corresponds to the formula $(Intend(i,\phi) \wedge Bel(i,\neg\phi))$ in the BDI framework. If this was allowed, then Alice the robot intends to cross the road, while at the same time believes that it is not doing it. In order to avoid such behaviour, the formula $Intend(i,\phi) \Rightarrow \neg Bel(i,\neg\phi)$ should be valid.

On the other hand Intention-Belief Incompleteness is allowed: an agent should be allowed to have an intention to bring about a state of affairs, but not necessarily believe that it is going to do it. Alice in this case may intend to cross the road, but not believe that it is doing it. Thus Intend$(i,\phi) \wedge \neg Bel(i,\phi)$ should be satisfiable, or in other words $Intend(i,\phi) \Rightarrow Bel(i,\phi)$ should not be valid in a BDI system.

Belief-Intention Incompleteness is another principle that should be allowed. It describes that an agent may believe that it can bring about ϕ, while not necessarily intending to do it. So Alice believes it can cross the road, but does not intend to do it. Hence, $Bel(i,\phi) \Rightarrow Intend(i,\phi)$ is required not to be valid. Similar comments can be made for the rest of the principles of Table 4.

Theories that involve attitudes such as beliefs, desires and intentions are prone to the side-effect problem. Thus, another way to evaluate the behaviour of a BDI agent is to examine whether or not it is affected by the side-effect problem and to what extent. Imagine an agent that intends to pay a visit to the dentist. The agent believes that a visit to the dentist implies suffering pain. Is it the case that since the agent intends ϕ and believes that $\phi \Rightarrow \psi$, it also intends ψ? In other words does an agent intend all the side-effects of its intentions? Obviously, this is not reasonable. The general form of this and two other variations of the side-effect problem are given below:

C1. $Intend(i,\phi_1) \wedge Bel(i,\phi_1 \Rightarrow \phi_2) \wedge \neg Intend(i,\phi_2)$

C2. $Intend(i,\phi_1) \wedge Des(i,\phi_1 \Rightarrow \phi_2) \wedge \neg Intend(i,\phi_2)$

C3. $Des(i,\phi_1) \wedge Bel(i,\phi_1 \Rightarrow \phi_2) \wedge \neg Des(i,\phi_2)$

These are known as the consequential closure principles (CC). It is easy to see that in order to avoid the various forms of the side-effect problem, these formulas should be satisfiable in the BDI system under consideration. However, their satisfiability depends upon the additional realism constraints.

Table 5 presents the satisfaction of the Asymmetry Thesis and the Consequential Closure principles for the three most generalised BDI systems, namely

Table 5. AT and CC principles in S-BDI, R-BDI and W-BDI

System	A1	A2	A3	A4	A5	A6	A7	A8	A9	C1	C2	C3
S-BDI	T	F	F	T	T	T	T	T	F	T	T	T
R-BDI	T	T	T	T	F	F	T	F	T	F	F	F
W-BDI	T	T	T	T	T	T	T	T	T	T	T	T

S-BDI, R-BDI and W-BDI. More details can be found in [26, 27]. In brief, S-BDI which is based on *strong realism*, does not satisfy three of the asymmetry thesis principles, whereas R-BDI does not satisfy any of the consequential closure and three of the asymmetry thesis principles. The BDI system based on *weak realism* is the only one that satisfies all principles.

7 Heterogeneous BDI Agents

Nowadays agent-based systems are being used in a variety of applications ranging from e-commerce to space mission control. Undoubtedly, these diverse in nature domains have different characteristics. As a result, they impose different constraints on the required behaviour. For instance, it seems unreasonable to assume that a stock market agent should exhibit the same characteristics in behaviour as that of an agent controlling a space mission. But even within the same domain, the requirements in the exhibited type of behaviour may vary. For instance, consider software agents that participate in electronic auctions. Such agents may not be of the same type. User A may require a risk-neutral type of agent, whereas user B a risk-prone one. Consequently, the need for heterogeneous agents stems from the diversity of the domains, as well as within domains.

This diversity needs to be reflected on the conceptualisation, design and implementation of agents. Agent theories in general and the BDI paradigm in particular can be viewed as specification languages for agents. As such they can be used to describe, design and validate the properties of agents and multi-agent systems. In the BDI paradigm heterogeneous agents can be described by adopting different realism constraints. These essentially characterise different types of agents and as we saw in the literature a cautious (*strong realism*), an enthusiastic (*realism*) and a balanced type of agent (*weak realism*) have been described. However these notions of realism present only three uniform ways that the various types of relations can be combined. A natural question arises: are these the only meaningful types of BDI agents? The answer comes [27] themselves: there may be additional meaningful types of BDI agents apart from the three considered. Motivated by this remark, we explore additional notions of realism. In the following sections we provide two ways of categorising BDI agents: the first one by considering the relations between sets of accessible worlds, and the second one by considering the relation between intentions and beliefs.

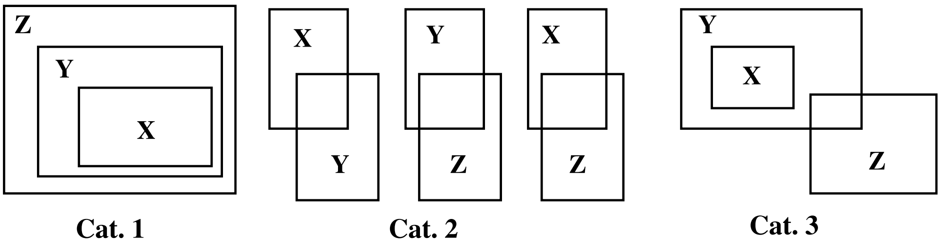

Fig. 2. Categories of realism.

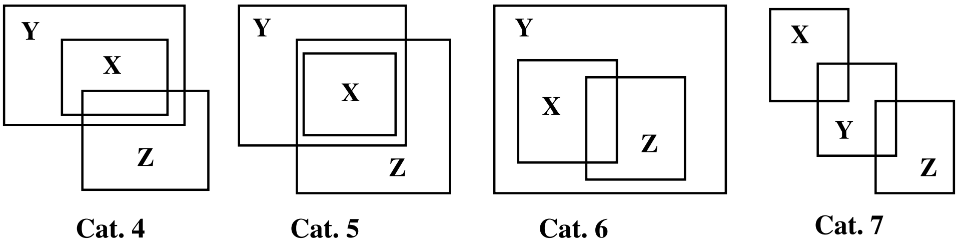

Fig. 3. Categories of realism (continued).

7.1 Categories of Realism

The main endeavour in this section is to categorise notions of realism for BDI agents according to the relations between the sets of accessible worlds. If Xm, Ym and Zm are three modalities such that $Xm \neq Ym \neq Zm$, and X, Y, Z are the respective sets of accessible worlds, then the interesting generic notions of realism based on the set relations are illustrated in Figures 2 and 3.

In the first category, all sets are related to each other via the subset relation; set X is a subset of Y, and Y is a subset of Z. Clearly *strong realism* and *realism* belong to this category. In category 2, all sets are related to each other via the consistency relation. *Weak realism* is the only notion comprising category 2. Category 3 includes those notions of realism in which set X is a subset of Y, while Z is connected to Y via the consistency relation. As is illustrated in Figure 2, X and Z are completely decoupled. In category 4, set X is a subset of Y and Z is related to X via the consistency relation. Category 5 consists of those notions of realism in which a set X is a subset of both sets Y and Z as depicted in Figure 3. In category 6, sets X and Z are related to each other via the consistency relation and they are both subsets of set Y. Finally, in category 7, set X is related to Y via the consistency relation, and Y is related to Z via the consistency relation as well. This category is different from category 2, since two sets are completely decoupled from one another.

Table 6 summarises the categories of realism, the relations that underpin them and the corresponding properties. This typology only involves set theo-

Table 6. Categories of realism: relations and axiom schemas

Category	Relation	Axiom Schema
Cat. 1	$X \subseteq Y \subseteq Z$	$Zm(i,\phi) \Rightarrow Ym(i,\phi) \Rightarrow Xm(i,\phi)$
Cat. 2	$X \cap Y \neq \emptyset$	$Xm(i,\phi) \Rightarrow \neg Ym(i,\neg\phi)$
	$Y \cap Z \neq \emptyset$	$Ym(i,\phi) \Rightarrow \neg Zm(i,\neg\phi)$
	$X \cap Z \neq \emptyset$	$Xm(i,\phi) \Rightarrow \neg Zm(i,\neg\phi)$
Cat. 3	$X \subseteq Y$	$Ym(i,\phi) \Rightarrow Xm(i,\phi)$
	$Y \cap Z \neq \emptyset$	$Ym(i,\phi) \Rightarrow \neg Zm(i,\neg\phi)$
Cat. 4	$X \subseteq Y$	$Ym(i,\phi) \Rightarrow Xm(i,\phi)$
	$X \cap Z \neq \emptyset$	$Xm(i,\phi) \Rightarrow \neg Zm(i,\neg\phi)$
Cat. 5	$X \subseteq Y$	$Ym(i,\phi) \Rightarrow Xm(i,\phi)$
	$X \subseteq Z$	$Zm(i,\phi) \Rightarrow Xm(i,\phi)$
Cat. 6	$X \subseteq Y$	$Ym(i,\phi) \Rightarrow Xm(i,\phi)$
	$Z \subseteq Y$	$Ym(i,\phi) \Rightarrow Zm(i,\phi)$
	$X \cap Z \neq \emptyset$	$Xm(i,\phi) \Rightarrow \neg Zm(i,\neg\phi)$
Cat. 7	$X \cap Y \neq \emptyset$	$Xm(i,\phi) \Rightarrow \neg Ym(i,\neg\phi)$
	$Y \cap Z \neq \emptyset$	$Ym(i,\phi) \Rightarrow \neg Zm(i,\neg\phi)$

retic relations. Notably, certain cases have been excluded. Such cases include two or more sets of accessible worlds being identical or two or more sets being completely unrelated; these were rendered to be of no interest.

As the reader can check, in total 28 distinct notions of realism, including *strong realism*, *realism* and *weak realism*, can be constructed. These are provided in Table 7. Needless to say, not all of the systems yield attractive properties for agents.

Structural relations between worlds can be adopted in addition to the set relations in order to refine the properties of each notion of realism. Thus, for each notion of realism three systems can be obtained: one generalised, one applying to inevitabilities (A) and one applying to options (E).

7.2 Bold and Circumspect BDI agents

So far we have considered categories of realism according to the relations between the sets of accessible worlds. In this section, we will attempt to categorise notions of realism according to a different criterion.

Notably, one of the decisive factors in the characterisation of a BDI agent as cautious or enthusiastic is the relation between intentions and beliefs. This seems quite reasonable since an agent's decisions on what actions to take are inevitably based upon its information about the world. Thus the cautious agent described by the notion of *strong realism* has the following property relating beliefs and intentions, $Intend(i,\phi) \Rightarrow Bel(i,\phi)$. This can be understood in general terms as stating that if an agent intends ϕ, then it believes it. This property sounds more intuitive when we consider its application to optional formulas:

$Intend(i, E(\phi)) \Rightarrow Bel(i, E(\phi))$

In other words, if an agent intends that ϕ is optionally true, then it believes it is optionally true. Thus an agent's intentions are grounded on its beliefs and

Table 7. Notions of realism per category

Cat.	Realism	Relation
Cat.1	R1-1	$\mathcal{I}_i(w,t) \subseteq \mathcal{D}_i(w,t) \subseteq \mathcal{B}_i(w,t)$
	R1-2	$\mathcal{D}_i(w,t) \subseteq \mathcal{I}_i(w,t) \subseteq \mathcal{B}_i(w,t)$
	R1-3	$\mathcal{I}_i(w,t) \subseteq \mathcal{B}_i(w,t) \subseteq \mathcal{D}_i(w,t)$
	R1-4	$\mathcal{B}_i(w,t) \subseteq \mathcal{I}_i(w,t) \subseteq \mathcal{D}_i(w,t)$
	R1-5	$\mathcal{D}_i(w,t) \subseteq \mathcal{B}_i(w,t) \subseteq \mathcal{I}_i(w,t)$
	R1-6	$\mathcal{B}_i(w,t) \subseteq \mathcal{D}_i(w,t) \subseteq \mathcal{I}_i(w,t)$
Cat.2	R2-1	$\mathcal{B}_i(w,t) \cap \mathcal{D}_i(w,t) \neq \emptyset$; $\mathcal{B}_i(w,t) \cap \mathcal{I}_i(w,t) \neq \emptyset$; $\mathcal{I}_i(w,t) \cap \mathcal{D}_i(w,t) \neq \emptyset$
Cat.3	R3-1	$\mathcal{D}_i(w,t) \subseteq \mathcal{B}_i(w,t)$; $\mathcal{B}_i(w,t) \cap \mathcal{I}_i(w,t) \neq \emptyset$
	R3-2	$\mathcal{I}_i(w,t) \subseteq \mathcal{B}_i(w,t)$; $\mathcal{B}_i(w,t) \cap \mathcal{D}_i(w,t) \neq \emptyset$
	R3-3	$\mathcal{B}_i(w,t) \subseteq \mathcal{D}_i(w,t)$; $\mathcal{D}_i(w,t) \cap \mathcal{I}_i(w,t) \neq \emptyset$
	R3-4	$\mathcal{I}_i(w,t) \subseteq \mathcal{D}_i(w,t)$; $\mathcal{D}_i(w,t) \cap \mathcal{B}_i(w,t) \neq \emptyset$
	R3-5	$\mathcal{B}_i(w,t) \subseteq \mathcal{I}_i(w,t)$; $\mathcal{I}_i(w,t) \cap \mathcal{D}_i(w,t) \neq \emptyset$
	R3-6	$\mathcal{D}_i(w,t) \subseteq \mathcal{I}_i(w,t)$; $\mathcal{I}_i(w,t) \cap \mathcal{B}_i(w,t) \neq \emptyset$
Cat.4	R4-1	$\mathcal{D}_i(w,t) \subseteq \mathcal{B}_i(w,t)$; $\mathcal{D}_i(w,t) \cap \mathcal{I}_i(w,t) \neq \emptyset$
	R4-2	$\mathcal{I}_i(w,t) \subseteq \mathcal{B}_i(w,t)$; $\mathcal{I}_i(w,t) \cap \mathcal{D}_i(w,t) \neq \emptyset$
	R4-3	$\mathcal{B}_i(w,t) \subseteq \mathcal{D}_i(w,t)$; $\mathcal{B}_i(w,t) \cap \mathcal{I}_i(w,t) \neq \emptyset$
	R4-4	$\mathcal{I}_i(w,t) \subseteq \mathcal{D}_i(w,t)$; $\mathcal{I}_i(w,t) \cap \mathcal{B}_i(w,t) \neq \emptyset$
	R4-5	$\mathcal{B}_i(w,t) \subseteq \mathcal{I}_i(w,t)$; $\mathcal{B}_i(w,t) \cap \mathcal{D}_i(w,t) \neq \emptyset$
	R4-6	$\mathcal{D}_i(w,t) \subseteq \mathcal{I}_i(w,t)$; $\mathcal{D}_i(w,t) \cap \mathcal{B}_i(w,t) \neq \emptyset$
Cat.5	R5-1	$\mathcal{D}_i(w,t) \subseteq \mathcal{B}_i(w,t)$; $\mathcal{D}_i(w,t) \subseteq \mathcal{I}_i(w,t)$
	R5-2	$\mathcal{I}_i(w,t) \subseteq \mathcal{B}_i(w,t)$; $\mathcal{I}_i(w,t) \subseteq \mathcal{D}_i(w,t)$
	R5-3	$\mathcal{B}_i(w,t) \subseteq \mathcal{D}_i(w,t)$; $\mathcal{B}_i(w,t) \subseteq \mathcal{I}_i(w,t)$
Cat.6	R6-1	$\mathcal{I}_i(w,t) \subseteq \mathcal{B}_i(w,t)$; $\mathcal{D}_i(w,t) \subseteq \mathcal{B}_i(w,t)$; $\mathcal{I}_i(w,t) \cap \mathcal{D}_i(w,t) \neq \emptyset$
	R6-2	$\mathcal{I}_i(w,t) \subseteq \mathcal{D}_i(w,t)$; $\mathcal{B}_i(w,t) \subseteq \mathcal{D}_i(w,t)$; $\mathcal{I}_i(w,t) \cap \mathcal{B}_i(w,t) \neq \emptyset$
	R6-3	$\mathcal{D}_i(w,t) \subseteq \mathcal{I}_i(w,t)$; $\mathcal{B}_i(w,t) \subseteq \mathcal{I}_i(w,t)$; $\mathcal{D}_i(w,t) \cap \mathcal{B}_i(w,t) \neq \emptyset$
Cat.7	R7-1	$\mathcal{B}_i(w,t) \cap \mathcal{D}_i(w,t) \neq \emptyset$; $\mathcal{D}_i(w,t) \cap \mathcal{I}_i(w,t) \neq \emptyset$
	R7-2	$\mathcal{B}_i(w,t) \cap \mathcal{I}_i(w,t) \neq \emptyset$; $\mathcal{I}_i(w,t) \cap \mathcal{D}_i(w,t) \neq \emptyset$
	R7-3	$\mathcal{I}_i(w,t) \cap \mathcal{B}_i(w,t) \neq \emptyset$; $\mathcal{B}_i(w,t) \cap \mathcal{D}_i(w,t) \neq \emptyset$

therefore it only adopts intentions that it believes to be options. On the other hand, the enthusiastic agent that is described by the notion of *realism* has the following property relating beliefs and intentions, $Bel(i, \phi) \Rightarrow Intend(i, \phi)$ or $Intend(i, \phi) \Rightarrow \neg Bel(i, \neg\phi)$, or an agent that intends ϕ at least does not believe its negation. This is also the axiom relating intentions and beliefs in *weak realism*. Restricting the axiom to optional states of affairs:

$Intend(i, E(\phi)) \Rightarrow \neg Bel(i, \neg E(\phi))$

In other words, an agent's optional intentions are consistent with its beliefs.

Although further axioms provide additional characteristics to the three notions of realism, most importantly an agent's intentions are those that guide its actions, and consequently the way they interact with beliefs play a major role in the characterisation of an agent as enthusiastic or cautious. Here we adopt the term "circumspect" for an agent whose beliefs are related to intentions via the schema: $Intend(i, \phi) \Rightarrow Bel(i, \phi)$, and the term "bold" for an agent whose beliefs and intentions are related via the schema $Intend(i, \phi) \Rightarrow \neg Bel(i, \neg\phi)$. Thus, each agent can be characterised as bold or circumspect according to this criterion.

7.3 Evaluating Varieties of Realism

Hence, the 28 general BDI systems that can be obtained from the respective notions of realism can be divided according to the relationship between belief- and intention-accessible worlds. However, this now begs two questions: i) how do we judge which of these notions of realism are better or more interesting than others and based on what criteria? and ii) how do they relate to the three original notions of realism?

Recall from previous sections that we can evaluate a BDI system by checking whether or not it satisfies the asymmetry thesis and the consequential closure principles. This will provide us with an indication of whether or not a particular BDI system captures properties of a rational agent or not and to what extent. Consequently, "interesting" and "better" are those notions of realism that characterise rational BDI agents. The original system based on *strong realism* does not satisfy three of the asymmetry thesis principles, while the one based on *realism* does not satisfy any of the consequential closure and three of the asymmetry thesis principles. In fact, the only notion of realism that satisfies all of them is that of *weak realism* which according to our typology comprises category 2. *Weak realism* characterises a "balanced" type of agent according to [27]. We can imagine a triangle in which *weak realism* is placed at the top indicating that it is the only system that satisfies all principles, while *strong realism* and *realism* are placed at the other two corners. Our aim is to move further up towards *weak realism* and improve on the other two notions, attempting to satisfy more of the desiderata for rational BDI agents while preserving the main characterisation of agents being circumspect or bold. Following this tactic we will review some of these BDI systems that we consider to be interesting and come closer to the desiderata for rational BDI agents as laid down in [3, 27].

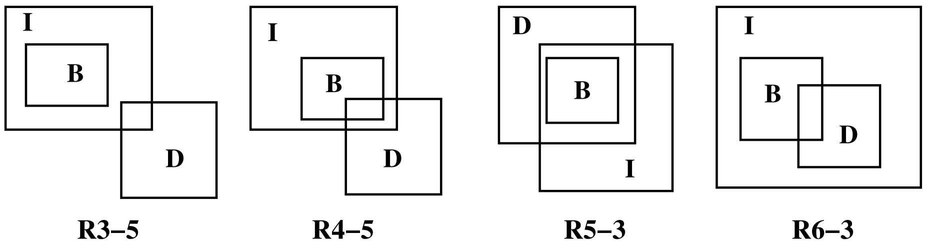

Fig. 4. Notions of realism for circumspect agents.

8 Circumspect Agents

The basic constraint that characterises circumspect agents in terms of set relations is that the set of belief-accessible worlds is a subset of the intention-accessible worlds. Accordingly, the main feature of such agents is that if they intend a state of affairs, then they believe it. The following sections will discuss four notions of realism characterising circumspect BDI agents based on this relation. These are illustrated in Figure 4. The naming scheme used is as follows: R*n*-*m* realism belongs to the general category n (see section 7.1) and it is the mth type of realism in this category (Table 7).

8.1 R3-5 Realism

The first notion of realism to be examined belongs to category 3 and is R3-5. The main relation that characterises circumspect agents is present, that is the set of belief-accessible worlds is a subset of the intention-accessible worlds. However, in this case the agent's desires are completely decoupled from its beliefs, and they are only consistent with its intentions. The set relations are depicted in Figure 4. Formally we have the following schemas relating the modalities:

$Intend(i,\phi) \Rightarrow Bel(i,\phi)$
$Intend(i,\phi) \Rightarrow \neg Des(i,\neg\phi)$

Lemma 6. *The properties for R3-5 realism are sound in all models that satisfy the semantic conditions:*
(i) $\mathcal{B}_i(w,t) \subseteq \mathcal{I}_i(w,t) : \forall w,w',t\ \mathcal{B}_i(w,t,w') \Rightarrow \mathcal{I}_i(w,t,w')$
(ii) $\mathcal{I}_i(w,t) \cap \mathcal{D}_i(w,t) \neq \emptyset : \forall w,t,\exists w'\ \mathcal{I}_i(w,t,w') \wedge \mathcal{D}_i(w,t,w')$

Proof. (i) Assume that $M(v,w,t) \models Intend(i,\phi)$ for an arbitrary $M(v,w,t)$. According to semantics we have $M(v,w',t) \models \phi$ for all w' such that $\mathcal{I}_i(w,t,w')$. Since $\mathcal{B}_i(w,t) \subseteq \mathcal{I}_i(w,t)$ we have $M(v,w'',t) \models \phi$ for all w'' such that $\mathcal{B}_i(w,t,w'')$. It now follows that $M(v,w,t) \models Bel(i,\phi)$.

(ii) Assume $M(v,w,t) \models Intend(i,\phi)$. Then for all w' such that $\mathcal{I}_i(w,t,w')$ we have $M(v,w',t) \models \phi$ and thus $M(v,w',t) \not\models \neg\phi$. Since $\mathcal{I}_i(w,t) \cap \mathcal{D}_i(w,t)$, there is at least one world w' such that $\mathcal{D}_i(w,t,w')$. It now follows that $M(v,w,t) \not\models Des(i,\neg\phi)$ and hence $M(v,w,t) \models \neg Des(i,\neg\phi)$ as required. □

The system consisting of the basic BDI system and the axioms for R3-5 realism will be called R3-5-BDI. Further constraints can be imposed in order to capture the properties described in Section 4.4. In particular, if the BI and DI conditions are imposed we have the following properties:

$Intend(i,\phi) \Rightarrow Bel(i, Intend(i,\phi))$ (beliefs about intentions)

$Intend(i,\phi) \Rightarrow Des(i, Intend(i,\phi))$ (desires about intentions)

The semantic condition BD (beliefs about desires) is not imposed since in this particular notion of realism the sets of the belief- and desire-accessible worlds are not related with each other.

Proposition 12. *The following properties are valid in R3-5-BDI if the BI and DI semantic conditions are satisfied:*
1) $Intend(i,\phi) \Rightarrow Bel(i, \neg Des(i,\neg\phi))$
2) $Intend(i,\phi) \Rightarrow Bel(i, Bel(i,\phi))$
3) $Intend(i,\phi) \Rightarrow Des(i, \neg Des(i,\neg\phi))$
4) $Intend(i,\phi) \Rightarrow Des(i, Bel(i,\phi))$

Proof. 1) Assume $Intend(i,\phi)$. By the belief of intentions property we have $Intend(i,\phi) \Rightarrow Bel(i, Intend(i,\phi))$ and by modus ponens $Bel(i, Intend(i,\phi))$ (*). From the R3-5 realism properties we have $Intend(i,\phi) \Rightarrow \neg Des(i,\neg\phi)$ and by necessitation for belief $Bel(i, Intend(i,\phi) \Rightarrow \neg Des(i,\neg\phi))$. By distribution of belief $Bel(i, Intend(i,\phi)) \Rightarrow Bel(i, \neg Des(i,\neg\phi))$ and from (*) and modus ponens $Bel(i, \neg Des(i,\neg\phi))$ is obtained.

2) Assume $Intend(i,\phi)$. By the belief of intentions property $Intend(i,\phi) \Rightarrow Bel(i, Intend(i,\phi))$, and by modus ponens $Bel(i, Intend(i,\phi))$ (*). From the R3-5 realism properties we have $Intend(i,\phi) \Rightarrow Bel(i,\phi)$ and by necessitation for belief $Bel(i, Intend(i,\phi) \Rightarrow Bel(i,\phi))$. By distribution of belief we have $Bel(i, Intend(i,\phi)) \Rightarrow Bel(i, Bel(i,\phi))$ and from (*) and modus ponens $Bel(i, Bel(i,\phi))$ is obtained.

3) Assume $Intend(i,\phi)$. By the desires of intentions property $Intend(i,\phi) \Rightarrow Des(i, Intend(i,\phi))$ and by modus ponens $Des(i, Intend(i,\phi))$ (*). From the R3-5 realism properties we have $Intend(i,\phi) \Rightarrow \neg Des(i,\neg\phi)$ and by necessitation for desires $Des(i, Intend(i,\phi) \Rightarrow \neg Des(i,\neg\phi))$. By distribution of desires $Des(i, Intend(i,\phi)) \Rightarrow Des(i, \neg Des(i,\neg\phi))$ and from (*) and modus ponens $Des(i, \neg Des(i,\neg\phi))$ is obtained.

4) Assume $Intend(i,\phi)$. By the desires of intentions property $Intend(i,\phi) \Rightarrow Des(i, Intend(i,\phi))$ and by modus ponens $Des(i, Intend(i,\phi))$ (*). From the R3-5 realism properties we have $Intend(i,\phi) \Rightarrow Bel(i,\phi)$ and by necessitation for desires $Des(i, Intend(i,\phi) \Rightarrow Bel(i,\phi))$. By distribution of desires we obtain $Des(i, Intend(i,\phi)) \Rightarrow Des(i, Bel(i,\phi))$ and from (*) and modus ponens $Des(i, Bel(i,\phi))$ is obtained. □

The $Intend(i,\phi) \Rightarrow Bel(i, Bel(i,\phi))$ property is also valid in R3-5-BDI even if the BI condition is not imposed. This follows simply by the B-S4 axiom for belief (positive introspection axiom). In fact, this property is valid in all BDI systems for circumspect agents due to the B-S4 axiom and the $Intend(i,\phi) \Rightarrow Bel(i,\phi)$ property that characterises circumspect agents.

We can further refine the basic properties for R3-5 realism by imposing additional structural relations between worlds. Thus, for the axioms to apply to inevitable states of affairs, we would adopt the following relations:

$\mathcal{B}_i(w,t) \subseteq_{sub} \mathcal{I}_i(w,t)$

$\mathcal{I}_i(w,t) \cap_{sup} \mathcal{D}_i(w,t) \neq \emptyset$

The first axiom now takes the form $Intend(i, A(\phi)) \Rightarrow Bel(i, A(\phi))$ which states that if an agent intends that ϕ is inevitably true, then it believes it is inevitably true. This system will be called R3-5-BDI$_A$. By changing the structural relations between worlds, we can restrict the application of the axioms to optional formulas:

$\mathcal{B}_i(w,t) \subseteq_{sup} \mathcal{I}_i(w,t)$

$\mathcal{I}_i(w,t) \cap_{sub} \mathcal{D}_i(w,t) \neq \emptyset$

For instance, the first axiom becomes $Intend(i, E(\phi)) \Rightarrow Bel(i, E(\phi))$. It now states that if an agent intends ϕ to be optionally true, then it believes that it is optionally true. This new system will be called R3-5-BDI$_E$ in accordance with our terminology.

This notion of realism describes an agent that is very careful regarding its choice of intentions since it only intends states of affairs to be optionally true that it believes to be optionally true. However, when it comes to its desires these are not restricted by its beliefs at all; they only have to be consistent with its intentions. These desires are not grounded on the agent's beliefs. The agent then chooses its intentions so that they are consistent with its desires. Thus, this agent has more degrees of freedom regarding its desires: it desires to become rich, even thought it may not believe it.

8.2 R4-5 Realism

The main difference between R4-5 and the previous notion of realism is that the agent's desires are now consistent with its beliefs (Figure 4). The relations between worlds and the respective properties are given below:

$\mathcal{B}_i(w,t) \subseteq \mathcal{I}_i(w,t)$ **:** $Intend(i, \phi) \Rightarrow Bel(i, \phi)$

$\mathcal{B}_i(w,t) \cap \mathcal{D}_i(w,t) \neq \emptyset$ **:** $Bel(i, \phi) \Rightarrow \neg Des(i, \neg\phi)$

It now follows that the property relating desires and intentions is that of consistency:

$Intend(i, \phi) \Rightarrow \neg Des(i, \neg\phi)$

The respective system will be called R4-5-BDI according to the adopted terminology. If the BI, BD and DI conditions are imposed (Section 4.4), then apart from the three basic properties we also have the following:

Proposition 13. *The following properties are valid in R4-5-BDI if the BI, BD and DI semantic conditions are satisfied:*
1) $Intend(i, \phi) \Rightarrow Bel(i, \neg Des(i, \neg\phi))$
2) $Intend(i, \phi) \Rightarrow Bel(i, Bel(i, \phi))$
3) $Des(i, \phi) \Rightarrow Bel(i, \neg Intend(i, \neg\phi))$
4) $Intend(i, \phi) \Rightarrow Des(i, \neg Des(i, \neg\phi))$
5) $Intend(i, \phi) \Rightarrow Des(i, Bel(i, \phi))$

Proof. The proof follows in a similar way to that of Proposition 12. □

Structural relations between worlds can be imposed in order to restrict the application of the axioms to inevitable or optional formulas. The two additional systems will be called R4-5-BDI$_A$ and R4-5-BDI$_E$.

While the agent still grounds its intentions on its beliefs, in this case its desires are also consistent with its beliefs and in turn with its intentions as well. In contrast to the previous notion of realism, the agent here is more conservative. Its desires need to be consistent with its beliefs and therefore if it has a desire to become rich, it does not believe its negation.

8.3 R5-3 Realism

Another notion of realism that characterises a circumspect agent is R5-3. While the basic relation between belief- and intention-accessible worlds is maintained, now the set of desire-accessible worlds is a superset of the belief-accessible worlds. At the same time the relation between the set of desire- and intention-accessible worlds is that of consistency as shown in Figure 4. Formally, we have the following relations and corresponding axioms:

$\mathcal{B}_i(w,t) \subseteq \mathcal{I}_i(w,t)$ **:** $Intend(i,\phi) \Rightarrow Bel(i,\phi)$

$\mathcal{B}_i(w,t) \subseteq \mathcal{D}_i(w,t)$ **:** $Des(i,\phi) \Rightarrow Bel(i,\ \phi)$

It then follows that the property describing the relation between intentions and desires is:

$Intend(i,\phi) \Rightarrow \neg Des(i,\neg\phi)$

This is a different type of agent than the ones described by the previous notions of realism. An agent's intentions are still grounded on its beliefs, whereas now the same holds for its desires. Thus, when the axioms are restricted to O-formulas, an agent can only desire ϕ to be optionally true, if it believes that it is optionally true, while this desire needs to be consistent with its intentions. This notion of realism comes closer to the original notion of *strong realism*, in that both desires and intentions are grounded on beliefs. Thus, an agent that has a desire to become rich, believes it as well, even though it may not necessarily adopt it as an intention.

If the BI, BD and DI conditions are imposed (Section 4.4), then apart from the three basic properties we also have the following:

Proposition 14. *The following properties are valid in R5-3-BDI if the BI, BD and DI semantic conditions are satisfied:*
1) $Intend(i,\phi) \Rightarrow Bel(i,\neg Des(i,\neg\phi))$
2) $Intend(i,\phi) \Rightarrow Bel(i,Bel(i,\phi))$
3) $Des(i,\phi) \Rightarrow Bel(i,\neg Intend(i,\neg\phi))$
4) $Intend(i,\phi) \Rightarrow Des(i,\neg Des(i,\neg\phi))$
5) $Intend(i,\phi) \Rightarrow Des(i,Bel(i,\phi))$

Proof. The proof follows in a similar way to that of Proposition 12. □

8.4 R6-3 Realism

In R6-3 realism both sets of belief- and desire-accessible worlds are subsets of the intention-accessible worlds. At the same time, the desire-, and belief-accessible worlds are related to each other via the consistency relation (Figure 4). The set relations and the corresponding properties are provided below:

$\mathcal{B}_i(w,t) \subseteq \mathcal{I}_i(w,t)$: $Intend(i,\phi) \Rightarrow Bel(i,\phi)$
$\mathcal{D}_i(w,t) \subseteq \mathcal{I}_i(w,t)$: $Intend(i,\phi) \Rightarrow Des(i,\ \phi)$
$\mathcal{B}_i(w,t) \cap \mathcal{D}_i(w,t) \neq \emptyset$: $Bel(i,\phi) \Rightarrow \neg Des(i,\neg\phi)$

According to this notion of realism, the agent's desires need to be consistent with its beliefs about the world. The agent's intentions on the other hand are now grounded on both its desires and beliefs. Such an agent intends a state of affairs and both believes and desires it. On the other hand if it desires to become rich, this needs to be consistent with its beliefs and intentions.

The belief about intentions, beliefs about desires and desires about intentions properties (Section 4.4) are adopted by imposing the BI, BD and DI conditions.

Proposition 15. *The following properties are valid in R6-3-BDI if the BI, BD and DI semantic conditions are satisfied:*
1) $Intend(i,\phi) \Rightarrow Bel(i, Des(i,\phi))$
2) $Intend(i,\phi) \Rightarrow Bel(i, Bel(i,\phi))$
3) $Des(i,\phi) \Rightarrow Bel(i,\neg Intend(i,\neg\phi))$
4) $Des(i,\phi) \Rightarrow Bel(i,\neg Bel(i,\neg\phi))$
5) $Intend(i,\phi) \Rightarrow Des(i, Des(i,\phi))$
6) $Intend(i,\phi) \Rightarrow Des(i, Bel(i,\phi))$

Proof. The proof follows in a similar way to that of Proposition 12. □

9 Bold Agents

The basic constraint that characterises bold agents in terms of set relations is that the set of intention-accessible worlds is a subset of the set of belief-accessible worlds or alternatively the two sets are related via the consistency relation. Hence, the main feature of such an agent is that it intends a state of affairs as long as it does not believe its negation, or in other words an agent adopts an intention as long as it does not contradict its beliefs. The following sections describe the four notions of realism for bold agents illustrated in Figure 5.

9.1 R3-6 Realism

The first notion of realism that characterises a bold agent is R3-6 and belongs to category 3. In R3-6 realism, the sets of intention- and belief-accessible worlds are related via the consistency relation and thus the basic property for bold agents holds. While the set of desire-accessible worlds is a subset of the intention-accessible worlds, it is completely decoupled from the set of belief-accessible

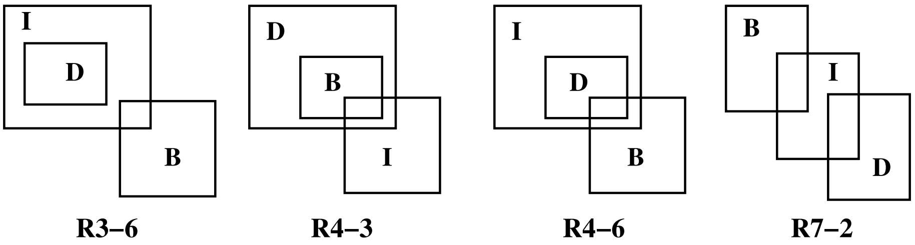

Fig. 5. Notions of realism for bold agents.

worlds (Figure 5). The following properties relate the modalities according to R3-6 realism

$Intend(i, \phi) \Rightarrow Des(i, \phi)$

$Intend(i, \phi) \Rightarrow \neg Bel(i, \neg\phi)$

Lemma 7. *The properties for R3-5 realism are sound in all models that satisfy the semantic conditions:*
(i) $\mathcal{D}_i(w,t) \subseteq \mathcal{I}_i(w,t) : \forall w, w', t\ \mathcal{D}_i(w,t,w') \Rightarrow \mathcal{I}_i(w,t,w')$
(ii) $\mathcal{I}_i(w,t) \cap \mathcal{B}_i(w,t) \neq \emptyset : \forall w, t, \exists w'\ \mathcal{I}_i(w,t,w') \wedge \mathcal{B}_i(w,t,w')$

Proof. (i) Assume that $M(v,w,t) \models Intend(i,\phi)$ for an arbitrary $M(v,w,t)$. According to semantics we have $M(v,w',t) \models \phi$ for all w' such that $\mathcal{I}_i(w,t,w')$. Since $\mathcal{D}_i(w,t) \subseteq \mathcal{I}_i(w,t)$ we have $M(v,w'',t) \models \phi$ for all w'' such that $\mathcal{D}_i(w,t,w'')$. It now follows that $M(v,w,t) \models Des(i,\phi)$.

(ii) Assume $M(v,w,t) \models Intend(i,\phi)$. Then for all w' such that $\mathcal{I}_i(w,t,w')$ we have $M(v,w',t) \models \phi$ and thus $M(v,w',t) \not\models \neg\phi$. Since $\mathcal{I}_i(w,t) \cap \mathcal{B}_i(w,t)$, there is at least one world w' such that $\mathcal{B}_i(w,t,w')$. It now follows that $M(v,w,t) \not\models Bel(i,\neg\phi)$ and hence $M(v,w,t) \models \neg Bel(i,\neg\phi)$ as required. □

The system consisting of the basic BDI system and the axioms for R3-6 realism is called R3-6-BDI.

Further constraints can be imposed in order to capture the properties described in Section 4.4. In particular, if the BI and DI conditions are imposed we have the following properties:

$Intend(i,\phi) \Rightarrow Bel(i, Intend(i,\phi))$ (beliefs about intentions)

$Intend(i,\phi) \Rightarrow Des(i, Intend(i,\phi))$ (desires about intentions)

The semantic condition BD (beliefs about desires) is not imposed since in this particular notion of realism the sets of the belief- and desire-accessible worlds are not related with each other.

Proposition 16. *The following properties are valid in R3-6-BDI if the BI and DI semantic conditions are satisfied:*
1) $Intend(i,\phi) \Rightarrow Bel(i, Des(i,\phi))$
2) $Intend(i,\phi) \Rightarrow Bel(i, \neg Bel(i,\neg\phi))$
3) $Intend(i,\phi) \Rightarrow Des(i, Des(i,\phi))$
4) $Intend(i,\phi) \Rightarrow Des(i, \neg Bel(i,\neg\phi))$

Proof. 1) Assume $Intend(i, \phi)$. By the belief of intentions property we have $Intend(i, \phi) \Rightarrow Bel(i, Intend(i, \phi))$ and by modus ponens $Bel(i, Intend(i, \phi))$ (*). From the R3-6 realism properties we have $Intend(i, \phi) \Rightarrow Des(i, \phi)$ and by necessitation for belief $Bel(i, Intend(i, \phi) \Rightarrow Des(i, \phi))$. By distribution of belief $Bel(i, Intend(i, \phi)) \Rightarrow Bel(i, Des(i, \phi))$ and from (*) and modus ponens $Bel(i, Des(i, \phi))$ is obtained.

2) Assume $Intend(i, \phi)$. By the belief of intentions property $Intend(i, \phi) \Rightarrow Bel(i, Intend(i, \phi))$ and by modus ponens $Bel(i, Intend(i, \phi))$ (*). From the R3-6 realism properties we have $Intend(i, \phi) \Rightarrow \neg Bel(i, \neg\phi)$ and by necessitation for belief $Bel(i, Intend(i, \phi) \Rightarrow \neg Bel(i, \neg\phi))$. By distribution of belief $Bel(i, Intend(i, \phi)) \Rightarrow Bel(i, \neg Bel(i, \neg\phi))$ and from (*) and modus ponens $Bel(i, \neg Bel(i, \neg\phi))$ is obtained.

3) Assume $Intend(i, \phi)$. By the desires of intentions property $Intend(i, \phi) \Rightarrow Des(i, Intend(i, \phi))$ and by modus ponens $Des(i, Intend(i, \phi))$ (*). From the R3-6 realism properties we have $Intend(i, \phi) \Rightarrow Des(i, \phi)$ and by necessitation for desires $Des(i, Intend(i, \phi) \Rightarrow Des(i, \phi))$. By distribution of desires $Des(i, Intend(i, \phi)) \Rightarrow Des(i, Des(i, \phi))$ and from (*) and modus ponens we obtain $Des(i, Des(i, \phi))$.

4) Assume $Intend(i, \phi)$. By the desires of intentions property $Intend(i, \phi) \Rightarrow Des(i, Intend(i, \phi))$ and by modus ponens $Des(i, Intend(i, \phi))$ (*). From the R3-6 realism properties we have $Intend(i, \phi) \Rightarrow \neg Bel(i, \neg\phi)$ and by necessitation for desires $Des(i, Intend(i, \phi) \Rightarrow \neg Bel(i, \neg\phi))$. By distribution of desires $Des(i, Intend(i, \phi)) \Rightarrow Des(i, \neg Bel(i, \neg\phi))$ and from (*) and modus ponens $Des(i, \neg Bel(i, \neg\phi))$ is obtained. □

The basic properties of R3-6 realism can be refined further by adopting additional structural relations between worlds. Thus for expressing attitudes towards inevitabilities the following conditions are imposed:

$\mathcal{D}_i(w, t) \subseteq_{sub} \mathcal{I}_i(w, t)$

$\mathcal{I}_i(w, t) \cap_{sup} \mathcal{B}_i(w, t) \neq \emptyset$

The first property takes the form $Intend(i, A(\phi)) \Rightarrow Des(i, A(\phi))$, which states that if an agent intends that ϕ is inevitably true, then it desires it to be inevitably true as well. This system is called R3-6-BDI$_A$. For expressing attitudes towards options the conditions below are imposed:

$\mathcal{D}_i(w, t) \subseteq_{sup} \mathcal{I}_i(w, t)$

$\mathcal{I}_i(w, t) \cap_{sub} \mathcal{B}_i(w, t) \neq \emptyset$

The first property becomes $Intend(i, E(\phi)) \Rightarrow Des(i, E(\phi))$. It states that if an agent intends that ϕ is optionally true, then it desires it to be optionally true. This system is named R3-6-BDI$_E$ according to our terminology.

This notion of realism describes a bold agent with respect to its intentions. However, the agent's desires are completely decoupled from its beliefs. As such, desires can be considered to represent states of affairs that although the agent would ideally like to bring about, it may never come to intend. So, an agent may desire to become rich, without necessarily adopting it as an intention, since this may not be consistent with its beliefs.

9.2 R4-3 Realism

Although as in the previous notion of realism the main property for bold agents holds by adopting the consistency relation between the two sets, in R4-3 realism desires are a superset of the agent's beliefs (Figure 5). Formally:

$\mathcal{B}_i(w,t) \subseteq \mathcal{D}_i(w,t)$ **:** $Des(i,\phi) \Rightarrow Bel(i,\phi)$

$\mathcal{B}_i(w,t) \cap \mathcal{I}_i(w,t) \neq \emptyset$ **:** $Bel(i,\phi) \Rightarrow \neg Intend(i,\neg\phi)$

R4-3-BDI consists of the basic BDI system and the axioms for R4-3 realism. It now follows that desires are related to intentions via the following property:

$Des(i,\phi) \Rightarrow \neg Intend(i,\neg\phi)$

Furthermore, the BI, BD and DI conditions can be imposed (Section 4.4) in order to obtain the beliefs about intentions, beliefs about desires and desires about intentions schemas.

Proposition 17. *The following properties are valid in R4-3-BDI if the BI, BD and DI semantic conditions are satisfied:*
1) $Intend(i,\phi) \Rightarrow Bel(i,\neg Des(i,\neg\phi))$
2) $Intend(i,\phi) \Rightarrow Bel(i,\neg Bel(i,\neg\phi))$
3) $Des(i,\phi) \Rightarrow Bel(i,\neg Intend(i,\neg\phi))$
4) $Des(i,\phi) \Rightarrow Bel(i,Bel(i,\phi))$
5) $Intend(i,\phi) \Rightarrow Des(i,\neg Des(i,\neg\phi))$
6) $Intend(i,\phi) \Rightarrow Des(i,\neg Bel(i,\neg\phi))$

Proof. The proof follows in a similar way to that of Proposition 16. □

As previously, the basic properties for R4-3 realism can be further refined by adopting additional structural relations between worlds.

According to this notion of realism, the agent's intentions need to be consistent with both its beliefs and desires. In this particular case, desires may be considered to be the agent's options and these options are grounded on its beliefs. Thus, an agent's desire to become rich means that it believes it as well. The agent's intentions have to be consistent with both its desires and beliefs about the world.

9.3 R4-6 Realism

Another notion of realism that characterises a bold agent is R4-6. This is in fact very closely related to R3-6. The sets of intention- and belief-accessible worlds are related via the consistency relation as are the sets of desire- and belief-accessible worlds (Figure 5). The relations between worlds and the respective properties are given below:

$\mathcal{D}_i(w,t) \subseteq \mathcal{I}_i(w,t)$ **:** $Intend(i,\phi) \Rightarrow Des(i,\phi)$

$\mathcal{D}_i(w,t) \cap \mathcal{B}_i(w,t) \neq \emptyset$ **:** $Des(i,\phi) \Rightarrow \neg Bel(i,\neg\phi)$

It now follows that intentions and beliefs are related via the property:

$Intend(i,\phi) \Rightarrow \neg Bel(i,\neg\phi)$

The belief about intentions, beliefs about desires and desires about intentions properties (Section 4.4) can be adopted by imposing the BI, BD and DI conditions.

Proposition 18. *The following properties are valid in R4-6-BDI if the BI, BD and DI semantic conditions are satisfied:*
1) $Intend(i, \phi) \Rightarrow Bel(i, Des(i, \phi))$
2) $Intend(i, \phi) \Rightarrow Bel(i, \neg Bel(i, \neg\phi))$
3) $Des(i, \phi) \Rightarrow Bel(i, \neg Intend(i, \neg\phi))$
4) $Des(i, \phi) \Rightarrow Bel(i, \neg Bel(i, \neg\phi))$
5) $Intend(i, \phi) \Rightarrow Des(i, Des(i, \phi))$
6) $Intend(i, \phi) \Rightarrow Des(i, \neg Bel(i, \neg\phi))$

Proof. The proof follows in a similar way to that of Proposition 16. □

An agent based on this notion of realism is bold regarding its intentions and thus if it intends a state of affairs, then it does not believe its negation. Moreover, if it intends a state of affairs, then it desires it as well. However, an agent's desire to become rich need only be consistent with its beliefs and intentions. It does not necessarily follow that if an agent has such a desire, this is adopted as an intention.

9.4 R7-2 Realism

Finally, the last notion of realism describing a bold agent is R7-2. According to Figure 5 the set of intention-accessible worlds is related to that of the belief-accessible worlds via the consistency relation, while the set of intention-accessible worlds is related to the set of desire-accessible worlds via the consistency relation as well. Formally we have the following:

$\mathcal{B}_i(w,t) \cap \mathcal{I}_i(w,t) \neq \emptyset$ **:** $Bel(i, \phi) \Rightarrow \neg Intend(i, \neg\phi)$
$\mathcal{I}_i(w,t) \cap \mathcal{D}_i(w,t) \neq \emptyset$ **:** $Intend(i, \phi) \Rightarrow \neg Des(i, \neg\phi)$

The belief about intentions and desires about intentions properties (Section 4.4) can be adopted by imposing the BI and DI conditions. The beliefs about desires is not considered here since the sets of belief- and desire-accessible worlds are not related with each other.

Proposition 19. *The following properties are valid in R7-2-BDI if the BI and DI semantic conditions are satisfied:*
1) $Intend(i, \phi) \Rightarrow Bel(i, \neg Des(i, \neg\phi))$
2) $Intend(i, \phi) \Rightarrow Bel(i, \neg Bel(i, \neg\phi))$
3) $Intend(i, \phi) \Rightarrow Des(i, \neg Des(i, \neg\phi))$
4) $Intend(i, \phi) \Rightarrow Des(i, \neg Bel(i, \neg\phi))$

Proof. The proof follows in a similar way to that of Proposition 16. □

An agent based upon R7-2 realism intends states of affairs that are consistent with its beliefs and desires. However, its desires are completely decoupled from its beliefs. Desires may again be considered as expressing what the agent would ideally like to bring about. An agent's desire to become rich in this particular notion of realism may not be consistent with its beliefs. However, if an agent has an intention to become rich this needs to be consistent with both its beliefs and desires. This is the preferred choice of realism of [27], in particular their preferred system is R7-2-BDI$_E$.

Table 8. Satisfaction of AA and CC in notions of realism for circumspect agents

System	A1	A2	A3	A4	A5	A6	A7	A8	A9	C1	C2	C3
R3-5-BDI	T	F	T	T	T	T	F	T	T	T	T	T
R4-5-BDI	T	F	T	T	T	T	T	T	T	T	T	T
R5-3-BDI	T	F	T	T	T	T	T	T	F	T	T	T
R6-3-BDI	T	F	F	T	T	T	T	T	T	T	T	T

10 Comparisons and Results

As has been shown so far, a number of notions of realism can be uncovered by combining the set theoretic relations. We distinguished BDI agents into two main categories according to the relation between intentions and beliefs: circumspect and bold and we presented four notions of realism for each of these categories. Since the concepts of a bold and circumspect agent share their main characteristic property with *strong realism* and *realism* respectively, our aim has been to uncover notions of realism that come closer to the desiderata for rational BDI agents as discussed in [3, 27]. We will now proceed with the evaluation of these notions of realism.

In order to put this work into context, recall that the general system based on *strong realism*, S-BDI, although it satisfies all Consequential Closure (CC) principles, it does not satisfy three of the Asymmetry Thesis (AT) principles. Table 8 summarises the satisfaction of the AT and the CC principles for the four general systems for circumspect agents that were presented. The main feature of circumspect agents is that if an agent intends ϕ then it believes it; this means that due to this very definition, A2 will never be satisfied for such agents. Evidently, the third column in Table 8 indicates that in all four systems A2 is not satisfied. However, otherwise it seems that all four notions for circumspect agents come closer to the desiderata for rational BDI agents than that of *strong realism*. Depending on the additional set relations, R3-5-BDI, R5-3-BDI and R6-3-BDI do not satisfy one AT principle in addition to A2. R4-5-BDI is the only system that does not satisfy one principle, namely A2. In conclusion, it seems that all four notions of realism for circumspect agents improve on *strong realism*, while at the same time maintaining their main attribute.

Table 9 summarises the satisfiability of the AT and CC principles for the four general systems for bold agents. As was discussed earlier in the paper, the general system based on *realism,* R-BDI, does not satisfy the CC principles and three of the AT principles. The first improvement that can be noticed by inspecting Table 9 is that all systems based on the notions of realism for bold agents satisfy the CC principles. This on its own is a major improvement on *realism* since all three general forms of the side-effect problem are avoided. Moreover, the first system R3-6-BDI does not satisfy two AT principles, whereas the other three do not satisfy one. In all, it seems that we have managed to improve on the notion of *realism* without however losing the main feature of bold agents.

However, we have not managed to improve on *weak realism*, since this constitutes the top of our triangle satisfying all principles. We have only managed

Table 9. Satisfaction of AA and CC in notions of realism for bold agents

System	A1	A2	A3	A4	A5	A6	A7	A8	A9	C1	C2	C3
R3-6-BDI	T	T	F	T	T	T	F	T	T	T	T	T
R4-3-BDI	T	T	T	T	T	T	T	T	F	T	T	T
R4-6-BDI	T	T	F	T	T	T	T	T	T	T	T	T
R7-2-BDI	T	T	T	T	T	T	F	T	T	T	T	T

to move up towards *weak realism* while at the same time remaining within our definitions of circumspect and bold agents. Additional notions of realism may be considered to have attractive properties for rational BDI agents depending on the application domain.

11 Conclusions

Drawing its inspiration from Bratman's philosophical work on practical reasoning, the BDI paradigm has been one of the most intensively studied frameworks for agents. One of the most interesting issues in this framework is that of the relations between the three attitudes. Although notions of realism for BDI agents were studied in previous work [25–27], no systematic attempt has been made to properly investigate other available options. The aim of this paper was to contribute to this area by thoroughly exploring notions of realism for BDI agents.

The contribution of this paper is two-fold. Firstly, we explored the dynamics between the three attitudes and we provided two ways of categorising notions of realism for BDI agents along two very different dimensions: according to the set relations between accessible worlds, and according to the relation between an individual agent's intentions and beliefs. The latter has led to the second aspect of our contribution: deriving their main feature from *realism* and *strong realism* , agents were distinguished into bold and circumspect respectively. Notions of realism for each of these categories were presented and as was shown, there are certain notions of realism that offer better properties than *realism* and *strong realism*. The work presented here supports the argument of [27] that there may not be a unique BDI system suitable for all applications; the designer is free to choose an appropriate BDI system according to the specific requirements of a particular application. Finally, the additional properties that were considered in [25] (Section 4.4) are re-visited here. It turns out the semantic conditions described in [25] are not correct. We thus provide new conditions that are able to capture the intended properties of beliefs about intentions, beliefs about desires and desires about intentions. Moreover, we consider these properties in the context of the three original notions of realism as well as in the context of the notions of realism for bold and circumspect agents.

There are a number of possible avenues for future development. The work presented here is based upon the possible worlds framework. However, although modal logics and possible worlds are a very powerful and convenient tool for constructing theories for reasoning agents they suffer from the logical omniscience problem [9]. Thus an agent believes, desires and intends all consequences of its beliefs, desires and intentions respectively. This has further repercussions that come in the form of the side-effect problem (see section 6). Moreover, the use of possible worlds for modelling intentions does not allow for relating intentions to one another. Agents often adopt intentions in support to other intentions [18], and in the current formalism this cannot be captured. Future work needs to address this issue. Finally, although this work offers an insight into the different relations between the main attitudes of the Belief-Desire-Intention model as well as a systematic categorisation of types of agents, it does not offer any criteria for choosing among the available types of agents. This is open to further investigation. A farther goal is to study the applicability of these models of agents to real applications such as trading agents for the stock market or auctions and agents with "personality". This would also contribute towards bridging the gap between theory and application.

References

1. Anscombe, G. E. M. (1963). *Intention*. Ithaca, N.Y.: Cornell University Press.
2. Audi, R. (1986). *Action, decision and intention*. Dordrecht/Boston: D. Reidel Publishing Company.
3. Bratman, M. E. (1987). *Intentions, plans and practical reason*. Cambridge, MA: MIT Press.
4. Chellas, B. (1980). *Modal logic: An introduction*. Cambridge, England: Cambridge University Press.
5. Cohen, P. R., & Levesque, H. J. (1990). Intention is choice with commitment. *Artificial intelligence, 42,* 213-261.
6. Davidson, D. (1980). Actions, reasons and causes. In *Essays on actions and events*. New York: Oxford University Press, pp. 3-19.
7. Dennett, D. C. (1987). *The intentional stance*. Cambridge, MA: MIT Press.
8. Emerson, E. A. (1990). Temporal and modal logic. In van Leeuweb, J. (Ed.), *Handbook of theoretical computer science: Formal models and semantics, Volume B*. Amsterdam and Cambridge, MA: Elsevier Publishers and MIT Press, pp. 995-1072.
9. Fagin, R., Halpern, J. Y., Moses, Y., & Vardi, M. Y. (1995). *Reasoning about Knowledge*. Cambridge, MA: MIT Press.
10. Fasli, M. (2000). *Commodious logics of agents*. Ph.D. Thesis. Department of Computer Science, University of Essex.
11. Fasli, M. (2002). Heterogeneous BDI agents. Technical report CSM-372. Department of Computer Science, University of Essex.
12. Fodor, J. A. (1978). Propositional attitudes. *The Monist, LXI,* 4, 501-523.
13. Hanser, M. (1998). Intention and teleology. *Mind, 107,* 381-402.
14. Hintikka, J. (1962). *Knowledge and belief*. Ithaca, N.Y.: Cornell University Press.
15. van der Hoek, W. (1990). Systems for knowledge and beliefs. In *Proceedings of Logics in Artificial Intelligence European Workshop*. LNAI:478. Berlin: Springer-Verlag, pp. 267-281.
16. van der Hoek, W., van Linder, J., & Meyer, J.-J. Ch. (1998). An integrated modal approach to rational agents. In Wooldridge, M., & Rao, A. (Eds.) *Foundations of rational agency*. Applied Logic Series 14. Dordecht: Kluwer, pp. 133-168.

17. Hughes, G. E., & Cresswell, M. J. (1968). *An introduction to modal logic*. London: Methew & Co Ltd.
18. Konolige, K., & Pollack, M. (1993). A representationalist theory of intention. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence*. pp. 390-395.
19. Kraus, S., & Lehmann, D. (1988). Knowledge, belief, and time. *Theoretical Computer Science, 58,* 155-174.
20. Kripke, S. A. (1963). Semantical analysis of modal logic. *Zeitschrift fur Mathematische Logik und Grundlagen der Mathematik, 9,* 67-96.
21. van Linder, B., van der Hoek, W., & Meyer, J.-J. Ch. (1996). Formalising motivational attitudes of agents: On preferences, goals and commitments. In Wooldridge, M., Muller, J.P., & Tambe, M. (Eds.) *Intelligent agents II: Agent theories, architectures and languages*. LNAI:1037. Berlin: Springer-Verlag, pp. 17-32.
22. McCarthy, J. (1979). Ascribing mental qualities to machines. In Ringle, M. (Ed.), *Philosophical perspectives in artificial intelligence*. Brighton, Sussex: The Harvester Press Limited, pp. 161-195.
23. Montague, R. (1974). The proper treatment of quantification in ordinary English. In Thomason, R. (Ed.), *Formal philosophy: Selected papers of Richard Montague*. New Haven and London: Yale University Press, pp. 247-270.
24. Quine, W. V. (1966). *The ways of paradox and other essays*. Cambridge, MA: Harvard University Press.
25. Rao, A., & Geogeff, M. (1991a). Modeling rational agents within a BDI architecture. In *Proceedings of the 2nd Principles of Knowledge Representation and Reasoning Conference*. San Mateo, CA.: Morgan Kaufmann Publishers, pp. 473-484.
26. Rao, A., & Georgeff, M. (1991b). Asymmetry thesis and side-effect problems in linear time and branching time intention logics. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence*. San Mateo, CA.: Morgan Kaufmann Publishers, pp. 498-504.
27. Rao, A., & Geogeff, M. (1998). Decision procedures for BDI logics. *Journal of Logic and Computation, 8,* 3, 293-342.
28. Salmon, N. & Soames, S. (1988). *Propositions and attitudes*. New York: Oxford University Press.
29. Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge: Cambridge University Press.
30. Thomason, R. (1980). A note on syntactical treatments of modality. *Synthese 44,* 391-395.
31. Voorbraak, F. (1990). The logic of objective knowledge and rational belief. In *Proceedings of Logics in Artificial Intelligence European Workshop*. LNAI:478. Berlin: Springer-Verlag, pp. 499-515.
32. Wooldridge, M., & Jennings, N. R. (1995). Intelligent agents: theory and practice. *Knowledge Engineering Review, 10,* (2), 115-152.
33. Wooldridge, M. (2000). *Reasoning about rational agents*. Cambridge, MA: MIT Press.