Relating Truth, Knowledge and Belief in epistemic states

Costas D. Koutras[1]
ckoutras@uop.gr

Yorgos Zikos[2]
zikos@sch.gr

[1]Dept. of Informatics and Telecommunications,
University of Peloponnese
end of Karaiskaki Street, 22100 Tripolis, Greece

[2]Graduate Programme in Logic, Algorithms and Computation
M.P.L.A., Dept. of Mathematics, University of Athens
Panepistimiopolis, 157 84 Ilissia, Greece

December 2, 2015

Abstract

Epistemic states are very useful in Knowledge Representation, in particular for defining logics of *minimal knowledge* in *NonMonotonic Reasoning*. The most successful example has been the introduction of *stable belief sets* by R. Stalnaker, a notion which heavily influenced the development of modal nonomonotonic logics. Up to now, the proposed epistemic states do not distinguish between *knowledge* and *belief*, focusing mostly on an analysis of a rational agent's *introspective power*. We define and investigate here a structure incorporating *what is true*, *what is known* and *what is believed* by a rational agent in possible worlds models. The notion of KB_R-structures introduced, provides a fine-grained modal analysis of an agent's epistemic state, actually one that differentiates knowledge from belief and accounts for an agent without full introspective power concerning knowledge. Many epistemic properties of this structure are proved and it is shown that belief collapses in the form of a Stalnaker stable set, while knowledge does not. Finally, a representation theorem is proved, which matches KB_R-structures to models of the logic **S4.2**, advocated by W. Lenzen as the 'correct' logic of knowledge, a statement further supported by the work of R. Stalnaker and other researchers.

1 Introduction

Epistemic Logic [Hin62, Len79] has been traditionally concerned with the rigorous analysis of the propositional attitudes '$agent_i$ *knows* φ' and '$agent_i$ *believes that* φ *holds*'. It grew up as an area of *Philosophical Logic* but it has been given a fresh new perspective and a strong motivation through its applications in *Computer Science* (for instance the *analysis of distributed systems* [FHMV03]) and *Artificial Intelligence* (*autoepistemic logics* [MT93], *multi-agent systems* [Woo09] and many others). In its current form, Epistemic Logic has been greatly benefited by the development of Modal Logic and, in particular, by the advent of '*possible worlds*' (or Kripke) *semantics*. Nowadays, many rich epistemic languages have been introduced and applied in various fields of computing; see [vB10] for a short presentation and many pointers to the literature. *Epistemic Logic* has recently met *Dynamic Logic* in an area which deals with the dynamic phenomena of *public announcements*, '*rumours*' and other actions which affect the knowledge state in a group of agents: *Dynamic Epistemic Logic* [vDvdHK07] deals with *logics of knowledge and change*.

Artificial Intelligence has provided a new, '*introspective*' perspective on modal epistemic reasoning. In *Knowledge Representation*, the issue of a '*good*' representation of a rational agent's (typically acting in a domain of interest and holding partial, incomplete information about the world) epistemic state is very important. A simple, yet very successful and influential notion is Stalnaker's definition of a **stable belief set** ([Sta93], [MT93]), which has played a significant role in the development of modal *Non-Monotonic Reasoning* (NMR). Succint and expressive logical definitions of an agent's epistemic state are of interest to other branches of Knowledge Representation too, such as *belief revision* and *reasoning about actions*.

In this paper we proceed to work on a detailed analysis of the **epistemic** and **doxastic theories** held by a rational agent, operating in a complex possible-worlds environment, under the realistic condition that the information acquired by the agent allows him to distinguish (at least) some of the possible worlds in the picture. This is definitely different from the **S5** picture of the Stalnaker stable sets, worked around the universal model paradigm, where no possible world is distinguishable for the others. Here, we actually place the (important for KR) question of the formal representation of an agent's knowledge and belief, under the lens of classical modal epistemic reasoning and revisit the notion of epistemic state(s) under a new, semantic perspective. Our **objective** is to describe the *epistemic* and *doxastic* status of a rational agent *without* full introspection (which has been strongly criticized in epistemic logic), taking a modal approach, which **differentiates knowledge from belief**. We introduce a notion of KB_R-structures, intending to capture the **interplay between truth, knowledge and belief** held by an agent operating in a domain modelled as a set of possible-worlds. We examine several proof-theoretic properties of KB_R-structures and provide a representation theorem for these structures, which proves an exact correspondence to the models of **S4.2**, the logic advocated by W. Lenzen as the '*correct*' logic of knowledge [Len79]. It is hardly surprising that the initial motivation of this research has been the ambition to define simple

variants of Stalnaker's stable sets inspired from interesting epistemic models, such as the models of **S4.2**.

The paper is organized as follows: in Section 2 we establish notation and terminology. In Section 3.1 we provide a motivating example for the epistemic states introduced in this paper. In the rest of Section 3 we define the KB_R-structures and examine their formal properties. In Section 4 we prove a representation theorem which links KB_R-structures with models of the logic **S4.2**. In Sections 5 we provide a detailed example for an epistemic situation, in view of our results in the previous sections. We conclude in Section 6 with some references to related work and some questions for further research.

2 Notation and Terminology

2.1 Modal Logic

In this section we gather the necessary background material and results: for the basics of *Modal Logic* and *modal Non-Monotonic Reasoning* the reader is referred to the books [BdRV01, Che80, HC96, MT93]. We assume a modal propositional language $\mathcal{L}_\Box$, endowed with an epistemic operator $\Box\varphi$, read as '*it is known that* φ *holds*'. Sentence symbols include $\top$ (for *truth*) and $\bot$ (for *falsity*). Some of the important axioms in epistemic/doxastic logic are:

K. $(\Box\varphi \wedge \Box(\varphi \supset \psi)) \supset \Box\psi$

T. $\Box\varphi \supset \varphi$ (axiom of true, justified knowledge)

4. $\Box\varphi \supset \Box\Box\varphi$ (axiom of positive introspection)

5. $\neg\Box\varphi \supset \Box\neg\Box\varphi$ (axiom of negative introspection)

G. $\neg\Box\neg\Box\varphi \supset \Box\neg\Box\neg\varphi$

The epistemic interpretation of **G** will be made clear below. **Modal logics** are sets of modal formulas containing classical propositional logic (i.e. containing all tautologies in the augmented language $\mathcal{L}_\Box$) and closed under rule

$$\mathbf{MP}.\frac{\varphi, \varphi \supset \psi}{\psi}$$

The smallest modal logic is denoted as **PC** (propositional calculus in the augmented language). **Normal** are called those modal logics, which contain all instances of axiom **K** and are closed under the **rule of generalization**

$$\mathbf{RN}.\ \frac{\varphi}{\Box\varphi}$$

By $\mathbf{KA_1} \ldots \mathbf{A_n}$ we denote the normal modal logic axiomatized by axioms $\mathbf{A_1}$ to $\mathbf{A_n}$. Well-known epistemic logics comprise **KT45** (**S5**) (a *strong logic of knowledge*) and

KT4G (**S4.2**). Throughout this paper we use the **notion of strong provability from a theory** I [MT93]. In the case of a normal modal logic Λ we write $I \vdash_\Lambda \varphi$ iff there is a Hilbert-style proof, where each step of the proof is a formula, which is a tautology in $\mathcal{L}_\Box$, or an instance of **K**, or an instance of an axiom of Λ, or a member of I, or a result of applying *Uniform Substitution*, **MP** or **RN** to formulas of previous steps. We say that a theory I is *consistent with logic* Λ (denoted as: $c\Lambda$) iff $I \nvdash_\Lambda \bot$. Theory Θ is *I-consistent with* Λ ($Ic\Lambda$) iff $(\forall n \in \mathbb{N})(\forall \varphi_0, \ldots, \varphi_n \in \Theta)\ I \nvdash_\Lambda \varphi_0 \wedge \ldots \wedge \varphi_n \supset \bot$, and theory Θ is *maximal I-consistent with* Λ ($mIc\Lambda$) iff Θ is $Ic\Lambda$ and $(\forall \psi \notin \Theta)\ \Theta \cup \{\psi\}$ is not I-consistent with Λ ($Iinc\Lambda$).

Furthermore, we say that I **is closed under Λ-consequence** iff $I = Cn_\Lambda(I)$. By definition, $Cn_\Lambda(I) = \{\varphi \in \mathcal{L}_\Box \mid I \vdash_\Lambda \varphi\}$. The notion of proof $\vdash_\Lambda$ depends on Λ. Except of modus ponens, in case of a normal modal logic Λ, it contains generalization. If propositional logic $\mathbf{PC}_\mathcal{L}$ is considered, and $I \subseteq \mathcal{L}$ (as in Prop.4.3 later on), then we say that I is closed under propositional consequence iff $I = Cn_{\mathbf{PC}_\mathcal{L}}(I)$. This time $Cn_{\mathbf{PC}_\mathcal{L}}(I) =_{\text{def.}} \{\varphi \in \mathcal{L} \mid I \vdash_{\mathbf{PC}_\mathcal{L}} \varphi\}$, and proof $\vdash_{\mathbf{PC}_\mathcal{L}}$ contains only modus ponens.

Normal modal logics are interpreted over **Kripke models**: a *Kripke model* $\mathfrak{M} = \langle W, R, V\rangle$ consists of a set of *possible worlds* (*states, situations*) W and a binary accessibility relation between them $R \subseteq W \times W$: whenever wRv, we say that world w '*sees*' world v, or that v is an alternative to w. The valuation V determines which propositional variables are true inside each possible world. Within a world w, the propositional connectives ($\neg$, $\supset$, $\wedge$, $\vee$) are interpreted classically, while $\Box\varphi$ is true at w iff it is true in every world '*seen*' by w (notation: $\mathfrak{M}, w \Vdash \Box\varphi$). The pair $\mathfrak{F} = \langle W, R\rangle$ is called the *frame* underlying $\mathfrak{M}$. A logic Λ is *determined* by a class of frames iff it is *sound* and *complete* with respect to this class; it is known that **S5** is determined by the class of frames with a *universal* accessibility relation, while **S4.2** is determined by the class of frames with a *reflexive, transitive* and *directed*[1] accessibility relation [Gol92].

2.2 Stable belief sets

The following notion has been very influential in NonMonotonic Reasoning. **Stable belief sets**, were introduced by R. Stalnaker in the early '80s [Sta93] as a formal representation of the epistemic state of an ideally rational agent, with full introspective capabilities. A set of formulas S in a monomodal epistemic language is a stable set if it is 'stable' under classical inference and epistemic introspection:

(i) $Cn_{\mathbf{PC}}(S) \subseteq S$

(ii) $\varphi \in S$ implies $\Box\varphi \in S$

(iii) $\varphi \notin S$ implies $\neg\Box\varphi \in S$

[1] i.e. $(\forall w, v \in W)(\exists u \in W)(wRu\ \&\ vRu)$.

2.3 A digression on Bimodal Epistemic Logics and the epistemic content of S4.2

W. Lenzen has advocated in [Len79] that **S4.2** is the 'correct' logic of knowledge and belief. His results are further supported by R. Stalnaker's work [Sta06] who has arrived at **S4.2** through a different (but equivalent) set of epistemic principles. Our perspective is very much influenced by W. Lenzen's work in [Len79], where many interesting formulations of knowledge and belief are discussed. To explain briefly the epistemic importance of **S4.2** we will move temporarily to a bimodal language in order to express axioms that capture the interplay between knowledge (K) and belief (B). Firstly let us explain that the 'basic' epistemic logic is $\mathbf{S4}_{\mathsf{K}}$ axiomatized by $\mathbf{K}_{\mathsf{K}}$. $\mathsf{K}\varphi \wedge \mathsf{K}(\varphi \supset \psi) \supset \mathsf{K}\psi$, $\mathbf{T}_{\mathsf{K}}$. $\mathsf{K}\varphi \supset \varphi$ and $\mathbf{4}_{\mathsf{K}}$. $\mathsf{K}\varphi \supset \mathsf{K}\mathsf{K}\varphi$. Another important logic is doxastic $\mathbf{KD45}_{\mathsf{B}}$, axiomatized by $\mathbf{K}_{\mathsf{B}}$, $\mathbf{4}_{\mathsf{B}}$ and the axioms $\mathbf{D}_{\mathsf{B}}$. $\mathsf{B}\varphi \supset \neg\mathsf{B}\neg\varphi$ and $\mathbf{5}_{\mathsf{B}}$. $\neg\mathsf{B}\varphi \supset \mathsf{B}\neg\mathsf{B}\varphi$.

The so called *bridge axioms* or *interaction axioms* attempt to capture the interaction between the two attitudes. Following is a list of the most important ones, with the name R. Stalnaker uses in [Sta06]; inside the parenthesis is the name used by W. Lenzen for the same axiom, if the name is different.

KB. $\mathsf{K}\varphi \supset \mathsf{B}\varphi$

Knowledge implies belief. (**B1**, *entailment property* in [Hal96])

(**B2.3**) $\mathsf{B}\varphi \supset \neg\mathsf{B}\neg\mathsf{K}\varphi$

Assuming that something is believed to be true, it cannot be the case that it is believed not to be known[2].

PIB. $\mathsf{B}\varphi \supset \mathsf{K}\mathsf{B}\varphi$

Positive introspection regarding belief. (**B2.4**)

NIB. $\neg\mathsf{B}\varphi \supset \mathsf{K}\neg\mathsf{B}\varphi$

Negative introspection regarding belief.

SB. $\mathsf{B}\varphi \supset \mathsf{B}\mathsf{K}\varphi$

'strong belief' – 'subjective certainty'. (**B2.1**)

The epistemic importance of S4.2. In the late '70s, W. Lenzen proved that assuming $\mathbf{S4}_{\mathsf{K}}$ for knowledge, $\mathbf{KD45}_{\mathsf{B}}$ for belief and some plausible axioms for their interplay, we arrive at a logic practically equivalent to **S4.2**, assuming that belief there is captured by a derived modal operator introduced by the axiom **DB**. $\mathsf{B}\varphi \equiv \neg\mathsf{K}\neg\mathsf{K}\varphi$ (which captures belief through knowledge); we call this version $\mathbf{S4.2}_{\mathsf{KB}}$. Similar results can be found in R. Stalnaker's work.

[2] According to W. Lenzen, a 'realist epistemologist' should, at least, accept this principle [Len79, p. 43].

Proposition 2.1

$$\mathbf{S4} + \mathbf{KD45}_{\mathsf{B}} + \mathbf{B1} + \mathbf{B2.3} + \mathbf{B2.4} = \mathbf{S4.2}_{\mathsf{KB}} \qquad \text{[Len79, Lenzen]}$$

$$\mathbf{S4} + \mathbf{CB} + \mathbf{KB} + \mathbf{SB} + \mathbf{PIB} + \mathbf{NIB} = \mathbf{S4.2}_{\mathsf{KB}} \qquad \text{[Sta06, Stalnaker]}$$

Ending this digression, we wish to remind the reader that our language is monomodal in this paper, we reserve $\Box\varphi$ for knowledge and we keep Lenzen's shorthhand for belief as $\neg\Box\neg\Box\varphi$. Some of the epistemic principles mentioned above, will be used below in our results.

2.4 Cluster analysis of transitive logics

The **cluster analysis** of **transitive logics** is well known [Gol92, Chap.8][Seg71]. We provide the necessary definitions and results below, with a bit of personal flavour in terminology.

Some useful facts. We will restrict ourselves to possible-worlds frames with a reflexive, transitive and directed relation (henceforth called **rtd-relation**), keeping in mind that in the class of reflexive and transitive frames, *directedness* is equivalent to *weak directedness*[3] [Gol92, p. 30]. The following definition for these relations, captures the notion of cluster, as a maximal subset of states, inside which the (restriction of the) accessibility relation is universal. Following this definition, we gather some properties of clusters inside rtd-relations.

Definition 2.2 *Let $R \subseteq W \times W$ be any (binary) rtd-relation on W, and $\varnothing \neq C \subseteq W$.*

(i) *C is called a* **cluster** *of R iff*
$(\forall s,t \in C)sRt$ *and* $(\forall u \in W \setminus C)(\exists v \in C)(\neg uRv$ *or* $\neg vRu)$

(ii) *The cluster C of R is called* final *iff* $(\forall u \in W \setminus C)(\forall v \in C)(uRv \ \& \ \neg vRu)$

Fact 2.3

(i) $(\forall s \in W)(\exists C : \textit{cluster})\ s \in C$

(ii) $(\forall\ \textit{clusters}\ C, C' \subseteq W)\ C \cap C' = \varnothing$

(iii) $(\forall\ \textit{clusters}\ C, C' \subseteq W)(\forall s \in C, s' \in C')(sRs' \Longrightarrow (\forall t \in C, t' \in C')\ tRt')$

(iv) $(\forall\ \textit{clusters}\ C, C' \subseteq W)(\forall s \in C, s' \in C')$
$\big((C \neq C' \ \&\ sRs') \Longrightarrow (\forall t \in C, t' \in C')\ \neg t'Rt\big)$

(v) *If a final cluster exists, it is unique. There is always a final cluster in finite models.*

[3] i.e. $(\forall w,v,u \in W)\big((wRv \ \&\ wRu) \Rightarrow (\exists t \in W)(vRt \ \&\ uRt)\big)$

It is customary to order clusters too, and we employ the following definition to make this concrete. As we will prove, there is no loss of generality in '*collapsing*' the clusters by defining a relation on the clusters' *indices* and we will work for simplicity with frames possessing a finite number of clusters (the indices will be members of $D = \{0, \ldots, n\}$). The lemma following the definition makes clear that the relation constructed inherits properties from its '*generating*' relation R.

Definition 2.4 *Let R be an rtd-relation on W. Then, a* pattern-relation $R_p \subseteq D \times D$ *of R is any relation on D s.t. $(\forall i, j \in D)$*

$$iR_pj \iff (\exists s \in C_i, t \in C_j)\ sRt$$

where $C_0, \ldots, C_n \subseteq W$ is an enumeration of the clusters of R.

Lemma 2.5 *Let R be an rtd-relation on W and R_p a pattern-relation of R (for clusters $C_0, \ldots, C_n \subseteq W$). Then,*

(i) $(\forall i, j \in D)\big(iR_pj \iff (\forall s \in C_i, t \in C_j)\ sRt\big)$

(ii) *R_p is also an rtd-relation.*

(iii) *All clusters of R_p are singletons.*

Proof. (i) It is immediate, by Def.2.4 and Fact 2.3(ii).

(ii) Follows easily by Def.2.4 and (i), since R is rtd.

(iii) First of all, in light of (ii), it is meaningful to refer to R_p-clusters, which contain R-clusters. Suppose, for the sake of contradiction, that there is an R_p-cluster with more than one elements. Let i, j be two of them. Since they belong to an R_p-cluster, by Def.2.2(i), iR_pj and jR_pi, hence, by (i), $(\forall s \in C_i, t \in C_j)\ (sRt\ \&\ tRs)$ (1)
But, by Fact 2.3(ii) (and since $C_j \neq \varnothing$), there exists a $u \in C_j \setminus C_i$, hence, by Def.2.2(i), there is a $v \in C_i$ s.t. $\neg uRv$ or $\neg vRu$, which contradicts to (1). And since clusters are by definition non-empty, they are singletons. ∎

The property (iii) entails another one, which will be useful below, so we will focus on rtd-relations endowed with (iii). These relations deserve a name.

Definition 2.6 *Every binary relation which is reflexive, transitive, directed and has only singleton clusters (i.e. every cluster consists of only one reflexive element) is called a* **simple rtd-relation** (**s-rtd**).

Lemma 2.7 *Let R be an s-rtd-relation on W. Then, there is an $f \in W$ s.t.*

$$(Gd) \qquad (\forall i \in W)\ \big(iRf\ \&\ (i \neq f \Rightarrow \neg fRi)\big)$$

PROOF. Let $F \subseteq W$ be the final cluster, guaranteed by Fact 2.3(v), and $i \in W$. Since every cluster is a singleton, let $F = \{f\}$. If $i = f$, then, since R is reflexive, iRf. If $i \neq f$, then, $i \in W \setminus F$, hence, by Fact 2.3(ii), iRf and $\neg fRi$. ∎

The *following* fact is now obvious.

Corollary 2.8 *Let R be an rtd-relation on W and R_p a pattern relation of R (for clusters $C_0, \ldots, C_n \subseteq W$). Then, R_p is an s-rtd-relation and satisfies (Gd), where $W = D$.*

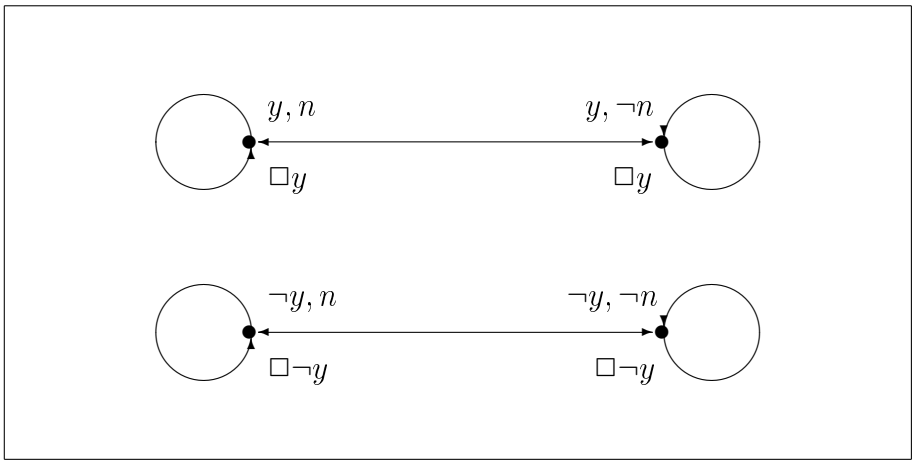
Let us mention here that it can be proved that **S4.2** is also determined by the subclass of its frames (finite or infinite) possessing a *final cluster*; the result is implicit in [KT07] and it can be also found at [KZ15].

3 KB_R-structures

3.1 Motivation

We will work on an example of M. Fitting from [Fit93]; similar examples can be found in various places at [vDvdHK07]. Assume that we are interested in the representation of knowledge (and ignorance) of an agent about the current raining conditions in New York and in Novosibirsk. We need two propositional variables (e.g. y, n), which represent "*it's raining in New York*" and "*it's raining in Novosibirsk*" respectively. Then, there are four different situations, i.e. combinations of truth values of y and n. Suppose that our agent resides in New York and knows whether it is raining there or not. Assume that it is raining there. Then, she would consider both situations y, n and $y, \neg n$ as two alternatives of the true state of the world. If she were in y, n, she could not be able to distinguish her situation from $y, \neg n$, and vice versa. And of course, y, n itself is indistinguishable from y, n. In this case, since y is true in every alternative situation for our agent (i.e. she knows y), $\Box y$ is true in y, n and in $y, \neg n$. Analogously, if it isn't raining in New York, situations $\neg y, n$ and $\neg y, \neg n$ are not distinguishable to each other by our agent, and $\Box \neg y$ is true in both of them. This epistemic model could be represented as in Figure 3.1 on page 8, provided that arrows connect indistinguishable situations.

This is the standard approach of an epistemic model, and the model is considered to be *symmetric*, i.e. that all arrows are bidirectional. This means in our model, for instance, that our agent's ignorance about the weather conditions in Novosibirsk is independent of what is really happening there, in other words, she knowns exactly the same facts independently of where she is located, within the indistinguishable part of the model. Generally, the standard approach assumes that all indistinguishable to eachother situations, 'see' one another, which is a result of the assumption that the information given to the agent is *the same* in all indistinguishable situations.



We claim that this assumption is not always true. For example, in the previous model one might think of a special case, where the heavy rain in Novosibirsk was reported in the news headlines, and our agent became aware of that. Then, assuming it isn't raining in New York, from $\neg y, n$, situation $\neg y, \neg n$ is distinguishable, whereas from $\neg y, \neg n$, situation $\neg y, n$ isn't, since we assume that in this case no comment about the (good) weather in Novosibirsk was made by the news agencies, and our agent doesn't know what's happening there.

Except of some special cases, which can not be covered by the assumption of uniform distribution of information within indistinguishable situations, there is another drawback of the standard assumption and its entailment that all indistinguishable situations 'see' each other. Suppose that an agent, being in a situation (let us name it) i, does not know φ. Then, there must be an indistinguishable from i situation j, where $\neg\varphi$ holds. Since every other indistinguishable from i situation k sees j, it will also in k be true that our agent does not know φ (a witness for that is j). Hence, our agent does know in i that she doesn't know φ. So, in every situation, it holds that if the agent *does not know something*, then *she is aware of her ignorance about that*. And this very fact, which is known as *negative introspection*, is not acceptable by the vast majority of philosophers.

In our approach, trying to find a remedy for these drawbacks, we will **assume** that **information is not uniformly distributed** all over the situations. We intend to establish a formal representation of knowledge sets, which will not necessarily be the same globally, but different for each situation (in fact, we will describe the properties of those sets, not necessarily for each situation, but for 'blocks' of indistinguishable situations). So, assuming that there are n different situations, we denote for any situation $i \in \{0, \ldots, n\}$ the agent's knowledge set as r_i. **To be able to define those sets, we have to consider sets T_i, which will contain all true formulas in situation** i. Our agent does not necessarily know every formula in T_i; and anything believed by her, might not be true. Furthermore, being in a situation i the agent might distinguish between her current situation and another, because she has some information, which allows her to do so. But she might also not be able not distinguish between her current

situation i and another j. As explained previously, if j is an alternative situation for i, then it is not necessarily true that i is an alternative situation for j, since being in j, our agent might be provided with extra information, which might allow her to distinguish between j and i.

Note also, that - in the general case - **the agent does not know in which situation she is located**. If we know that the agent is in situation i and that, say, j, k and l are alternative situations for i (i.e. indistinguishable from i), she might not know that she is in i. She rather knows that i, j, k and l are all indistinguishable situations. Speaking about indistinguishable situations *from* i means that *we do know* that if our agent were in i, she would consider these situations together with i as alternative variations of her present situation, which is unknown to her.

Furthermore, we could know – since **we enjoy the "eagle's view"** – that if our agent were in situation j, she would have the information to distinguish between her situation and, say, k, but this is something that she does not know. Only if she actually were in j she would know that. Now, assume that in the previous example our agent is aware of the fact that in all alternative situations (included the unknown to her, current situation i) a formula φ is true (i.e. $\varphi \in T_i \cap T_j \cap T_k \cap T_l$). Then, it is natural to say that she is sure about φ, that she *knows* φ. Therefore, given a relation $R \subseteq \{0, \ldots, n\}$, representing all couples of indistinguishable situations (i.e. iRj means that j is an alternative situation for i), we will define in the next section, r_i as $\bigcap_{iRj} T_j$.

As mentioned previously, in our modal language $\mathcal{L}_\Box$ **the modality denotes knowledge**. Hence, we have two ways of denoting knowledge of φ: using formula $\Box\varphi$, and saying that $\varphi \in \mathsf{r}_i$. To be consistent with our intuitions, we have to demand that

$$\text{if } \varphi \in \mathsf{r}_i, \text{ then } \Box\varphi \in T_i \tag{3.0.i}$$

(i.e. if our agent knows φ, then, obviously, it is true that she does know it!), and

$$\text{if } \varphi \notin \mathsf{r}_i, \text{ then } \neg\Box\varphi \in T_i \tag{3.0.ii}$$

One might wonder why don't we simply demand $\varphi \in \mathsf{r}_i$ iff $\Box\varphi \in T_i$. Then, $\varphi \notin \mathsf{r}_i$ would simply entail $\Box\varphi \notin T_i$, which seams to be natural, since "*it is not true that I know* φ" looks equivalent to "*it is true that I do not know* φ"! This equivalence is obviously true, if we see each situation i as a unique state of affairs, as we did hitherto. But in a more general case, we could consider bunches of situations (possibly, infinite many situations in a bunch), where all situations of the same bunch are indistinguishable to each other, i.e. for every situation s of a bunch, any other of the same bunch, is an alternative one for s. From now on we will call those bunches, *clusters* and we will denote them as i, j, k etc. The situations themselves will be denoted as s, t, u etc.

We intend to define those clusters in a such way, that if some situation s of a cluster i considers situation t of any other cluster as an alternative one, then every other situation of i will consider t as an alternative one. And if we say that φ is true in cluster i, obviously, we would like to mean that φ is true in every situation of i, i.e. that T_i contains all formulas valid in i. Hence, $\Box\varphi \notin T_i$ does not necessarily entail

that $\neg\Box\varphi \in T_i$. But the inverse is true. That's why we chose the stronger property: $\varphi \notin \Gamma_i \Rightarrow \neg\Box\varphi \in T_i$. Note also that now, R does not anymore relate situations, rather than clusters, in the sense that iRj means that our agent, being in any situation s of i considers as indistinguishable from s any situation of j.

We will also adopt the option of **defining belief through knowledge**. To do so, we will follow the idea introduced by W. Lenzen [Len79], who argued that the following definition of belief is acceptable even by the '*most scrupulous epistemologist*': **an agent believes φ iff she does not know that she doesn't know** φ (i.e. $\neg\Box\neg\Box\varphi$ defines '*believing in* φ'). Now, our agent knows that she doesn't know φ iff $\varphi \notin \Gamma_j$ for every alternative situation j for i, hence, she would believe φ iff $\varphi \in \Gamma_j$ for some alternative situation j for i. Therefore – assuming that the belief sets, containing everything believed by our agent in any situation of i, will be denoted as Δ_i – it is consistent with Lenzen's definition to identify Δ_i as $\bigcup_{iRj} \Gamma_j$. As noted above, there exists a direct way to speak about "*believing*" φ: $\neg\Box\neg\Box\varphi$. So, to be consistent with our intuitions, we have to define the theories T_i and Δ_i in such a way, that they will satisfy the following conditions:

$$\text{if } \varphi \in \Delta_i, \text{ then } \neg\Box\neg\Box\varphi \in T_i \tag{3.0.iii}$$

(i.e. if our agent believes in φ, then, it is true that she does not know that she doesn't know it), and

$$\text{if } \varphi \notin \Delta_i, \text{ then } \Box\neg\Box\varphi \in T_i \tag{3.0.iv}$$

Let us now sum up, everything we have discussed so far. We began, considering an agent, who might be in some situation, and who accepts as possible from there, all other situations, which she can not distinguish. We presumed that there is a relation R connecting those situations, in the sense that, sRt iff situation t is indistinguishable from s. We decided that R should be reflexive, transitive and directed (rtd), and we saw that in that case, there are clusters of situations, which (situations) are indistinguishable from eachother. Then, by defining the pattern relation R_p of R, which connects all clusters of R, we proved that it is an s-rtd-relation, i.e. rtd and, additionally, it has only singleton-clusters, which entails that there is one "final" element (property (Gd)). So, henceforth, we will focus on this pattern relation, which links clusters to eachother, and everytime we mention R, we refer to the pattern relation, which is s-rtd.

We also declared that we want T_i to be the set of all valid formulas in all situations of cluster i. We found out that $\Gamma_i = \bigcap_{iRj} T_j$ should be the set of all formulas known in cluster i, and $\Delta_i = \bigcup_{iRj} \Gamma_j$ the set of all formulas believed there. We were interested only to clusters, rather that to single situations, since in all situations of a cluster, exactly the same formulas are known. This is immediate, since to '*know in a situation* s' means '*true in all indistinguishable situations from* s' and every situation in a cluster considers as indistinguishable exactly the same set of situations.

Hence, assuming that we were given an s-rtd-relation between clusters of situations, we should describe all requirement the T_i's should meet, so that the intuitive properties (3.0.i)–(3.0.iv) about knowledge and belief are true. This is exactly what we are going to do in next section.

3.2 Definition of KB_R-structures

Let us have in mind that $D = \{0, \ldots, n\}$ contains the (indices of the) clusters of the epistemic situations considered, and that $T_0, \ldots, T_n$ are the corresponding theories, containing exactly all formulas, valid there. Firstly, we describe all those properties, which these theories should satisfy, and privide the overall structure a name.

Definition 3.1 *Let $R \subseteq D \times D$ be an s-rtd-relation on D and $T_0, \ldots, T_n \subseteq \mathcal{L}_\Box$ be consistent theories s.t. $(\forall i \in D)$*

(PC_i) $\quad \mathbf{PC}_{\mathcal{L}_\Box} \subseteq T_i$ *and* T_i *is closed under* **MP**

(P_i) $\quad (\forall \varphi \in \mathcal{L}_\Box)(\varphi \in \bigcap_{iRj} T_j \Rightarrow \Box\varphi \in T_i)$

(N_i) $\quad (\forall \varphi \in \mathcal{L}_\Box)(\varphi \notin T_i \Rightarrow \neg\Box\varphi \in \bigcap_{jRi} T_j)$

Furthermore, for any $i \in D$, we define Γ_i and Δ_i as

$$\Gamma_i = \bigcap_{iRj} T_j \quad \text{and} \quad \Delta_i = \bigcup_{iRj} \Gamma_j$$

Then, the ordered triple $\langle (T_i), (\Gamma_i), (\Delta_i) \rangle^R_{i \in D}$ is called a **$\mathbf{KB_R}$-structure**. *In fact, it is a triple consisting of n-tuples of theories.*

The following simple example demonstrates that Stalnaker stable sets correspond to a trivial case of our setting, i.e. one that originates from a simple cluster.

Example 3.2 Consider $D = \{0\}$, a consistent theory $T_0 \subseteq \mathcal{L}_\Box$, and the corresponding $KB_{\{(0,0)\}}$-structure $\langle T_0, \Gamma_0, \Delta_0 \rangle^{\{(0,0)\}}$ (for the trivial s-rtd-relation over D, $\{(0,0)\}$). Then, by Def. 3.1, T_0 satisfies: $(\forall \varphi \in \mathcal{L}_\Box)$

(PC_0) $\quad \mathbf{PC}_{\mathcal{L}_\Box} \subseteq T_0$ and T_0 is closed under **MP**

(P_0) $\quad \varphi \in T_0 \Rightarrow \Box\varphi \in T_0$

(N_0) $\quad \varphi \notin T_0 \Rightarrow \neg\Box\varphi \in T_0$

T_0 is a stable set according to Stalnaker's definition. Furthermore, $\Gamma_0 = \Delta_0 = T_0$.

Example 3.3 Let us consider now the s-rtd-relation $R = \{(0,0), (1,1), (1,0)\}$ and the corresponding KB_R-structure $\langle (T_i), (\Gamma_i), (\Delta_i) \rangle^R_{i \in D}$. Then, Def.3.1 says that T_0 and T_1 are meant to be consistent and to satisfy all conditions listed below: $(\forall \varphi \in \mathcal{L}_\Box)$

$(PC_{0,1})$ $\quad \mathbf{PC}_{\mathcal{L}_\Box} \subseteq T_0, T_1$ and T_0, T_1 are closed under **MP**

(P_0) $\quad \varphi \in T_0 \Rightarrow \Box\varphi \in T_0$

(N_0) $\quad \varphi \notin T_0 \Rightarrow \neg\Box\varphi \in T_0 \;\&\; \neg\Box\varphi \in T_1$

(P_1) $\quad \varphi \in T_0 \;\&\; \varphi \in T_1 \Rightarrow \Box\varphi \in T_1$

(N_1) $\quad \varphi \notin T_1 \Rightarrow \neg\Box\varphi \in T_1$

Furthermore, $\mathsf{r}_0 = T_0$, $\mathsf{r}_1 = T_0 \cap T_1$, $\vartriangle_0 = T_0$ and $\vartriangle_1 = T_0 \cup (T_0 \cap T_1) = T_0$. The fact that $\vartriangle_0 = \vartriangle_1 = T_0$ is not a coincidence, but a result of some properties, which are satisfied by R, and which will be proved below (Fact 3.14).

The next Fact shows that everything in Definition 3.1 is consistent with what we said in section 3.1.

Fact 3.4 $(\forall i \in D)\big((P_i) \iff (3.0.i) \;\&\; (N_i) \iff (3.0.ii)\big)$
Definition 3.1 entails properties (3.0.iii), (3.0.iv) and $(\forall i \in D)(\forall \varphi \in \mathcal{L}_\Box)$

$$\varphi \in \mathsf{r}_i \iff \Box\varphi \in T_i \quad \textit{and} \quad \varphi \in \vartriangle_i \iff \neg\Box\neg\Box\varphi \in T_i \tag{3.4.v}$$

Proof. The equivalence of (P_i) and (3.0.i) is immediate, by definition of r_i. Next, assume that $(\forall i \in D)(N_i)$ holds and let $\varphi \notin \mathsf{r}_i$. Then, by definition of r_i, there is a $j \in D$ s.t. iRj and $\varphi \notin T_j$, and by (N_j), $\neg\Box\varphi \in T_i$.

Conversely, assume that $(\forall i \in D)$ (3.0.ii) holds and let $\varphi \notin T_i$ and $j \in D$ s.t. jRi. Suppose for the sake of contradiction, that $\varphi \in \mathsf{r}_j$. Then, by definition of r_j, $\varphi \in \bigcap_{jRk} T_k$, and since jRi, $\varphi \in T_i$, which is a contradiction. Hence, $\varphi \notin \mathsf{r}_j$, so, by (3.0.ii), $\neg\Box\varphi \in T_j$. Therefore, $\neg\Box\varphi \in \bigcap_{jRi} T_j$, and (N_i) is true.

For (3.0.iii), assume that $\varphi \in \vartriangle_i$. Then, by definition of $\vartriangle_i$, there is an $i \in D$ s.t. iRj and $\varphi \in \mathsf{r}_j$. Hence, by (3.0.i), $\Box\varphi \in T_j$, and since T_j is consistent, $\neg\Box\varphi \notin T_j$, so, $\neg\Box\varphi \notin \bigcap_{iRj} T_j$, therefore, by definition of r_i, $\neg\Box\varphi \notin \mathsf{r}_i$, and by (3.0.ii), $\neg\Box\neg\Box\varphi \in T_i$.

For (3.0.iv), let $\varphi \notin \vartriangle_i$. Then, by definition of $\vartriangle_i$, for all $j \in D$ s.t. iRj, $\varphi \notin \mathsf{r}_j$, hence, by (3.0.ii), $(\forall j \in D)(iRj \Rightarrow \neg\Box\varphi \in T_j)$, so, $\neg\Box\varphi \in \bigcap_{iRj} T_j$, i.e., by definition of r_i, $\neg\Box\varphi \in \mathsf{r}_i$, and finally, by (3.0.i), $\Box\neg\Box\varphi \in T_i$.

For (3.4.v), if $\varphi \notin \mathsf{r}_i$, then, by (3.0.ii), $\neg\Box\varphi \in T_i$, hence, since every T_i is consistent, $\Box\varphi \notin T_i$. Therefore, using also (3.0.i), $\varphi \in \mathsf{r}_i \Leftrightarrow \Box\varphi \in T_i$. Furthermore, if $\varphi \notin \vartriangle_i$, then, by (3.0.iv), $\Box\neg\Box\varphi \in T_i$, hence, since T_i is consistent, $\neg\Box\neg\Box\varphi \notin T_i$. So, by (3.0.iii), $\varphi \in \vartriangle_i \Leftrightarrow \neg\Box\neg\Box\varphi \in T_i$. ∎

3.3 Epistemic properties of $\mathbf{KB_R}$-structures

Even without any restrictions to R, Def. 3.1 would endow all theories appearing there with axiom **K**, as the first lemma verifies. Further on in our discussion in the motivation section, it would be desirable that the properties of R would lead to the incorporation of some intuitively acceptable properties of knowledge and belief in r_i and $\vartriangle_i$. The following

lemmata state that **reflexivity** leads to two desirable properties: the *entailment thesis* (**knowledge implies belief**) and the property requiring that **knowledge implies certainty**.

Lemma 3.5 *Let* $\langle (T_i), (\mathsf{r}_i), (\vartriangle_i) \rangle^R_{i \in D}$ *be any* KB_R*-structure. Then,*

$$(\forall i \in D)(\forall \varphi, \psi \in \mathcal{L}_\Box)\ \mathbf{K} \in T_i$$

Proof. If $\neg\Box\varphi \in T_i$ or $\neg\Box(\varphi \supset \psi) \in T_i$, then, by (PC_i), $\mathbf{K} \in T_i$.
If $\neg\Box\varphi \notin T_i$ and $\neg\Box(\varphi \supset \psi) \notin T_i$, then, by (3.0.ii), $\varphi \in \mathsf{r}_i$ and $\varphi \supset \psi \in \mathsf{r}_i$, hence, by definition of r_i, $\varphi \in \bigcap_{iRj} T_j$ and $\varphi \supset \psi \in \bigcap_{iRj} T_j$, so, by (PC_j), $\psi \in \bigcap_{iRj} T_j$, therefore, by (P_i), $\Box\psi \in T_i$, and finally, by (PC_i), again $\mathbf{K} \in T_i$. ∎

Lemma 3.6 *Let* $\langle (T_i), (\mathsf{r}_i), (\vartriangle_i) \rangle^R_{i \in D}$ *be any* KB_R*-structure. Then,* $(\forall i \in D)\ \mathsf{r}_i \subseteq T_i \cap \vartriangle_i$ *(i.e. everything our agent knows is true, and she believes it).*

Proof. Assume $i \in D$ and $\varphi \in \mathsf{r}_i$. By definition of r_i, $\varphi \in \bigcap_{iRj} T_j$, and since iRi, $\varphi \in T_i$. Furthermore, since iRi, $\varphi \in \bigcup_{iRj} \mathsf{r}_j = \vartriangle_i$. ∎

Lemma 3.7 *Let* $\langle (T_i), (\mathsf{r}_i), (\vartriangle_i) \rangle^R_{i \in D}$ *be any* KB_R*-structure. Then,*

$$(\forall i \in D)(\forall \varphi \in \mathcal{L}_\Box)\ \mathbf{T} \in T_i$$

Proof. If $\neg\Box\varphi \in T_i$, then, by (PC_i), $\mathbf{T} \in T_i$.
If $\neg\Box\varphi \notin T_i$, then, by (3.0.ii), $\varphi \in \mathsf{r}_i$, i.e. by definition of r_i, $\varphi \in \bigcap_{iRj} T_j$, and since R is reflexive, iRi, so, $\varphi \in T_i$, hence, by (PC_i), again $\mathbf{T} \in T_i$. ∎

Not really surprisingly, **transitivity** entails **positive introspection concerning knowledge**, as a context rule. Next, Lemma 3.8 along with the definition of $\vartriangle_i$ entail Lemma 3.9.

Lemma 3.8 *Let* $\langle (T_i), (\mathsf{r}_i), (\vartriangle_i) \rangle^R_{i \in D}$ *be any* KB_R*-structure. Then,* $(\forall i \in D)$

$$(PI_i) \qquad (\forall \varphi \in \mathcal{L}_\Box)(\varphi \in \mathsf{r}_i \Rightarrow \Box\varphi \in \mathsf{r}_i)$$

Proof. Suppose that $\varphi \in \mathsf{r}_i$, i.e. by definition of r_i, $\varphi \in \bigcap_{iRj} T_j$. Then,

$$(\forall j \in D)(iRj \Rightarrow \varphi \in T_j) \quad (*)$$

Let now $k \in D$ s.t. iRk, and $l \in D$ s.t. kRl. Since R is transitive, iRl, so, by $(*)$, $\varphi \in T_l$. Hence, $\varphi \in \bigcap_{kRl} T_l$, subsequently, by (P_k), $\Box\varphi \in T_k$. Therefore, $\Box\varphi \in \bigcap_{iRk} T_k$, i.e. by definition of r_i, $\Box\varphi \in \mathsf{r}_i$. ∎

Lemma 3.9 *Let* $\langle (T_i), (\mathsf{r}_i), (\vartriangle_i) \rangle^R_{i \in D}$ *be any* KB_R*-structure. Then,*

$$(\forall i \in D)(\forall \varphi \in \mathcal{L}_\Box)(\varphi \in \vartriangle_i \Rightarrow \Box\varphi \in \vartriangle_i)$$

Note that Lemma 3.9 in light of (3.4.v) (see section 3.1) shows that *if our agent believes something, then she believes that she knows it* (which is similar to Lenzen's property (B2.1) [Len79]). Transitivity of R is embedded in every theory of Def. 3.1 through axiom **4**. Finally, Lemma 3.11 is technically useful in the next section.

Lemma 3.10 *Let* $\langle (T_i), (\mathsf{r}_i), (\vartriangle_i)\rangle^R_{i\in D}$ *be any* KB_R*-structure. Then,*

$$(\forall i \in D)(\forall \varphi \in \mathcal{L}_\Box)\ \mathbf{4} \in T_i$$

Proof. If $\neg\Box\varphi \in T_i$, then, by (PC_i), $\mathbf{4} \in T_i$.
If $\neg\Box\varphi \notin T_i$, then, by (3.0.ii), $\varphi \in \mathsf{r}_i$, and by Lemma 3.8, $\Box\varphi \in \mathsf{r}_i$, hence, by (3.0.i), $\Box\Box\varphi \in T_i$, so, by (PC_i), again $\mathbf{4} \in T_i$. ∎

Lemma 3.11 *Let* $\langle (T_i), (\mathsf{r}_i), (\vartriangle_i)\rangle^R_{i\in D}$ *be any* KB_R*-structure. Then,* $(\forall i, j \in D)$

$$iRj \Rightarrow \mathsf{r}_i \subseteq \mathsf{r}_j$$

Proof. Let $\varphi \in \mathsf{r}_i$ and $k \in D$ s.t. jRk. Then, since R is transitive, iRk, and by definition of r_i, $\varphi \in T_k$, hence, by definition of r_j, $\varphi \in \mathsf{r}_j$. ∎

Finally, **directedness** of R leads to properties, similar to Lenzen's (B2.3) and (B2.4) [Len79, p.43-44]. The *former* one, which should be acceptable by a "*realistic epistemologist*", says that **if an agent believes something, then she can not believe that she doesn't know it**. The *latter* property, which should be acceptable – according to Lenzen – by a "*simplifier*", states that **if an agent believes something, then she knows that she believes it**.

Lemma 3.12 *Let* $\langle (T_i), (\mathsf{r}_i), (\vartriangle_i)\rangle^R_{i\in D}$ *be any* KB_R*-structure. Then,*
$(\forall i \in D)(\forall \varphi \in \mathcal{L}_\Box)$

$$(B2.3)\quad \varphi \in \vartriangle_i \Rightarrow \neg\Box\varphi \notin \vartriangle_i \qquad \textit{and} \qquad (B2.4)\quad \varphi \in \vartriangle_i \Rightarrow \neg\Box\neg\Box\varphi \in \mathsf{r}_i$$

Proof. Since both implications have the same premise, we start proving both of them, assuming that $\varphi \in \vartriangle_i$. Then, by definition of $\vartriangle_i$, there is a $j \in D$ s.t. iRj and $\varphi \in \mathsf{r}_j$. Let now $l \in D$ s.t. iRl. Then, since R is weakly directed [4], there must be an $m \in D$ s.t. jRm and lRm. Furthermore, assume that $s \in D$ be s.t. mRs. Since jRm and since R is transitive, jRs, hence, since $\varphi \in \mathsf{r}_j = \bigcap_{jRk} T_k$, $\varphi \in T_s$. So, $\varphi \in \bigcap_{mRs} T_s$, and by (P_m), $\Box\varphi \in T_m$, and since T_m is consistent, $\neg\Box\varphi \notin T_m$.

For (B2.3). It has been proved so far, that there is an $m \in D$ s.t. lRm and $\neg\Box\varphi \notin T_m$. Consequently, by definition of r_l, $\neg\Box\varphi \notin \mathsf{r}_l$, hence, $(\forall l \in D)(iRl \Rightarrow \neg\Box\varphi \notin \mathsf{r}_l)$, so, by definition of $\vartriangle_i$, $\neg\Box\varphi \notin \vartriangle_i$.

For (B2.4). Since $\neg\Box\varphi \notin T_m$ and since lRm, by (N_m), $\neg\Box\neg\Box\varphi \in T_l$, hence, $\neg\Box\neg\Box\varphi \in \bigcap_{iRl} T_l$, and by definition of r_i, $\neg\Box\neg\Box\varphi \in \mathsf{r}_i$. ∎

[4] As we have said in section 2, since R is reflexive and transitive, being directed is equivalent to being weakly directed.

Now, let us focus on the last presumption for R: being a **simple** rtd-relation. Then, by Lemma 2.7, property (Gd) is true for R. Without loss of generality, we will tacitly assume that the '*final*' element of R is 0, i.e. that (Gd) appears in the following form:

$$(Gd) \qquad (\forall i \in D)\ (iR0\ \&\ (i > 0 \Rightarrow \neg 0Ri))$$

This property endows every theory of Def. 3.1 with axiom **G** and leads to next two results.

Lemma 3.13 *Let $\langle (T_i), (\mathsf{r}_i), (\vartriangle_i)\rangle^R_{i\in D}$ be any KB_R-structure. Then,*

$$(\forall i \in D)(\forall \varphi \in \mathcal{L}_\Box)\ \mathbf{G} \in T_i$$

Proof. If $\Box\neg\Box\varphi \in T_i$, then, by (PC_i), $\mathbf{G} \in T_i$.

If $\Box\neg\Box\varphi \notin T_i$, then, by (P_i), there is a $j \in D$ s.t. iRj and $\neg\Box\varphi \notin T_j$, hence, since (by (Gd)) $jR0$, by (N_0), $\varphi \in T_0$, therefore, because T_0 is consistent, $\neg\varphi \notin T_0$, and by (N_0), $\neg\Box\neg\varphi \in \bigcap_{jR0} T_j$, so, by (Gd), $(\forall j \in D)\neg\Box\neg\varphi \in T_j$, hence of course, $\neg\Box\neg\varphi \in \bigcap_{iRj} T_j$, and by (P_i), $\Box\neg\Box\neg\varphi \in T_i$, consequently, by (PC_i), again $\mathbf{G} \in T_i$. ∎

Fact 3.14 *Let $\langle (T_i), (\mathsf{r}_i), (\vartriangle_i)\rangle^R_{i\in D}$ be any KB_R-structure. Then, $(\forall i \in D)$*

(i) $\vartriangle_i = \mathsf{r}_0 = T_0$

(ii) $\vartriangle_i$ *is a stable theory according to Stalnaker's definition*

Proof. (i) Suppose that $\varphi \in \vartriangle_i$. Then, by definition of $\vartriangle_i$, there is a $j \in D$ s.t. iRj and $\varphi \in \mathsf{r}_j$, hence, by definition of r_j, $\varphi \in \bigcap_{jRk} T_k$, and since – by (Gd) – $jR0$, $\varphi \in T_0$.

Conversely, assume that $\varphi \in T_0$. But, by definition of r_0 and (Gd), $\mathsf{r}_0 = \bigcap_{0Rj} T_j = T_0$, hence, $\varphi \in \mathsf{r}_0$. But, again by (Gd), $iR0$, hence, $\varphi \in \bigcup_{iRj} \mathsf{r}_j$, so, by definition of $\vartriangle_i$, $\varphi \in \vartriangle_i$.

It has been proved that $\vartriangle_i = T_0$, but also that $\mathsf{r}_0 = T_0$.

(ii) (PC_0), (P_0) and (N_0) guarantee that T_0 is Stalnaker stable. Then, so is every $\vartriangle_i$, by (i). ∎

Now, it is immediate that our belief sets follow the **principle of consistency of belief**, i.e. that **if an agent believes φ, she can not believe $\neg\varphi$**.

Lemma 3.15 *Let $\langle (T_i), (\mathsf{r}_i), (\vartriangle_i)\rangle^R_{i\in D}$ be any KB_R-structure. Then,*
$(\forall i \in D)(\forall \varphi \in \mathcal{L}_\Box)$

$$\varphi \in \vartriangle_i \Rightarrow \neg\varphi \notin \vartriangle_i$$

Proof. If $\varphi \in \vartriangle_i$, then, by Fact 3.14, $\varphi \in T_0$, hence, by consistency of T_0, $\neg\varphi \notin T_0$, and again by Fact 3.14, $\neg\varphi \notin \vartriangle_i$. ∎

All the previous lemmata seem to justify the choice of the KB_R notion in Def. 3.1: KB_R-structures contain **K**(nowledge) theories (the Γ_i's), and **B**(elief) theories (the Δ_i's). According to Fact 3.14, one of the Γ_i's coincides with everything believed in any situation. Without loss of generality, it is assumed that this one is Γ_0. In the following section we will present a model-theoretic characterization of KB_R-structures. To do so, we need the next important result, in which we employ a notion of **strong provability**.

Lemma 3.16 *If $\langle (T_i), (\Gamma_i), (\Delta_i) \rangle^R_{i \in D}$ is any KB_R-structure, then, $(\forall i \in D)$*

(i) Γ_i *is closed under strong* **S4.2** *provability, i.e.* $\Gamma_i = \{\varphi \in \mathcal{L}_\Box \mid \Gamma_i \vdash_{\mathbf{S4.2}} \varphi\}$.

(ii) Γ_i *is a consistent with* **S4.2** *theory (c***S4.2***-theory).*

Proof. (i) It is obvious that, if $\varphi \in \Gamma_i$, then $\Gamma_i \vdash_{\mathbf{S4.2}} \varphi$. Conversely, suppose that $\Gamma_i \vdash_{\mathbf{S4.2}} \varphi$. It will be proved, by induction on the length of $\Gamma_i \vdash_{\mathbf{S4.2}} \varphi$, that $\varphi \in \Gamma_i$.

Ind.Basis. For length of proof equal to 1. Let $j \in D$ be s.t. iRj.
If $\varphi \in \mathbf{PC}_{\mathcal{L}_\Box}$, then, by (PC_j), $\varphi \in T_j$. If φ is an instance of **K** or **T** or **4** or **G**, then, by Lemmata 3.5, 3.7, 3.10 and 3.13 respectively, $\varphi \in T_j$. Hence, in any case $\varphi \in T_j$, and since iRj, by definition of Γ_i, $\varphi \in \Gamma_i$.

Ind.Step. If ψ and $\psi \supset \varphi$ are formulas of the proof in previous steps, then, by Ind.Hypothesis, $\psi \in \Gamma_i$ and $\psi \supset \varphi \in \Gamma_i$, i.e. $(\forall j \in D)(iRj \Rightarrow (\psi \in T_j \ \& \ \psi \supset \varphi \in T_j)$, and so, by (PC_j), $\varphi \in T_j$, hence, by definition of Γ_i, $\varphi \in \Gamma_i$.
If $\varphi = \Box\psi$ and ψ is a formula of the proof in a previous step, then, by Ind.Hypothesis, $\psi \in \Gamma_i$ and so, by (PI_i) (Lemma 3.8), $\Box\psi \in \Gamma_i$.

(ii) Suppose, for the sake of contradiction, that Γ_i was an *inc***S4.2**-theory. Then $\Gamma_i \vdash_{\mathbf{S4.2}} \bot$, hence, by (i), $\bot \in \Gamma_i$, i.e. by definition of Γ_i and (Gd), $\bot \in T_0$, hence, T_0 is inconsistent, which is a contradiction, by Def.3.1. ∎

4 S4.2 representation of KB_R-structures

First of all, we need to define the theories we will use.

Definition 4.1 *Assume any Kripke model $\mathfrak{M} = \langle W, R, V \rangle$ and any $C \subseteq W$. Then,*

$$Th_{\mathfrak{M}}(C) =_{def} \{\varphi \in \mathcal{L}_\Box \mid (\forall w \in C)\ \mathfrak{M}, w \Vdash \varphi\}$$

$$K_{\mathfrak{M}}(C) =_{def} \{\varphi \in \mathcal{L}_\Box \mid (\forall w \in C)\ \mathfrak{M}, w \Vdash \Box\varphi\}$$

$$B_{\mathfrak{M}}(C) =_{def} \{\varphi \in \mathcal{L}_\Box \mid (\forall w \in C)\ \mathfrak{M}, w \Vdash \neg\Box\neg\Box\varphi\}$$

Intuitively, $Th_{\mathfrak{M}}(C)$ is the theory containing formulas, which are true in every situation of C, $K_{\mathfrak{M}}(C)$ is *everything our agent knows in every situation of* C, and $B_{\mathfrak{M}}(C)$ is *everything she believes in*, in every situation of C. Our first result states that in the case of an epistemic **S4.2**-model, everything she knows and everything she believes in, can

be captured syntactically by the notion of KB_R-structures. Furthermore, everything she believes in, is the same in all clusters, and coincides with everything she knows in the final cluster. For an example, see Section 5.

Theorem 4.2 *Let $\mathfrak{M} = \langle W, R, V\rangle$ be any* **S4.2***-model with clusters $C_i \subseteq W$ $(i \in D)$, where C_0 is the final cluster. Then, there is a relation $P \subseteq D \times D$ such that $\langle (Th_{\mathfrak{M}}(C_i)), (K_{\mathfrak{M}}(C_i)), (B_{\mathfrak{M}}(C_i))\rangle^{P}_{i\in D}$ is a KB_P-structure and $B_{\mathfrak{M}}(C_i) = K_{\mathfrak{M}}(C_0)$.*

Proof. According to Definition 3.1 we should:

(a) prove that all $Th_{\mathfrak{M}}(C_i)$ are consistent theories, and

(b) find a simple, reflexive, transitive and directed relation $P \subseteq D \times D$ s.t. $(\forall i \in D)$

(c) $K_{\mathfrak{M}}(C_i) = \bigcap_{iPj} T_j$ and (PC_i), (P_i) and (N_i) hold for $Th_{\mathfrak{M}}(C_i)$.
As far as the $B_{\mathfrak{M}}(C_i)$'s is concerned, by Fact 3.14, it suffices to prove that $(\forall i \in D)$

(d) $B_{\mathfrak{M}}(C_i) = \bigcup_{iPj} K_{\mathfrak{M}}(C_j)$.

Here are the proofs of (a) to (d).

(a) For convenience, let us denote each $Th_{\mathfrak{M}}(C_i)$ as T_i. Obviously, $T_i \neq \varnothing$ and $\varphi \in T_i \Rightarrow \neg\varphi \notin T_i$, so they are consistent.

(b) Since R is rtd, by Def.2.4, it is meaningful to refer to its pattern-relations. So, let $P \subseteq D \times D$ be a pattern-relation of R (for clusters $C_0, \ldots, C_n$), i.e.

$$iPj \overset{\text{def}}{\iff} (\exists w \in C_i)(\exists v \in C_j)\ wRv$$

Then, by Corollary 2.8, P is an s-rtd-relation.

(c) By Def.4.1, we have to show that for any $i \in D, \varphi \in \mathcal{L}_{\Box}$

$$(\forall w \in C_i)\ \mathfrak{M}, w \Vdash \Box\varphi \iff (\forall j \in D)(iPj \Rightarrow (\forall v \in C_j)\ \mathfrak{M}, v \Vdash \varphi)$$

For $(\Rightarrow)$, consider any $j \in D$ s.t. iPj and any $v \in C_j$. Then, assuming any $w \in C_i$, by Lemma 2.5(i), wRv, hence, by premise, $\mathfrak{M}, v \Vdash \varphi$. For $(\Leftarrow)$, take any $w \in C_i$ and any $v \in W$ s.t. wRv. Then, since $\bigcup_{j\in D} C_j = W$, there is a $j \in D$ s.t. $v \in C_j$, hence, by definition of P, iPj, so, by premise, $\mathfrak{M}, v \Vdash \varphi$, therefore, $\mathfrak{M}, w \Vdash \Box\varphi$.

(PC_i) This propery is obvious, since all formulas of $\mathbf{PC}_{\mathcal{L}_{\Box}}$ are valid in every Kripke model, and since closure of T_i under **MP** is actually the definition of truth of $\varphi \supset \psi$ in a Kripke model.

(P_i) Assume that $\varphi \in \bigcap_{iPj} T_j$ i.e. $\forall j \in D$ s.t. iPj, $(\forall w \in C_j)\ \mathfrak{M}, w \Vdash \varphi$. Let $w \in C_i$ and $v \in W$ s.t. wRv. Then, there is a $j \in D$ s.t. $v \in C_j$, so, by definition of P, iPj, hence, by assumption, $\mathfrak{M}, v \Vdash \varphi$, consequently, $\mathfrak{M}, w \Vdash \Box\varphi$, so $\Box\varphi \in T_i$.

(N_i) Suppose that $\varphi \notin T_i$ i.e. there exists a $w \in C_i$ s.t. $\mathfrak{M}, w \Vdash \neg\varphi$. Let $j \in D$ be s.t. jPi and $v \in C_j$. Then, by Lemma 2.5(i), vRw, hence, since $\mathfrak{M}, w \Vdash \neg\varphi$, $\mathfrak{M}, v \Vdash \neg\Box\varphi$, so, $\neg\Box\varphi \in T_j$, and $\neg\Box\varphi \in \bigcap_{jPi} T_j$.

(d) By Def. 4.1, it suffices to show that for any $i \in D, \varphi \in \mathcal{L}_{\Box}$

$$(\forall w \in C_i)\ \mathfrak{M}, w \Vdash \neg\Box\neg\Box\varphi \iff (\exists j \in D)(iPj\ \&\ (\forall v \in C_j)\ \mathfrak{M}, v \Vdash \Box\varphi)$$

For $(\Rightarrow)$, after considering any $w \in C_i$, using the premise, it must be a $u \in W$ s.t. wRu and $\mathfrak{M}, u \Vdash \Box\varphi$. Since $\bigcup_{j\in D} C_j = W$, there is a $j \in D$ s.t. $u \in C_j$, hence, by definition of P, iPj. Furthermore, consider any $v \in C_j$. Since C_j is a cluster, uRv. Finally, let $s \in W$ s.t. vRs. R is transitive, so uRs, hence (because $\mathfrak{M}, u \Vdash \Box\varphi$) $\mathfrak{M}, s \Vdash \varphi$ i.e. $\mathfrak{M}, v \Vdash \Box\varphi$.

For $(\Leftarrow)$, consider any $w \in C_i$ and any $v \in C_j$, where j is the integer, whose existence is guaranteed by the premise. Then, $\mathfrak{M}, v \Vdash \Box\varphi$ and, since iPj, by Lemma 2.5(i), wRv, hence, $\mathfrak{M}, w \Vdash \neg\Box\neg\Box\varphi$. ∎

Similarly to parts (a), (b) and (c) of the proof of Theorem 4.2 one can prove that, having fixed modal-free, consistent and closed under propositional consequence theories $S_0, \ldots, S_n$ and an s-rtd-relation P, we can find a KB_P-structure $\langle (T_i), (\ulcorner_i), (\vartriangle_i)\rangle^P_{i\in D}$ such that the non-modal part of the theories $T_0, \ldots, T_n$, is exactly $S_0, \ldots, S_n$ respectively.

Proposition 4.3 *Let $S_0, \ldots, S_n \subseteq \mathcal{L}$ be modal-free, consistent and closed under propositional consequence theories, and $P \subseteq D \times D$ an s-rtd-relation. Then, there exists a KB_P-structure $\langle (T_i), (\ulcorner_i), (\vartriangle_i)\rangle^P_{i\in D}$ s.t. $T_i \cap \mathcal{L} = S_i \quad (i \in D)$.*

Proof. Consider the model $\mathfrak{M} = \langle W, R, V\rangle$, where

- $W = \bigcup_{i\in D} C_i$, where
$$C_i = \{(i, w) \in D \times (\Phi \mapsto \{\mathbf{t}, \mathbf{f}\}) \mid (\forall\varphi \in S_i)\ \overline{w}(\varphi) = \mathbf{t}\} \quad (i \in D)$$
- $R = \bigcup_{iPj} C_i \times C_j$
- $V(p) = \bigcup_{i\in D}\{(i, w) \in C_i \mid w(p) = \mathbf{t}\} \quad (p \in \Phi)$

Every C_i consists of all (indexed by i) propositional valuations which satisfy S_i (note that Φ is the set of all propositional variables and $(\Phi \mapsto \{\mathbf{t}, \mathbf{f}\})$ is the set of all functions from Φ to $\{\mathbf{t}, \mathbf{f}\}$). Let us now fix any $i \in D$. First of all, let us point out that the notion $\overline{w}(\varphi)$ is meaningful, since $S_i \subseteq \mathcal{L}$, and so, $\varphi \in \mathcal{L}$.

Furthermore, $C_i \neq \varnothing$, since S_i is consistent, and hence, by the completeness Theorem for propositional logic, S_i is satisfiable. Next, using the definitions of V and of propositional valuations, by a trivial induction on the complexity of φ, we can prove that, $(\forall(i, w) \in W)(\forall\varphi \in \mathcal{L})$
$$\mathfrak{M}, (i, w) \Vdash \varphi \iff \overline{w}(\varphi) = \mathbf{t} \tag{4.3.i}$$

Now, consider any $\varphi \in S_i$. Then, by the definition of C_i, $(\forall(i, w) \in C_i)\ \overline{w}(\varphi) = \mathbf{t}$, so, by (4.3.i), since $\varphi \in \mathcal{L}$, $(\forall(i, w) \in C_i)\ \mathfrak{M}, (i, w) \Vdash \varphi$, hence, $\varphi \in Th_{\mathfrak{M}}(C_i)$, and of course, $\varphi \in Th_{\mathfrak{M}}(C_i) \cap \mathcal{L}$.

Conversely, let $\varphi \in Th_{\mathfrak{M}}(C_i) \cap \mathcal{L}$. Then, $(\forall(i, w) \in C_i)\ \mathfrak{M}, (i, w) \Vdash \varphi$, consequently, by (4.3.i), since $\varphi \in \mathcal{L}$, $(\forall(i, w) \in C_i)\ \overline{w}(\varphi) = \mathbf{t}$, so, by the definition of C_i, $S_i \vDash \varphi$, hence, by the completeness Theorem for propositional logic, $S_i \vdash_{\mathbf{PC}} \varphi$, and since S_i is closed under propositional consequence, $\varphi \in S$.

Hence, the assertion will follow for theories $Th_{\mathfrak{M}}(C_i)$, by proving additionally that $\langle (Th_{\mathfrak{M}}(C_i)), (\mathsf{r}_i), (\vartriangle_i)\rangle^R_{i\in D}$ is a KB_P-structure. But, by the construction of R, since P is rtd, so is R. Then, it is meaningful to refer to its pattern-relations. Let us focus on P. Firstly, P is a binary relation on D. Next, for any $i, j \in D$, assume that iPj. Then, by definition of R, $(\forall (i,w) \in C_i, (j,v) \in C_j)\ (i,w)R(j,v)$, and since $C_i, C_j \neq \varnothing$, $(\exists (i,w) \in C_i, (j,v) \in C_j)\ (i,w)R(j,v)$.

Conversely, suppose that $(\exists (i,w) \in C_i, (j,v) \in C_j)\ (i,w)R(j,v)$. Then, by definition of R, there must be $i', j' \in D$ and $C_{i'}, C_{j'}$ s.t. $i'Pj'$ and $(i,w) \in C_{i'}, (j,v) \in C_{j'}$. But then, by the definition of the C_i's, $i' = i$ and $j' = j$, hence, iPj. Furthermore, by construction of R, all C_i $(i \in D)$ are clusters of R. Hence, all this shows, by Def.2.4, that P is a pattern-relation of R (for clusters $C_0, \ldots, C_n$). Now, we continue our proof exactly as in parts (a), (b) and (c) (only for (PC_i), (P_i) and (N_i)) of the proof of Theorem 4.2, and we conclude that $\langle (Th_{\mathfrak{M}}(C_i)), (\mathsf{r}_i), (\vartriangle_i)\rangle^R_{i\in D}$ is a KB_P-structure. ∎

As an application of Proposition 4.3, let us consider again Ohlbach's s-rtd-relation R of example 3.3. Furthermore, consider $p \in \Phi$, $S_0 = Cn_{\mathbf{PC}_{\mathcal{L}}}(\{p\})$ and $S_1 = Cn_{\mathbf{PC}_{\mathcal{L}}}(\varnothing)$. It is easy to see that $S_0 = Cn_{\mathbf{PC}_{\mathcal{L}}}(S_0)$ and $S_1 = Cn_{\mathbf{PC}_{\mathcal{L}}}(S_1)$. Clearly, both are satisfiable, hence, by the soundness theorem for propositional logic, they are consistent. So, by Proposition 4.3, there is a KB_R-structure $\langle (T_0, T_1), (\mathsf{r}_0, \mathsf{r}_1), (\vartriangle_0, \vartriangle_1)\rangle^R$ s.t. $T_0 \cap \mathcal{L} = S_0$ and $T_1 \cap \mathcal{L} = S_1$. Hence, $p \in T_0$ and $p \notin T_1$ (for otherwise, $p \in S_1$, so $\vdash_{\mathbf{PC}_{\mathcal{L}}} p$, hence p would be a tautology, which is absurd). Then, since $p \notin T_1$, by definition of r_1, $p \notin \mathsf{r}_1$. But, $p \in T_0$, hence, by (P_0), $\Box p \in T_0$, and since T_0 is consistent, $\neg\Box p \notin T_0$, so, $\neg\Box p \notin \mathsf{r}_1$. Therefore, $p \notin \mathsf{r}_1 \nRightarrow \neg\Box p \in \mathsf{r}_1$. This counterexample verifies the next lemma, which is most welcomed.

Lemma 4.4 *There are KB_R-structures, whose knowledge-part (some r_i's) does not satisfy the negative introspection property concerning knowledge.*

Our next goal is to prove the converse of Theorem 4.2, i.e. for a given KB_R-structure, there is an epistemic **S4.2**-model, in which everything an agent knows and believes, is described by the KB_R-structure given, and furthermore, everything she believes, is described by one of the knowledge-theories in structure KB_R. The model we are searching for, will be a construction similar to the well known canonical model for a modal logic, and it will be based on the normal modal logic **S4.2**, which we will denote as Λ.

Definition 4.5 *Let $\langle (T_i), (\mathsf{r}_i), (\vartriangle_i)\rangle^R_{i\in D}$ be any KB_R-structure. The* canonical model *for it, is Kripke model $\mathfrak{M}^c = \langle W^c, R^c, V^c\rangle$, where*

(i) $W^c = \bigcup_{i\in D} W^c_i$, *where* $W^c_i = \{(i,\Theta) \in D \times \mathcal{P}(\mathcal{L}_\Box) \mid \Theta : m\mathsf{r}_i c\Lambda\}$ $(i \in D)$

(ii) $\big(\forall (i,\Theta),(j,\mathsf{z}) \in W^c\big)\big((i,\Theta)R^c(j,\mathsf{z}) \iff$
$(iRj \quad \& \quad (\forall \varphi \in \mathcal{L}_\Box)(\Box\varphi \in \Theta \Rightarrow \varphi \in \mathsf{z}))\big)$

(iii) $(\forall p \in \Phi)(V^c(p) = \{(i,\Theta) \in W^c \mid p \in \Theta\})$

The following lemmata will be useful in our main theorem. The first two are presented without proofs, since they are well-known, classical results.

Lemma 4.6 *Let Λ be any normal modal logic, I a $c\Lambda$-theory, Θ a $mIc\Lambda$-theory and φ, ψ $\mathcal{L}_\Box$-formulae. Then,*

(i) Θ *is closed under* **MP**

(ii) *either* $\varphi \in \Theta$ *or* $\neg\varphi \in \Theta$

(iii) $I \vdash_\Lambda \varphi \iff (\forall \mathsf{z} : mIc\Lambda)\varphi \in \mathsf{z}$

(iv) $\varphi \wedge \psi \in \Theta \iff (\varphi \in \Theta$ *and* $\psi \in \Theta)$

Lemma 4.7 (Lindenbaum) *Let Λ be any normal modal logic, I a $c\Lambda$-theory and T an $Ic\Lambda$-theory. Then, there is a $mIc\Lambda$ theory Θ s.t. $T \subseteq \Theta$.*

Remark 4.8 Firstly, notice that W^c is the disjoint union of all $m\Gamma_i c\Lambda$ theories with indexes in D. Furthermore, by Lemma 3.16(ii), every Γ_i $(i \in D)$ is $c\Lambda$, hence, to refer to $m\Gamma_i c\Lambda$-theories is meaningful, and $\Gamma_i \nvdash_\Lambda \bot$, so, $\{\top\}$ is $\Gamma_i c\Lambda$, and by Lindenbaum's Lemma, there exists a $m\Gamma_i c\Lambda$-theory (which, by the way, contains $\{\top\}$), therefore, every $W^c_i \neq \varnothing$ $(i \in D)$.

Lemma 4.9 (Truth Lemma) $(\forall \varphi \in \mathcal{L}_\Box)(\forall (i, \Theta) \in W^c)(\mathfrak{M}^c, (i, \Theta) \Vdash \varphi \Leftrightarrow \varphi \in \Theta)$

Proof. The proof runs by induction on the complexity of φ. The *induction basis* follows immediately from Def. 4.5(iii). For the *induction step*, the first part, concerning $\varphi \supset \psi$, follows trivially from the induction hypothesis using items (i) to (iv) of Lemma 4.6. Now, for the second part of the induction step, the $\Box\varphi$ case:

$\mathfrak{M}^c, (i, \Theta) \Vdash \Box\varphi$ iff $(\forall (j, \mathsf{z}) \in W^c)((i, \Theta)R^c(j, \mathsf{z}) \Rightarrow \mathfrak{M}^c, (j, \mathsf{z}) \Vdash \varphi)$ iff (by Ind.Hyp.) $(\forall (j, \mathsf{z}) \in W^c)((i, \Theta)R^c(j, \mathsf{z}) \Rightarrow \varphi \in \mathsf{z})$. It suffices to show that this is equivalent to the fact that $\Box\varphi \in \Theta$.

$(\Rightarrow)$: Suppose that $\Box\varphi \notin \Theta$. Notice that, since $(i, \Theta) \in W^c$, Θ is $m\Gamma_i c\Lambda$. Now, let us define $\mathsf{H} = \{\psi \in \mathcal{L}_\Box \mid \Box\psi \in \Theta\}$ and $\mathsf{I} = \{\neg\varphi\} \cup \mathsf{H}$. Suppose, for the sake of contradiction, that I was $\Gamma_i inc\Lambda$ i.e. there exist $\psi_1, \ldots, \psi_n \in \mathsf{I}$ s.t. $\Gamma_i \vdash_\Lambda \psi_1 \wedge \ldots \wedge \psi_n \supset \bot$.

- if $n = 1$ and $\psi_1 = \neg\varphi$ i.e. $\Gamma_i \vdash_\Lambda \varphi$, then, by (RN), $\Gamma_i \vdash_\Lambda \Box\varphi$, hence, since Θ is $m\Gamma_i c\Lambda$, by Lemma 4.6(iii), $\Box\varphi \in \Theta$, which is a contradiction.

- if $\psi_1, \ldots, \psi_n \in \mathsf{H}$, then $\Gamma_i \vdash_\Lambda \psi_1 \wedge \ldots \wedge \psi_n \supset \phi$, since $\bot \supset \varphi \in \mathbf{PC}$.
  if $n > 1$ and $\psi_1, \ldots, \psi_{n-1} \in \mathsf{H}$ and $\psi_n = \neg\varphi$, then $\Gamma_i \vdash_\Lambda \psi_1 \wedge \ldots \wedge \psi_{n-1} \supset \phi$.
  So, in both cases, there are $\psi_1, \ldots, \psi_n \in \mathsf{H}$ with $n \geq 1$ s.t. $\Gamma_i \vdash_\Lambda \psi_1 \wedge \ldots \wedge \psi_n \supset \varphi$. Hence, by **RN** and using **K** (and by a trivial induction), $\Gamma_i \vdash_\Lambda \Box\psi_1 \wedge \ldots \wedge \Box\psi_n \supset \Box\varphi$, so, by Lemma 4.6(iii), $\Box\psi_1 \wedge \ldots \wedge \Box\psi_n \supset \Box\varphi \in \Theta$. But, since $\psi_1, \ldots, \psi_n \in \mathsf{H}$, by definition, $\Box\psi_1, \ldots, \Box\psi_n \in \Theta$, therefore, by Lemma 4.6(iv),(i), $\Box\varphi \in \Theta$, which is again a contradiction.

So, I is a $\mathsf{r}_i c\Lambda$-theory, and by Lindenbaum's lemma, there is a $m\mathsf{r}_i c\Lambda$-theory z s.t. $\mathsf{I} \subseteq \mathsf{z}$, hence, $\neg\varphi \in \mathsf{z}$, which entails, by Lemma 4.6(ii), that $\varphi \notin \mathsf{z}$.

Furthermore, since R is reflexive, iRi, and $(\forall \psi \in \mathcal{L}_\Box)$, if $\Box\psi \in \Theta$, then, by definition, $\psi \in \mathsf{H}$, hence, $\psi \in \mathsf{I}$, so, $\psi \in \mathsf{z}$. Therefore, by Def.4.5(ii), $(i, \Theta)R^c(i, \mathsf{z})$.

($\Leftarrow$) Suppose that $\Box\varphi \in \Theta$ and let $(j, \mathsf{z}) \in W^c$ be s.t. $(i, \Theta)R^c(j, \mathsf{z})$. Then, by Def.4.5(ii), $\varphi \in \mathsf{z}$. ∎

Lemma 4.10 *Let $\langle (T_i), (\mathsf{r}_i), (\Delta_i)\rangle^R_{i\in D}$ be any KB_R-structure, and $\mathfrak{M}^c$ its canonical model. Then, $(\forall i \in D)(\forall \varphi \in \mathcal{L}_\Box)$*

$$\mathsf{r}_i \vdash_\Lambda \varphi \iff (\forall (i, \Theta) \in W^c_i)\ \mathfrak{M}^c, (i, \Theta) \Vdash \Box\varphi$$

Proof. For ($\Rightarrow$), assume that $\mathsf{r}_i \vdash_\Lambda \varphi$. Then, by (RN), $\mathsf{r}_i \vdash_\Lambda \Box\varphi$, hence, by Lemma 4.6(iii), $(\forall \Theta : m\mathsf{r}_i c\Lambda)\ \Box\varphi \in \Theta$, so, by Truth Lemma, $(\forall (i, \Theta) \in W^c_i)\ \mathfrak{M}^c, (i, \Theta) \Vdash \Box\varphi$.

For ($\Leftarrow$), suppose that $\mathsf{r}_i \nvdash_\Lambda \varphi$ and, for the sake of contradiction, that $\{\neg\Box\varphi\}$ was $\mathsf{r}_i inc\Lambda$. Then, $\mathsf{r}_i \vdash_\Lambda \Box\varphi$, hence, since $\mathbf{T} \in \Lambda$, $\mathsf{r}_i \vdash_\Lambda \varphi$, which is a contradiction. So, $\{\neg\Box\varphi\}$ is $\mathsf{r}_i c\Lambda$, therefore, by Lindenbaum's Lemma, there is a $\Theta : m\mathsf{r}_i c\Lambda$, i.e. a pair $(i, \Theta) \in W^c_i$ s.t. $\neg\Box\varphi \in \Theta$, hence, by Truth Lemma, $\mathfrak{M}^c, (i, \Theta) \nVdash \Box\varphi$. ∎

Now, we are ready to prove a representation theorem for KB_R-structures.

Theorem 4.11 *Let $\langle (T_i), (\mathsf{r}_i), (\Delta_i)\rangle^R_{i\in D}$ be any KB_R-structure. Then, there exists an* **S4.2***-model $\mathfrak{M} = \langle W, R, V\rangle$ and $C_i \subseteq W$ s.t. $(\forall i \in D)$*

$$\mathsf{r}_i = K_\mathfrak{M}(C_i) \qquad \Delta_i = B_\mathfrak{M}(C_i) = \mathsf{r}_0$$

Proof. Consider the canonical model $\mathfrak{M}^c$ for KB_R-structure $\langle (T_i), (\mathsf{r}_i), (\Delta_i)\rangle^R_{i\in D}$ and set $C_i =_{\text{def}} W^c_i$. First of all, we will check whether $\mathfrak{M}^c$ is indeed an **S4.2**-model.

For reflexivity, fix any $i \in D$ and consider any $(i, \Theta) \in W^c_i$ and a $\varphi \in \mathcal{L}_\Box$ s.t. $\Box\varphi \in \Theta$. Then, since $\mathsf{r}_i \vdash_\Lambda \mathbf{T}$ and since Θ is $m\mathsf{r}_i c\Lambda$, by Lemma 4.6(iii), $\mathbf{T} \in \Theta$, hence, by Lemma 4.6(i), $\varphi \in \Theta$. Furthermore, since R is reflexive, iRi, hence, by Def. 4.5(ii), $(i, \Theta)R^c(i, \Theta)$.

For transitivity, let $(i, \Theta) \in W^c_i$, $(j, \mathsf{z}) \in W^c_j$ and $(k, \mathsf{H}) \in W^c_k$ be s.t. $(i, \Theta)R^c(j, \mathsf{z})$ and $(j, \mathsf{z})R^c(k, \mathsf{H})$. Furthermore, consider any $\varphi \in \mathcal{L}_\Box$ s.t. $\Box\varphi \in \Theta$. Then, since $\mathsf{r}_i \vdash_\Lambda \mathbf{4}$ and since Θ is $m\mathsf{r}_i c\Lambda$, by Lemma 4.6(iii), $\mathbf{4} \in \Theta$, hence, by Lemma 4.6(i), $\Box\Box\varphi \in \Theta$, so, since $(i, \Theta)R^c(j, \mathsf{z})$, $\Box\varphi \in \mathsf{z}$, and since $(j, \mathsf{z})R^c(k, \mathsf{H})$, $\varphi \in \mathsf{H}$. Furthermore, since R is transitive and since iRj and jRk, iRk. Hence, by Def. 4.5(ii), $(i, \Theta)R^c(k, \mathsf{H})$.

For directedness, we will firstly prove that

$$(\forall (i, \Theta) \in W^c_i)(\forall (0, \mathsf{H}) \in W^c_0)\ (i, \Theta)R^c(0, \mathsf{H}) \tag{4.11.i}$$

Let $\varphi \in \mathcal{L}_\Box$ be s.t. $\Box\varphi \in \Theta$. Suppose, for the sake of contradiction, that $\neg\Box\varphi \in \mathsf{r}_i$. Then, $\mathsf{r}_i \vdash_\Lambda \neg\Box\varphi$, hence (since Θ is $m\mathsf{r}_i c\Lambda$), by Lemma 4.6(iii), $\neg\Box\varphi \in \Theta$, so, Θ would be

inconsistent, and hence, $\mathsf{r}_i inc \Lambda$, which is a contradiction. So, $\neg\Box\varphi \notin \mathsf{r}_i$, i.e. by definition of r_i, there is a $k \in D$ s.t. iRk and $\neg\Box\varphi \notin T_k$. But then, by (2) of Fact 3.4, $\varphi \in \mathsf{r}_k$. Furthermore, by (Gd), $kR0$, hence, by Lemma 3.11, $\mathsf{r}_k \subseteq \mathsf{r}_0$, so, $\varphi \in \mathsf{r}_0$, consequently, $\mathsf{r}_0 \vdash_\Lambda \varphi$, and since H is $m\mathsf{r}_0 c\Lambda$, by Lemma 4.6(iii), $\varphi \in \mathsf{H}$. It has been proved that, if $\Box\varphi \in \Theta$, then $\varphi \in \mathsf{H}$. Additionally, $iR0$, hence, by Def. 4.5(ii), $(i, \Theta)R^c(0, \mathsf{H})$. So, (4.11.i) has been proved.

Now, consider any $(i, \Theta) \in W_i^c$, $(j, \mathsf{Z}) \in W_j^c$ and fix a $(0, \mathsf{H}) \in W_0^c$. Then, by (4.11.i), $(i, \Theta)R^c(0, \mathsf{H})$ and $(j, \mathsf{Z})R^c(0, \mathsf{H})$, which entails directedness of $\mathfrak{M}^c$.

We come now to theories of Knowledge $K_{\mathfrak{M}^c}$. By Lemma 3.16(i), $\varphi \in \mathsf{r}_i$ iff $\mathsf{r}_i \vdash_\Lambda \varphi$. Furthermore, $\mathsf{r}_i \vdash_\Lambda \varphi$ iff, by Lemma 4.10, $(\forall (i, \Theta) \in C_i)\ \mathfrak{M}^c, (i, \Theta) \Vdash \Box\varphi$ iff, by Def. 4.1, $\varphi \in K_{\mathfrak{M}^c}(C_i)$. Hence,

$$\mathsf{r}_i = K_{\mathfrak{M}^c}(C_i) \tag{4.11.ii}$$

Finally, we will focus on theories of Belief $B_{\mathfrak{M}^c}$. Consider any $\varphi \in B_{\mathfrak{M}^c}(C_i)$. Then, $(\forall (i, \Theta) \in W_i^c)\ \mathfrak{M}^c, (i, \Theta) \Vdash \neg\Box\neg\Box\varphi$, i.e. there exists a $(j, \mathsf{Z}) \in W_j^c$ s.t. $(i, \Theta)R^c(j, \mathsf{Z})$ and $\mathfrak{M}^c, (j, \mathsf{Z}) \Vdash \Box\varphi$. Now, assume that $(0, \mathsf{H}) \in W_0^c$ and let $(k, \mathsf{H}') \in W_k^c$ s.t. $(0, \mathsf{H})R^c(k, \mathsf{H}')$, i.e. by Def. 4.5(ii), $0Rk$, hence, by (Gd), $k = 0$. This means that $(k, \mathsf{H}') = (0, \mathsf{H}') \in W_0^c$. But then, by (4.11.i), $(j, \mathsf{Z})R^c(0, \mathsf{H}')$. So, and since $\mathfrak{M}^c, (j, \mathsf{Z}) \Vdash \Box\varphi$, it is true that $\mathfrak{M}^c, (k, \mathsf{H}') \Vdash \varphi$, hence, $\mathfrak{M}^c, (0, \mathsf{H}) \Vdash \Box\varphi$, consequently, $(\forall (0, \mathsf{H}) \in W_0^c)\ \mathfrak{M}^c, (0, \mathsf{H}) \Vdash \Box\varphi$, which entails, by Def. 4.1, $\varphi \in K_{\mathfrak{M}^c}(W_0^c)$, and by (4.11.ii), $\varphi \in \mathsf{r}_0$.

Conversely, suppose that $\varphi \in \mathsf{r}_0$, i.e. again by (4.11.ii), $\varphi \in K_{\mathfrak{M}^c}(W_0^c)$. Consider any $(i, \Theta) \in W_i^c$. Then, by (4.11.i), there is a $(0, \mathsf{H}) \in W_0^c$ s.t. $(i, \Theta)R^c(0, \mathsf{H})$ (in fact, it is true for all elements of W_0^c, which is, by Remark 4.8, non-empty). Hence, since $\varphi \in K_{\mathfrak{M}^c}(W_0^c)$, by Def. 4.1, $\mathfrak{M}^c, (0, \mathsf{H}) \Vdash \Box\varphi$, so, $\mathfrak{M}^c, (i, \Theta) \Vdash \neg\Box\neg\Box\varphi$, i.e. by Def. 4.1, $\varphi \in B_{\mathfrak{M}^c}(C_i)$. Therefore, proof of $B_{\mathfrak{M}^c}(C_i) = \mathsf{r}_0$ $(i \in D)$ is complete. And of course, by Fact 3.14(i), $\Delta_i = \mathsf{r}_0$ $(i \in D)$. ∎

5 A detailed example

Let us present here an epistemic model example, where an agent (A) is provided with information, which depends on the current agent's situation. Suppose that (A) is interested in the current raining conditions in Stockholm, Athens and Paris, and tries to get some information from a friend of hers, which is a meteorologist (M). The source of information for (A) is only (M). We assume that (M) responds to (A)'s struggle for information very reluctantly, as follows.

1. If it is raining only in Athens, then (M) tells (A) that "It's raining in Stockholm or in Athens".

2. If it is raining in Stockholm and not in all three cities, then (M) says "It's raining in Stockholm or in two cities overall".

3. If it isn't raining in Stockholm nor in Athens, then (M) says "It isn't raining in Stockholm nor in Athens, or it is raining in Athens and Paris".

4. If it is raining in Athens and Paris, then (M) is very talkative this time and announces "It's raining in Athens and Paris"!

Assuming a language with only three propositional variables, namely s, a and p (corresponding to the facts that it is raining in Stockholm, Athens or Paris, respectively), and considering the assertions above, we can construct the epistemic model shown in Figure 1 on page 23. This is a typical **S4.2**-model with clusters C_0 to C_3. C_0 is the final cluster. Since each T_i contains all formulas true (everywhere) in C_i, and taking in account the assertions 1 to 4 (in this order), we conclude that $\neg s \wedge a \wedge \neg p \in T_3$, $s \wedge \neg(a \wedge p) \in T_2$, $\neg s \wedge \neg a \in T_1$, and $a \wedge p \in T_0$.

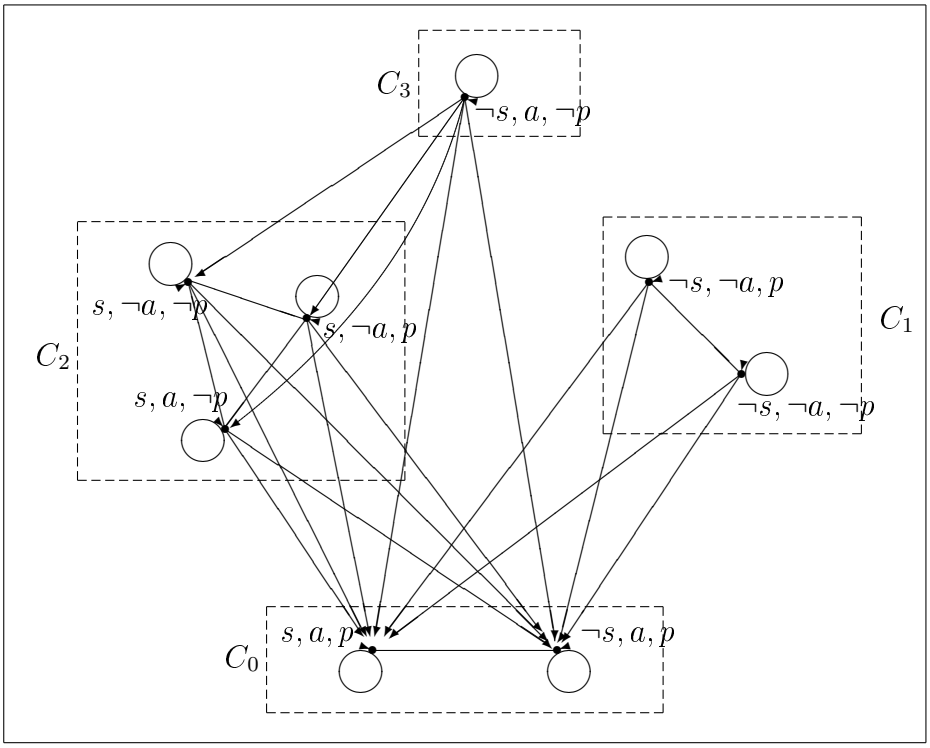


Figure 1:

Now, Theorem 4.2 and Definition 3.1 state that everything that is true (T_i) in each cluster satisfies the following properties (the s-rtd-relation on $D = \{0, 1, 2, 3\}$ guaranteed by Theorem 4.2 is $P = \{(0,0), (1,1), (1,0), (2,2), (2,0), (3,3), (3,2), (3,0)\}$):

$(PC_{0,1,2,3})$ $\quad \mathbf{PC}_{\mathcal{L}_\Box} \subseteq T_0, T_1, T_2, T_3$ and T_0, T_1, T_2, T_3 are closed under **MP**

(P_0) $\quad \varphi \in T_0 \Rightarrow \Box\varphi \in T_0$

(N_0) $\quad \varphi \notin T_0 \Rightarrow \neg\Box\varphi \in T_0 \;\&\; \neg\Box\varphi \in T_1 \;\&\; \neg\Box\varphi \in T_2 \;\&\; \neg\Box\varphi \in T_3$

(P_1) $\quad \varphi \in T_0 \;\&\; \varphi \in T_1 \Rightarrow \Box\varphi \in T_1$

(N_1) $\varphi \notin T_1 \Rightarrow \neg\Box\varphi \in T_1$

(P_2) $\varphi \in T_0$ & $\varphi \in T_2 \Rightarrow \Box\varphi \in T_2$

(N_2) $\varphi \notin T_2 \Rightarrow \neg\Box\varphi \in T_2$ & $\neg\Box\varphi \in T_3$

(P_3) $\varphi \in T_0$ & $\varphi \in T_2$ & $\varphi \in T_3 \Rightarrow \Box\varphi \in T_3$

(N_3) $\varphi \notin T_3 \Rightarrow \neg\Box\varphi \in T_3$

Furthermore, Theorem 4.2 and Definition 3.1 say that everything an agent knows (r_i) in a cluster C_i is exactly: $\mathsf{r}_0 = T_0$, $\mathsf{r}_1 = T_0 \cap T_1$, $\mathsf{r}_2 = T_0 \cap T_2$, and $\mathsf{r}_3 = T_0 \cap T_2 \cap T_3$, and everything she believes in any situation is the same and builds set T_0.

Now, taking into accountthe properties above, some examples follow, of what our agent *knows* and *believes* in this epistemic model.

- Since each T_i is closed under propositional consequence ((PC_0) to (PC_3)),

$$\begin{array}{rcccccc} s \vee a & \in & T_3, & T_2, & & T_0 \\ s \vee (a \wedge p) & \in & & T_2, & & T_0 \\ (\neg s \wedge \neg a) \vee (a \wedge p) & \in & & & T_1, & T_0 \\ a \wedge p & \in & & & & T_0 \end{array}$$

  Hence, $s \vee a \in \mathsf{r}_3$, $s \vee (a \wedge p) \in \mathsf{r}_2$, $(\neg s \wedge \neg a) \vee (a \wedge p) \in \mathsf{r}_1$, and $a \wedge p \in \mathsf{r}_0$.

- If it is raining only in Athens, our agent is in C_3, and since $\neg s \wedge a \wedge \neg p \in T_3$, by (PC_3), $\neg p \in T_3$, hence, since T_3 is consistent, $p \notin T_3$, therefore, $a \wedge p \notin T_3$, and by definition of r_3, $a \wedge p \notin \mathsf{r}_3$, i.e. she **is not sure** that it is raining in Athens and Paris (which is good, since it isn't true!). But, by definition of $\vartriangle_3$, since $a \wedge p \in \mathsf{r}_0$ and $(3,0) \in P$, $a \wedge p \in \vartriangle_3$, i.e. she **believes** that it is raining in Athens and Paris.

- In the same situation as above, $\neg s \wedge a \wedge \neg p \notin \mathsf{r}_3$ (since one can easily see that $\neg s \wedge a \wedge \neg p \notin T_0$, and $(3,0) \in P$), i.e. our agent doesn't know that it is raining only in Athens; she **doesn't know in which situation she is located**. Furthermore, since $\neg s \wedge a \wedge \neg p \notin T_0$ and $(3,0), (2,0), (0,0) \in P$, by (N_0), $\neg\Box(\neg s \wedge a \wedge \neg p) \in T_0 \cap T_2 \cap T_3$, hence, $\neg\Box(\neg s \wedge a \wedge \neg p) \in \mathsf{r}_3$, i.e. she **is aware of her ignorance** about the fact that it is raining only in Athens.

- By construction of this model, it is not necessarily true that it is raining in Athens in every situation of C_2, i.e. $a \notin T_2$, hence, since $(3,2) \in P$, $a \notin \mathsf{r}_3$, which means that in the situation of C_3 our agent **does not know** that it is raining in Athens, although it is true. Furthermore, $a \in T_0$, so, by (P_0), $\Box a \in T_0$, hence, $\neg\Box a \notin T_0$, and since $(3,0) \in P$, $\neg\Box a \notin \mathsf{r}_3$, i.e. our agent **does not know that she doesn't know** that it is raining in Athens; she believes she might know it! (this is also verified by the fact that $\Box a \in T_0 = \mathsf{r}_0 = \vartriangle_3$). So, in this case our agent does not possess negatine introspection.

- In any situation of C_2, since $s \wedge \neg(a \wedge p) \in T_2$, $a \wedge p \notin T_2$, hence, $a \wedge p \notin \mathsf{r}_2$, i.e. she **doesn't know** that it is raining in Athens and Paris, which is rather expectable, since this fact is simply false in every situation of C_2. But, $a \wedge p \in T_0$, so, by (P_0), $\Box(a \wedge p) \in T_0$, therefore, $\neg\Box(a \wedge p) \notin T_0$, and since $(2,0) \in P$, $\neg\Box(a \wedge p) \notin \mathsf{r}_2$, hence, she **does not know that she doesn't know** this fact. And, $\Box(a \wedge p) \in T_0 = \mathsf{r}_0 = \mathsf{a}_2$, hence, she believes (falsely) that she knows that it is raining in both cities. This time, our agent is again not negative-introspective, but in a more 'severe' situation, since she believes that she knows something, which is wrong.

6 Related Work - Further Research

The identification of logical theories, which capture the epistemic content of a rational agent's view of the world, is a very important topic in *Knowledge Representation*. A very important notion has been the notion of a *stable belief set*, introduced by R.Stalnaker [Sta93] and further investigated in modal non-monotonic reasoning [MT93]. The original motivation of this paper (rather distinctly far from the final result) has been the idea to derive logically interesting notions of stable epistemic states out of a model-theoretic starting point, and prove that they posses intuitive syntactic characterizations. This seems natural to do: stable belief sets can be represented as **S5** theories or sets of beliefs held inside a **KD45** situation [Hal97],[MT93]. In a previous paper [KZ10] we obtained interesting syntactic variations of epistemic states and proved representating theorems, in terms of possible-world models for non-normal modal logics. It (still) seems natural to investigate the other way around: to define epistemic theories in terms of possible worlds models for interesting epistemic logics (such as $\mathbf{S4.2}, \mathbf{S4.4}$), and then match this definitions to closure under intuitive *context-rules*, such as the ones encountered in Stalnaker's initial definition. On the way, it became clear to us that, from a purely epistemological viewpoint that takes into account the information available to the agent inside each situation, the **S5**-like analysis of epistemic reasoning is too simple to furnish a realistic view (although there exists a compensation, in terms of various handy technical properties). Thus, we took a step back to start from the very beginning: the notion of accessibility between possible worlds, its epistemic content and logical interpretations. This led us to the semantic analysis discussed in section 3.1 and to the origination of KB_R-structures.

The KB_R-structures introduced here represent a somewhat complex, yet interesting, description of the epistemic status of a rational (but not fully introspective) agent, allowing a differentiation of knowledge from belief. It would be interesting to embed them in core KR techniques, such as default reasoning or belief revision; actually it is a very challenging (albeit complex) task to define reasoning procedures that will take into account the subtle differences between knowledge and belief. Such a task is bound to be complex but it will be necessarily useful to deviate from the currently dominating model of a *logically omniscient, fully introspective* agent. As a short-term goal, it is definitely interesting to identify the computational complexity of reasoning with KB_R-structures.

Acknowledgments. The paper is a fully revised and expanded version of the extended abstract [KZ11] which appeared in ECSQARU 2011. The second author gratefully acknowledges financial support by the *Greek Ministry of Education, Lifelong Learning and Religious Affairs* under the scheme of educational leave.

References

[BdRV01] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic.* Number 53 in Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 2001.

[Che80] B. F. Chellas. *Modal Logic, an Introduction.* Cambridge University Press, 1980.

[FHMV03] R. Fagin, J. Halpern, Y. Moses, and M. Vardi. *Reasoning about Knowledge.* MIT Press, 2003.

[Fit93] M. C. Fitting. *Basic Modal Logic*, pages 368–448. Volume 1 of Gabbay et al. [GHA93], 1993.

[GHA93] D. M. Gabbay, C. J. Hogger, and J. A.Robinson, editors. *Handbook of Logic in Artificial Intelligence and Logic Programming.* Oxford University Press, 1993.

[Gol92] R. Goldblatt. *Logics of Time and Computation.* Number 7 in CSLI Lecture Notes. Center for the Study of Language and Information, Stanford University, 2nd edition, 1992.

[Hal96] J. Halpern. Should knowledge entail belief? *Journal of Philosophical Logic*, 25(5):483–494, 1996.

[Hal97] J. Halpern. A theory of knowledge and ignorance for many agents. *Journal of Logic and Computation*, 7(1):79–108, 1997.

[HC96] G. E. Hughes and M. J. Cresswell. *A New Introduction to Modal Logic.* Routledge, 1996.

[Hin62] J. Hintikka. *Knowledge and Belief: an Introduction to the Logic of the two notions.* Cornell University Press, Ithaca, NY, 1962.

[JN10] Tomi Janhunen and Ilkka Niemelä, editors. *Logics in Artificial Intelligence - 12th European Conference, JELIA 2010, Helsinki, Finland, September 13-15, 2010. Proceedings*, volume 6341 of *Lecture Notes in Computer Science.* Springer, 2010.

[KT07] M. Kaminski and M. L. Tiomkin. The modal logic of cluster-decomposable kripke interpretations. *Notre Dame Journal of Formal Logic*, 48(4):511–520, 2007.

[KZ10] C. D. Koutras and Y. Zikos. Stable belief sets revisited. In Janhunen and Niemelä [JN10], pages 221–233.

[KZ11] Costas D. Koutras and Yorgos Zikos. Relating truth, knowledge and belief in epistemic states. In Liu [Liu11], pages 374–385.

[KZ15] C. D. Koutras and Y. Zikos. A note on the properties of **S4.2**. Technical report, Graduate Programme in Logic, Algorithms and Computation, December 2015.

[Len79] W. Lenzen. Epistemologische Betrachtungen zu [S4,S5]. *Erkenntnis*, 14:33–56, 1979.

[Liu11] Weiru Liu, editor. *Symbolic and Quantitative Approaches to Reasoning with Uncertainty - 11th European Conference, ECSQARU 2011, Belfast, UK, June 29-July 1, 2011. Proceedings*, volume 6717 of *Lecture Notes in Computer Science*. Springer, 2011.

[MT93] V. W. Marek and M. Truszczyński. *Nonmonotonic Logic: Context-dependent Reasoning*. Springer-Verlag, 1993.

[Seg71] K. Segerberg. *An essay in Clasical Modal Logic*. Filosofiska Studies, Uppsala, 1971.

[Sta93] R. Stalnaker. A note on non-monotonic modal logic. *Artificial Intelligence*, 64:183–196, 1993. Revised version of the unpublished note originally circulated in 1980.

[Sta06] R. Stalnaker. On logics of knowledge and belief. *Philosophical Studies*, 128(1):169–199, 2006.

[vB10] J. van Benthem. *Modal Logic for Open Minds*. CSLI Publications, 2010.

[vDvdHK07] H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*. Springer, 2007.

[Woo09] M. Wooldridge. *An Introduction to MultiAgent Systems*. John Wiley & Sons, 2009.