Stable belief sets, revisited

Costas D. Koutras
Department of Computer Science and Technology,
University of Peloponnese,
end of Karaiskaki Street, 22100 Tripolis, Greece
ckoutras@uop.gr

Yorgos Zikos
Graduate Programme in Logic, Algorithms and Computation,
M.P.L.A., Dept. of Mathematics,
University of Athens
Panepistimioupolis, 157 84 Ilissia, Greece
zikos@sch.gr

May 8, 2010

Abstract

Stable belief sets were introduced by R. Stalnaker in the early '80s, as a formal representation of the epistemic state for an ideal introspective agent. This notion motivated Moore's *autoepistemic logic* and greatly influenced *modal nonmonotonic reasoning*. Stalnaker stable sets possess an undoubtly simple and intuitive definition and can be elegantly characterized in terms of **S5** universal models or **KD45** situations. However, they do model an extremely perfect introspective reasoner and suffer from a KR version of the *logical omniscience problem*. In this paper, we vary the context rules underlying the positive and/or negative introspection conditions in the original definition of R. Stalnaker, to obtain variant notions of a stable epistemic state, which appear to be more plausible under the epistemic viewpoint. For these alternative notions of stable belief set, we obtain *representation theorems* using *possible world models* with *non-normal (impossible) worlds* and *neighborhood modal models*. En route, we identify some modal axioms which appear to be of some interest in KR and develop the proof theory of some regular and classical modal logics with a notion of strong provability. This stream of research resembles the questions posed and (partly) settled in classical (monotonic) epistemic reasoning about logical omniscience, now examined under the perspective of Knowledge Representation.

1 Introduction

Classical epistemic reasoning has been born and bred within the realm of Philosophical Logic and always had a modal flavour, already from its early inception in Hintikka's seminal work [Hin62]. The *epistemic/doxastic logic* stream of research was very active for more than two decades and mainly revolved around constructing and discussing axiomatic systems which accurately describe the phenomena of *knowledge* and *belief*, from the perspective of a philosopher 'externally' reasoning about other entities' knowledge [Len79]. Many axiomatic systems have been proposed and several problems around this axiomatic approach to knowledge and belief have been identified and discussed (see [Len78, HM92]); in more recent years, epistemic and doxastic modal logics have found important applications in Knowledge Representation and Computer Science [FHMV03].

AI has created a completely new battlefield for epistemic reasoning, through the attempts to construct *nonmonotonic logics* in Knowledge Representation. The perspective of KR is much different, as the objective now is to describe 'internally' the epistemic capabilities of an intelligent agent reasoning on his/her own beliefs. The use of modal languages and the import of techniques from classical epistemic reasoning have been employed from as early as the beginning of the '80s, when nonmonotonic logics have been announced. Modal nonmonotonic reasoning has been introduced through the work of D. McDermott and J. Doyle [MD80], with the use of a fixpoint construction which has been seriously criticized initially. *Stable belief sets* were introduced by R. Stalnaker at the same time; the short note [Sta93] was written as a commentary on modal nonmonotonic logic and proposed the notion of a *stable set of beliefs* as a formal representation of the epistemic state of an ideally rational agent, with full introspective capabilities. Assuming a propositional language, endowed with a modal operator $\Box\varphi$, interpreted as '*φ is believed*', a set of formulas S 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 $\neg\Box\varphi \in S$

This notion proved to be of major importance in nonmonotonic modal logics. According to [Sta93], R. Moore has written that this notion ' .. *was a very important influence on the development of autoepistemic logic*' [Moo85]; it also played a role in the logical investigations of Marek, Schwarz and Truszczyński on the McDermott & Doyle family of modal nonmonotonic modal logics [MT93]. Actually, the definition of stable sets was the first important step towards the idea of constructing epistemic logic(s) in nonmonotonic reasoning, without any appeal to classical modal logic (known as the '*Modality Si, Modal Logic No!*' motto of J. McCarthy).

The syntactic definition of stable sets is very natural and intuitive. Further research quickly revealed that they possess interesting properties while they do also admit simple and elegant semantic characterizations: they can be represented as the theories of universal (**S5**) Kripke models, or alternatively, as the set of beliefs of an agent residing in a **KD45** situation (see [MT93, Chapt.8], [Hal97a]).

It is not hard to see however, that Stalnaker's stable sets model an extremely perfect reasoner. In a sense, the situation is reminiscent of the '*logical omniscience*' problem in classical epistemic logic: normal modal logics of knowledge describe a reasoner who knows all the logical consequences of his/her beliefs; more on this, in section 2. Actually, the situation in Stalnaker's stable sets is a bit more uncomfortable: all tautologies are known and a stable set is a theory maximally consistent with provability in **S5**. This raises some important philosophical and technical questions in modal non-monotonic reasoning, observed in [Hal97a] and addressed from a fine viewpoint in the work of Marek, Schwarz and Truszczyński [MST93].

So, stable sets are defined by calling for closure under (classical propositional logic and) suitable context rules, intended to capture positive and negative introspection on self beliefs. They are characterized by (and represented as theories of) well-known epistemic possible-worlds models, which have emerged in logics of classical epistemic reasoning (**S5**, **KD45**). It is absolutely natural to investigate whether one can define in a natural way, variants of this notion which represent a less ideal and less omniscient agent, while retaining some of their interesting and useful properties; in this direction it is interesting from the KR viewpoint to work on the following two questions, related to the interplay between syntax and semantics of stable epistemic states:

- can we weaken the positive and/or negative introspection conditions (seen henceforth as context-dependent rules) in Stalnaker's original definition and still obtain a plausible (and perhaps, more pragmatic) notion of stable epistemic state? For such an emerging notion, does there exist a good model-theoretic representation?

- can we suitably replace **S5** and **KD45** in the semantic characterization of stable sets, with a possible-worlds model (possibly with *non-normal worlds* or a *neighborhood model*) determining some other classical modal logic and prove that the emerging notion of an epistemic state admits a syntactic definition in terms of (closure under) natural positive and negative introspection conditions?

In this paper, we work on the first of these two questions, actually the most important from the KR viewpoint. We vary conditions (ii) and (iii) in Stalnaker's definition to obtain three weaker notions of an epistemic state. We obtain semantic characterizations for the notions of stable sets we define; not surprisingly, we have to employ *impossible worlds* and *neighborhood modal models*. In Section 2 we gather the necessary technical background needed for our results, establishing notation and terminology. In Section 3 we very briefly mention some results we have obtained on the *determination* of *classical* and *regular modal logics*, with a notion of *strong provability* from premises. These results are later used for obtaining our representation theorems. Sections 4 and 5 form the core of our results: we define, examine and characterize weaker notions of a stable epistemic state. In Section 6 we comment on related work and discuss open questions for future research.

2 Background Material

In this section we gather the necessary background material and results. For the basics of Modal Logic the reader is refered to the books [BdRV01, Che80, HC96] and for the essentials of modal nonmonotonic logics to [MT93]. We assume a modal propositional language $\mathcal{L}_\Box$, endowed with an epistemic operator $\Box\varphi$, read as '*it is believed that* φ *holds*'. Sentence symbols include $\top$ (for *truth*) and $\bot$ (for *falsity*).

Some of the important axioms in epistemic/doxastic logic are:

K. $(\Box\varphi \wedge \Box(\varphi \supset \psi)) \supset \Box\psi$ [1]

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

D. $\Box\varphi \supset \neg\Box\neg\varphi$ or $\neg(\Box\varphi \wedge \Box\neg\varphi)$ (consistent belief)

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

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

w5. $(\varphi \wedge \neg\Box\varphi) \supset \Box\neg\Box\varphi$ (weak negative introspection)

p5. $(\neg\Box\varphi \wedge \neg\Box\neg\varphi) \supset \Box\neg\Box\varphi$ (weak negative introspection)

Modal logics are sets of modal formulae containing classical propositional logic (i.e. containing all tautologies in the augmented language $\mathcal{L}_\Box$) and closed under rule **MP**.$\frac{\varphi, \varphi\supset\psi}{\psi}$. The smallest modal logic is denoted as **PC** (propositional calculus in the augmented language). A set T of formulae is called *consistent* iff $(\forall n \in \mathbb{N}, \forall \varphi_0, \ldots, \varphi_n \in T)$ $\varphi_0 \wedge \ldots \wedge \varphi_n \supset \bot \notin \mathbf{PC}$; otherwise, T is called *inconsistent*. *Normal* are called those modal logics, which contain all instances of axiom **K** and are closed under rule

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

By $\mathbf{KA_1 \ldots A_n}$ 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 **KD45** (a *logic of consistent belief*).

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 W and a binary relation between them $R \subseteq W \times W$: whenever wRv, we say that world w '*sees*' world v. 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: $(\mathfrak{M}, w \Vdash \Box\varphi$ iff $(\forall v \in W)(wRv \Rightarrow \mathfrak{M}, v \Vdash \varphi))$.

A logic Λ is *determined* by a class of models iff it is *sound* and *complete* with respect to this class; it is known that **S5** is determined by the class of Kripke models with a *universal* relation, while **KD45** is determined by the class of models where each world

[1] In our notation **K** is the axiomatic scheme $(\Box\varphi \wedge \Box(\varphi \supset \psi)) \supset \Box\psi$ i.e. $\mathbf{K} = US\big((\Box p \wedge \Box(p \supset q)) \supset \Box q\big)$, where $US(\varphi)$ is the set of all universal substitution instances of φ.

w '*sees*' a '*cluster*' (i.e. a universally connected subset) of worlds; every model of this class has the form $\langle \{w\} \cup W, (\{w\} \cup W) \times W, V \rangle$).

Normal modal epistemic logics suffer from the so-called *logical omniscience* problem, which can be attributed to axiom **K** and rule **RN**. Because of the latter, all tautologies are known. Also, because of the axiom **K**, logical consequences of knowledge constitute knowledge, something unreasonable in realistic situations. Note however that axiom **K** and axioms as simple as **N**.$\Box\top$ are unavoidable in Kripke models and ubiquitous in normal modal logics.

A first step towards solving the logical omniscience problem is by defining *regular* modal logics which contain **K**, but substitute rule **RN** for rule

$$\textbf{RM}.\ \frac{\varphi \supset \psi}{\Box\varphi \supset \Box\psi}$$

We denote by $\mathbf{KA_1 \ldots A_{nR}}$ the regular modal logic axiomatized by axioms $\mathbf{A_1}$ to $\mathbf{A_n}$. Regular modal logics are interpreted on a strange species of possible world models, introduced by Kripke too; we will call them *q-models* here ($\mathfrak{M} = \langle W, N, R, V \rangle$). We now have two kinds of worlds: *normal* worlds (N), which behave in the way we described above and *non-normal* (also called *queer* or *impossible*) worlds ($W \setminus N$), where nothing is known/believed ($\Box\varphi$ is *never* true there) and everything is consistent to our state of affairs ($\neg\Box\neg\varphi$ is *always* true there). Within a world w, the propositional connectives are interpreted classically and $\Box\varphi$ is true at w iff $w \in N$ and $(\forall v \in W)(wRv \Rightarrow \mathfrak{M}, v \Vdash \varphi)$).

This however does not avoid the effect of **K**: to be able to eliminate **K** we have to resort to *neighborhood* (also called *Montague* or *minimal* in [Che80]) semantics. In this kind of models, which we will call *n-models*, each world does not 'see' other worlds but it is associated to possible '*neighborhoods*' (subsets) of possible worlds: an n-model is a triple $\mathfrak{N} = \langle W, E, V \rangle$, where W is any set of worlds, E is any function assigning to any world, its sets of 'neighboring' worlds (i.e. $E : W \to \mathcal{P}(\mathcal{P}(W))$) and V is again a valuation. The interpretation of any formula is exactly as in Kripke models, except of the formulas of the form $\Box\varphi$; such a formula is true at w iff the set of worlds where φ holds, belong to the possible neighborhoods of w: $\overline{V}(\varphi) = \{v \in W \mid \mathfrak{N}, v \Vdash \varphi\} \in E(w)$. *Theory* of a (Kripke, q- or n-) model $\mathfrak{M}$ (denoted as $Th(\mathfrak{M})$) is the set of all formulae being true in every world of $\mathfrak{M}$.

Having a q-model, we can define a pointwise equivalent n-model:

Definition 2.1 *Let* $\mathfrak{M} = \langle W, N, R, V \rangle$ *be a q-model and* $\mathfrak{N}_{\mathfrak{M}} = \langle W, E, V \rangle$ *the n-model, where* $E(w) = \{X \subseteq W \mid R_w \subseteq X\}$[2]*, if* $w \in N$*, and* $E(w) = \varnothing$*, if* $w \in W \setminus N$*.* $\mathfrak{N}_{\mathfrak{M}}$ *is called* the equivalent n-model produced by $\mathfrak{M}$*.*

This notion of 'equivalence' seems to be appropriate, because of the following result:

Proposition 2.2 *Let* $\mathfrak{M} = \langle W, N, R, V \rangle$ *be a Kripke q-model. Then*

$$(\forall\varphi \in \mathcal{L}_\Box)(\forall w \in W)(\mathfrak{M}, w \Vdash \varphi \iff \mathfrak{N}_{\mathfrak{M}}, w \Vdash \varphi)$$

[2] $R_w = \{v \in W \mid wRv\}$

PROOF. By induction on the complexity of φ. Induction base and boolean cases of induction step are obvious. So, let us focus on $\Box\varphi$.

$$\begin{aligned}
\mathfrak{M}, w \Vdash \Box\varphi &\iff w \in N \wedge (\forall v \in W)(wRv \Rightarrow \mathfrak{M}, v \Vdash \varphi) \\
&\overset{Ind.Hyp.}{\iff} w \in N \wedge (\forall v \in W)(v \in R_w \Rightarrow \mathfrak{N}_{\mathfrak{M}}, v \Vdash \varphi) \\
&\iff w \in N \wedge R_w \subseteq \overline{V}(\varphi) \\
&\overset{Def.2.1}{\iff} \overline{V}(\varphi) \in E(w) \\
&\iff \mathfrak{N}_{\mathfrak{M}}, w \Vdash \Box\varphi
\end{aligned}$$

■

The directed graph $\mathfrak{F} = \langle W, R \rangle$, underlying a (Kripke, q-, or n-) model, is called a *frame*. A modal logic Λ is called *classical* iff it is closed under the rule

$$\textbf{RE}. \ \frac{\varphi \equiv \psi}{\Box\varphi \equiv \Box\psi}$$

See [Che80] for results on the characterization of classical modal logics in terms of Montague semantics. By $\mathbf{A_1} \ldots \mathbf{A_{nC}}$ we denote the classical modal logic axiomatized by axioms $\mathbf{A_1}$ to $\mathbf{A_n}$.

It is convenient in our paper to consider the following context-dependent versions of the modal rules mentioned up to this point: assuming a set S of modal formulae, we denote the rules

$$\mathbf{RN_c}. \ \frac{\varphi \in S}{\Box\varphi \in S} \qquad \mathbf{NI_c}. \ \frac{\varphi \notin S}{\neg\Box\varphi \in S}$$

$$\mathbf{RM_c}. \ \frac{\varphi \supset \psi \in S}{\Box\varphi \supset \Box\psi \in S} \qquad \mathbf{RE_c}. \ \frac{\varphi \equiv \psi \in S}{\Box\varphi \equiv \Box\psi \in S}$$

Stalnaker stable sets are closed under propositional reasoning (i), under rule $\mathbf{RN_c}$ (ii) and rule $\mathbf{NI_c}$ (iii). The following theorem gathers some of their useful properties; see [MT93] for a proof.

Theorem 2.3

(i) If a set S is stable, then it is closed under strong **S5** *provability. In particular, it contains every instance of* **K**, **T**, **4**, *and* **5**.

(ii) A set S is stable iff it is the theory of a Kripke model with a universal accessibility relation.

(iii) A set S is stable iff it is the set of formulae believed in a world w of a **KD45**-*model, i.e. S is stable iff there is a* **KD45**-*model $\mathfrak{M} = \langle W, R, V \rangle$ and $(\exists w \in W)$ $S = \{\varphi \in \mathcal{L}_\Box \mid \mathfrak{M}, w \Vdash \Box\varphi\}$.*

3 A digression: Regular and Classical Modal Logics

To be able to characterize the stable sets introduced in the subsequent sections, we have to work on the proof theory of regular and classical modal logics with a notion of strong provability from premises. The results are original, in the sense that they have not been developed elsewhere, yet they are quite lengthy to be included in this extended abstract and they are left for the full paper.

Regular modal logics Firstly, we employ the axioms:

$\mathbf{4}_\top$. $\Box\varphi \supset \Box(\Box\top \supset \Box\varphi)$

$\mathbf{B}_\top$. $(\varphi \wedge \Box\top) \supset \Box\neg\Box\neg\varphi$

$\mathbf{5}_\top$. $\neg\Box\varphi \wedge \Box\top \supset \Box\neg\Box\varphi$

The first of them appears in [Seg71] and all of them seem useful in our KR investigations. Furthermore, for a q-frame $\mathfrak{F} = \langle W, N, R\rangle$, we employ following property:

$(\mathsf{U_q})$ $(\forall w \in N)(\forall v \in W)wRv$

The notion of strong provabillity from premises in a regular modal logic is defined as usual.

Definition 3.1 *If* $\{A_0, \ldots, A_n\} \subseteq \mathcal{L}_\Box$ *is a set of axioms of regular modal logic* Λ *(i.e.* $\Lambda = \mathbf{KA_0} \ldots \mathbf{A_{nR}}$ *is the smallest regular modal logic containing* $A_0, \ldots, A_n$*) and* $I \subseteq \mathcal{L}_\Box$ *is a set of premises, then for any formula* φ *we say that there is an* RM-proof of φ from premises I in Λ $(I \vdash_\Lambda \varphi)$ *iff there is a Hilbert-style proof, where each step of the proof is either a formula in* $\mathbf{PC} \cup \mathbf{K} \cup US(A_0) \cup \ldots \cup US(A_n) \cup I$ *or a result of applying* **MP** *or* **RM** *to formulas of previous steps and the last formula in this proof is* φ*.*

The consistency of theories is also defined as usual.

Definition 3.2 *A theory* $I \subseteq \mathcal{L}_\Box$ *is called* consistent with regular modal logic Λ *(abbr. c*Λ*-theory) iff* $I \nvdash_\Lambda \bot$*; otherwise,* I *is called* inconsistent with Λ *(abbr. inc*Λ*-theory).*
Supposed that I *is a c*Λ*-theory, a set of formulae* T *is called* I-consistent with Λ *(abbr. Ic*Λ*-theory) iff* $(\forall n \in \mathbb{N}, \forall \varphi_0, \ldots, \varphi_n \in T)$ $I \nvdash_\Lambda \varphi_0 \wedge \ldots \wedge \varphi_n \supset \bot$*; otherwise,* T *is called* I-inconsistent with Λ *(abbr. Iinc*Λ*-theory).*
T *is called* maximal I-consistent with Λ *(abbr. mIc*Λ*-theory) iff* T *is* $Ic\Lambda$ *and* $(\forall \psi \notin T)$ $T \cup \{\psi\}$ *is* $Iinc\Lambda$*.*

Following lemma contains useful properties for maximal consistent theories.

Lemma 3.3 *Let* I *be a c*Λ*-theory and* Γ *a mIc*Λ*-theory. Then*

(i) Γ *is closed under (MP)*

(ii) $(\forall\varphi \in \mathcal{L}_\Box)(\varphi \in \Gamma$ *or* $\neg\varphi \in \Gamma)$

(iii) $(\forall\varphi \in \mathcal{L}_\Box)(I \vdash_\Lambda \varphi \Rightarrow \varphi \in \Gamma)$

(iv) $(\forall\varphi \in \mathcal{L}_\Box)(\varphi \wedge \psi \in \Gamma \Leftrightarrow (\varphi \in \Gamma \textit{ and } \psi \in \Gamma))$

PROOF. Consider any $\varphi, \psi \in \mathcal{L}_\Box$.
(i)
Suppose that $\varphi, \varphi \supset \psi \in \Gamma$ and, for the sake of contradiction, that $\psi \notin \Gamma$. Then, $\Gamma \cup \{\psi\}$ would be an $Iinc\Lambda$-theory, i.e. there are $n \geq 0$ and $\varphi_1, \ldots, \varphi_n \in \Gamma$ s.t. $I \vdash_\Lambda \varphi_1 \wedge \ldots \wedge \varphi_n \wedge \psi \supset \bot$, hence, $I \vdash_\Lambda \varphi_1 \wedge \ldots \wedge \varphi_n \wedge \varphi \wedge (\varphi \supset \psi) \supset \bot$, i.e. Γ is $Iinc\Lambda$, which is a contradiction.
(ii)
Suppose, for the sake of contradiction, that $\varphi, \neg\varphi \notin \Gamma$. Then, $\Gamma \cup \{\varphi\}$, $\Gamma \cup \{\neg\varphi\}$, would be both $Iinc\Lambda$-theories, i.e. there are $n \geq 0$, $m \geq 0$ and $\varphi_1, \ldots, \varphi_n \in \Gamma$ s.t. $I \vdash_\Lambda \varphi_1 \wedge \ldots \wedge \varphi_n \wedge \varphi \supset \bot$ and $\psi_1, \ldots, \psi_m \in \Gamma$ s.t. $I \vdash_\Lambda \psi_1 \wedge \ldots \wedge \psi_m \wedge \neg\varphi \supset \bot$.
– If $n > 0$ or $m > 0$, then $I \vdash_\Lambda \varphi_1 \wedge \ldots \wedge \varphi_n \wedge \psi_1 \wedge \ldots \wedge \psi_m \supset \neg\varphi \wedge \varphi$, therefore, $I \vdash_\Lambda \varphi_1 \wedge \ldots \wedge \varphi_n \wedge \psi_1 \wedge \ldots \wedge \psi_m \supset \bot$, so Γ is $Iinc\Lambda$, which is a contradiction.
– If $n = 0$ and $m = 0$, then $I \vdash_\Lambda \varphi \wedge \neg\varphi$, i.e. $I \vdash_\Lambda \bot$, so I is $inc\Lambda$, which is again a contradiction.
(iii)
Suppose that $I \vdash_\Lambda \varphi$ and, for the sake of contradiction, that $\varphi \notin \Gamma$. Then, $\Gamma \cup \{\varphi\}$ would be an $Iinc\Lambda$-theory, i.e. there are $n \geq 0$ and $\varphi_1, \ldots, \varphi_n \in \Gamma$ s.t. $I \vdash_\Lambda \varphi_1 \wedge \ldots \wedge \varphi_n \wedge \varphi \supset \bot$.
– If $n > 0$, then $I \vdash_\Lambda \varphi_1 \wedge \ldots \wedge \varphi_n \supset \neg\varphi$, and, since $I \vdash_\Lambda \varphi$, $I \vdash_\Lambda \varphi_1 \wedge \ldots \wedge \varphi_n \supset \bot$, i.e. Γ is $Iinc\Lambda$, which is a contradiction.
– If $n = 0$, then $I \vdash_\Lambda \varphi \supset \bot$, and, since $I \vdash_\Lambda \varphi$, $I \vdash_\Lambda \bot$, so I is $inc\Lambda$, which is again a contradiction.
(iv)
$(\Rightarrow)$ Suppose that $\varphi \wedge \psi \in \Gamma$. Since $I \vdash_\Lambda \varphi \wedge \psi \supset \varphi$, by (iii), $\varphi \wedge \psi \supset \varphi \in \Gamma$, hence, by (i), $\varphi \in \Gamma$. In exactly the same way, it can be proved that $\psi \in \Gamma$.
$(\Leftarrow)$ Suppose that $\varphi, \psi \in \Gamma$. Since $I \vdash_\Lambda \varphi \supset (\psi \supset \varphi \wedge \psi)$, by (iii), $\varphi \supset (\psi \supset \varphi \wedge \psi) \in \Gamma$, hence, by (i), $\varphi \wedge \psi \in \Gamma$. ∎

Aiming to construct a model, whose theory contains exactly all formulae, which can be proved from I in Λ, we firstly prove following lemmata:

Lemma 3.4 *Let I be a $c\Lambda$-theory. Then, there exist a nonempty $Ic\Lambda$-theory.*

PROOF. Since I is $c\Lambda$, there is a $\varphi \in \mathcal{L}_\Box$ s.t. $I \nvdash_\Lambda \varphi$. Hence, $\{\neg\varphi\}$ is $Ic\Lambda$. ∎

Lemma 3.5 (Lindenbaum) *Let I be a $c\Lambda$-theory and T an $Ic\Lambda$-theory. Then, there is a $mIc\Lambda$ theory Γ s.t. $T \subseteq \Gamma$.*

PROOF. Since the infinite set Φ of propositional variables of our language $\mathcal{L}_\Box$ is countable, there is an enumeration $\varphi_0, \varphi_1, \varphi_2, \ldots$ of $\mathcal{L}_\Box$. Now, let us define recursively following sequence of sets

$$\begin{aligned} T_0 &= T \\ T_{n+1} &= \begin{cases} T_n \cup \{\varphi_n\} & \text{if } T_n \cup \{\varphi_n\} \text{ is } Ic\Lambda \\ T_n \cup \{\neg\varphi_n\} & \text{otherwise} \end{cases} \end{aligned}$$

(a)
Firstly, we will prove by induction on n, that $(\forall n \in \mathbb{N})(T_n \text{ is } Ic\Lambda)$. It suffices to show (in the ind. step) that if $T_n \cup \{\varphi_n\}$ is $Iinc\Lambda$, then $T_n \cup \{\neg\varphi_n\}$ is $Ic\Lambda$. So, if $T_n \cup \{\varphi_n\}$ is $Iinc\Lambda$, then there are $m \geq 0$ and $\psi_1,\ldots,\psi_m \in T_n$ s.t. $I \vdash_\Lambda \psi_1 \wedge \ldots \wedge \psi_m \wedge \varphi_n \supset \bot$
(if $I \vdash_\Lambda \psi_1 \wedge \ldots \wedge \psi_m \supset \bot$, then T_n would be $Iinc\Lambda$, which is contradictory to ind. hypothesis, hence, φ_n must appear in the conjunction). Now, suppose, for the sake of contradiction, that $T_n \cup \{\neg\varphi_n\}$ were $Iinc\Lambda$. Then, there would be $p \geq 0$ and $\chi_1,\ldots,\chi_p \in T_n$ s.t. $I \vdash_\Lambda \chi_1 \wedge \ldots \wedge \chi_p \wedge \neg\varphi_n \supset \bot$ (as before, $\neg\varphi_n$ must appear in the conjunction).
– if $m > 0$ or $p > 0$, then $I \vdash_\Lambda \psi_1 \wedge \ldots \wedge \psi_m \wedge \chi_1 \wedge \ldots \wedge \chi_p \supset \bot$, i.e. T_n is $Iinc\Lambda$, which is a contradiction, by ind. hypothesis.
– if $m = 0$ and $p = 0$, then $I \vdash_\Lambda \neg\varphi_n$ and $I \vdash_\Lambda \varphi_n$, hence, $I \vdash_\Lambda \bot$, i.e. I is $inc\Lambda$, which is also a contradiction.
(b)
It can be proved, by a trivial induction, that $(\forall i, j \in \mathbb{N})(i \leq j \Rightarrow T_i \subseteq T_j)$
(c)
Now, let us define $\Gamma = \bigcup_{n\in\mathbb{N}} T_n$. Suppose, for the sake of contradiction, that Γ is $Iinc\Lambda$, i.e. there are $m \geq 0$ and $\psi_0, \ldots, \psi_m \in \Gamma$ s.t. $I \vdash_\Lambda \psi_0 \wedge \ldots \wedge \psi_m \supset \bot$. Since $\psi_0, \ldots, \psi_m$ appear in the enumeration of $\mathcal{L}_\Box$, there must be $k_0, \ldots, k_m \in \mathbb{N}$ s.t. $\varphi_{k_0} = \psi_0, \ldots, \varphi_{k_m} = \psi_m$. Furthermore, since $\varphi_{k_0}, \ldots, \varphi_{k_m} \in \Gamma$, all $T_{k_0} \cup \{\varphi_{k_0}\}, \ldots, T_{k_m} \cup \{\varphi_{k_m}\}$ are $Ic\Lambda$ and $\varphi_{k_0} \in T_{k_0+1}, \ldots, \varphi_{k_m} \in T_{k_m+1}$, hence, by (b), $\varphi_{k_0}, \ldots, \varphi_{k_m} \in T_{\max\{k_0,\ldots,k_m\}+1}$, consequently, $T_{\max\{k_0,\ldots,k_m\}+1}$ is $Iinc\Lambda$, which is a contradiction, by (a).
(d)
Let now $\varphi \in \mathcal{L}_\Box \setminus \Gamma$. Since φ appears in the enumeration of $\mathcal{L}_\Box$, there must be a $k \in \mathbb{N}$ s.t. $\varphi_k = \varphi$. Then, since $\varphi \notin \Gamma$, $T_k \cup \{\varphi_k\}$ is $Iinc\Lambda$ and $\neg\varphi \in T_{k+1}$, so $\neg\varphi \in \Gamma$. But then, since $I \vdash_\Lambda \varphi \wedge \neg\varphi \supset \bot$, $\Gamma \cup \{\varphi\}$ is $Iinc\Lambda$.
So, it has been proved that $T = T_0 \subseteq \Gamma$ and, by (c), (d), Γ is a $mIc\Lambda$-theory. ■

Last two lemmata do guarantee that the model defined next, does exist.

Definition 3.6 *Let Λ be any regular modal logic and I be any $c\Lambda$-theory. The* canonical model $\mathfrak{M}^{\Lambda,I}$ *for Λ and I is the Kripke q-model, which is defined as the quadruple $\langle W^{\Lambda,I}, N^{\Lambda,I}, R^{\Lambda,I}, V^{\Lambda,I}\rangle$, where:*

(i) $W^{\Lambda,I} = \{\mathsf{\Gamma} \subseteq \mathcal{L}_\Box \mid \mathsf{\Gamma} : mIc\Lambda\}$

(ii) $N^{\Lambda,I} = \{\mathsf{\Gamma} \in W^{\Lambda,I} \mid \Box\top \in \mathsf{\Gamma}\}$

(iii) $(\forall \mathsf{\Gamma}, \mathsf{\Delta} \in W^{\Lambda,I})(\mathsf{\Gamma} R^{\Lambda,I} \mathsf{\Delta}$ *iff* $(\forall \varphi \in \mathcal{L}_\Box)(\Box\varphi \in \mathsf{\Gamma} \Rightarrow \varphi \in \mathsf{\Delta}))$

(iv) $(\forall p \in \Phi)(V^{\Lambda,I}(p) = \{\mathsf{\Gamma} \in W^{\Lambda,I} \mid p \in \mathsf{\Gamma}\})$

Frame $\mathfrak{F}^{\mathbf{\Lambda,I}} = \langle W^{\Lambda,I}, N^{\Lambda,I}, R^{\Lambda,I}\rangle$ underlying $\mathfrak{M}^{\mathbf{\Lambda,I}}$ is called the canonical frame *for Λ and I.*

In a case of a normal modal logic Λ, $\Box\top \in \Lambda$. Hence, every proof $I \vdash_\Lambda \varphi$ is equivalent to a proof using **RN** instead of **RM** and vice versa. Furthermore, by Lem.3.3(iii), $(\forall \mathsf{\Gamma} \in W^{\Lambda,I})\ \Box\top \in \mathsf{\Gamma}$, hence, by Def.3.6(ii), $N^{\Lambda,I} = W^{\Lambda,I}$. So,

Fact 3.7 *If Λ is a normal modal logic, then $N^{\Lambda,I} = W^{\Lambda,I}$ and $\mathfrak{M}^{\Lambda,I}$ coincides with the canonical model defined for normal modal logics (and the corresponding $c\Lambda$-theories) in bibliography.*

Now, we come to the key-lemma towards proving that the theory of $\mathfrak{M}^{\Lambda,I}$ contains exactly all formulae, which can be proved from I in Λ:

Lemma 3.8 (Truth Lemma) *Let Λ be a regular modal logic and I be a $c\Lambda$-theory. Then,* $(\forall\varphi \in \mathcal{L}_\Box)(\forall\Gamma \in W^{\Lambda,I})(\mathfrak{M}^{\Lambda,I}, \Gamma \Vdash \varphi \iff \varphi \in \Gamma)$

Proof. By induction on the complexity of φ. Induction base follows from Def.3.6(iv) and $\varphi \supset \psi$ – part of induction step follows immediatelly from induction hypothesis using (i) to (iv) of Lem.3.3. Now, to the $\Box\varphi$ – case.
$\mathfrak{M}^{\Lambda,I}, \Gamma \Vdash \Box\varphi$ iff $(\forall\Delta \in W^{\Lambda,I})(\Gamma R^{\Lambda,I}\Delta \Rightarrow \mathfrak{M}^{\Lambda,I}\Delta \Vdash \varphi) \wedge \Gamma \in N^{\Lambda,I}$ iff (by Ind.Hyp.) $(\forall\Delta \in W^{\Lambda,I})(\Gamma R^{\Lambda,I}\Delta \Rightarrow \varphi \in \Delta) \wedge \Gamma \in N^{\Lambda,I}$
It suffices to show that this is equivalent to the fact that $\Box\varphi \in \Gamma$.
$(\Rightarrow)$
Suppose that $\Box\varphi \notin \Gamma$ and $\Gamma \in N^{\Lambda,I}$. Since Γ is a $mIc\Lambda$-theory, by Lem.3.3(ii), $\neg\Box\varphi \in \Gamma$. Now, let us define $\Delta = \{\psi \in \mathcal{L}_\Box \mid \Box\psi \in \Gamma\}$ and $\Theta = \{\neg\varphi\} \cup \Delta$. Suppose, for the sake of contradiction, that Θ is $Iinc\Lambda$ i.e. there exist $\psi_1, \ldots, \psi_n \in \Theta$ s.t. $I \vdash_\Lambda \psi_1 \wedge \ldots \wedge \psi_n \supset \bot$.

- if $n = 1$ and $\psi_1 = \neg\varphi$ i.e. $I \vdash_\Lambda \neg\varphi \supset \bot$, then $I \vdash_\Lambda \top \supset \varphi$, and, by (RM), $I \vdash_\Lambda \Box\top \supset \Box\varphi$. Then, by Lem.3.3(iii), $\Box\top \supset \Box\varphi \in \Gamma$. But $\Gamma \in N^{\Lambda,I}$, so, by Def.3.6(ii) and Lem.3.3(i), $\Box\varphi \in \Gamma$, which is a contradiction, since $\neg\Box\varphi \in \Gamma$ and Γ is an $Ic\Lambda$-theory.

- if $\psi_1, \ldots, \psi_n \in \Delta$, then $I \vdash_\Lambda \psi_1 \wedge \ldots \wedge \psi_n \supset \phi$, since $\bot \supset \varphi \in$ **PC**.
  if $n > 1$ and $\psi_1, \ldots, \psi_{n-1} \in \Delta$ and $\psi_n = \neg\varphi$, then $I \vdash_\Lambda \psi_1 \wedge \ldots \wedge \psi_{n-1} \supset \phi$.
  So, in both cases, there are $\psi_1, \ldots, \psi_n \in \Delta$ with $n \geq 1$ s.t. $I \vdash_\Lambda \psi_1 \wedge \ldots \wedge \psi_n \supset \varphi$.
  Hence, by **RM**, $I \vdash_\Lambda \Box(\psi_1 \wedge \ldots \wedge \psi_n) \supset \Box\varphi$ (1)
  But, $I \vdash_\Lambda \psi_1 \supset (\psi_2 \supset \psi_1 \wedge \psi_2)$, so, by **RM**, $I \vdash_\Lambda \Box\psi_1 \supset \Box(\psi_2 \supset \psi_1 \wedge \psi_2)$, and, by **K**, $I \vdash_\Lambda \Box\psi_1 \supset (\Box\psi_2 \supset \Box(\psi_1 \wedge \psi_2))$ i.e. $I \vdash_\Lambda \Box\psi_1 \wedge \Box\psi_2 \supset \Box(\psi_1 \wedge \psi_2)$. Hence, by a trivial induction, $I \vdash_\Lambda \Box\psi_1 \wedge \ldots \wedge \Box\psi_n \supset \Box(\psi_1 \wedge \ldots \wedge \psi_n)$, and by (1), $I \vdash_\Lambda \Box\psi_1 \wedge \ldots \wedge \Box\psi_n \supset \Box\varphi$, so, by Lem.3.3(iii), $\Box\psi_1 \wedge \ldots \wedge \Box\psi_n \supset \Box\varphi \in \Gamma$ (2)
  But, since $\psi_1, \ldots, \psi_n \in \Delta$, $\Box\psi_1, \ldots, \Box\psi_n \in \Gamma$, therefore, by Lem.3.3(iv), $\Box\psi_1 \wedge \ldots \wedge \Box\psi_n \in \Gamma$, and finally, by (2) and Lem.3.3(i), $\Box\varphi \in \Gamma$, which is again a contradiction.

So, Θ is an $Ic\Lambda$-theory, and by Lindenbaum's lemma (3.5), there is a $mIc\Lambda$-theory Ξ s.t. $\Theta \subseteq \Xi$. Hence, $\neg\varphi \in \Xi$, which entails, by Lem.3.3(ii), that $\varphi \notin \Xi$.
Furthermore, $(\forall\psi \in \mathcal{L}_\Box)$ if $\Box\psi \in \Gamma$, then $\psi \in \Delta$ i.e. $\psi \in \Theta$ i.e. $\psi \in \Xi$. Therefore, by Def.3.6(iii), $\Gamma R^{\Lambda,I}\Xi$.
So, the contrapositive was proved.
$(\Leftarrow)$
Suppose that $\Box\varphi \in \Gamma$. Then, for any $\Delta \in W^{\Lambda,I}$, if $\Gamma R^{\Lambda,I}\Delta$, then by Def.3.6(iii), $\varphi \in \Delta$. Furthermore, $I \vdash_\Lambda \varphi \supset \top$, hence, by **RM**, $I \vdash_\Lambda \Box\varphi \supset \Box\top$. Then, by Lem.3.3(iii), $\Box\varphi \supset \Box\top \in \Gamma$. But $\Box\varphi \in \Gamma$, so, by Lem.3.3(i), $\Box\top \in \Gamma$, consequently, by Def.3.6(ii), $\Gamma \in N^{\Lambda,I}$. ■

Using Truth Lemma 3.8, we can prove (see Appendix A.1) the following characterization of a useful regular modal logic, namely $\mathbf{S5'_R} = \mathbf{KT4}_\top\mathbf{B}_{\top\mathbf{R}}$

Theorem 3.9 $\mathbf{S5'_R}$ *is strongly complete with respect to all q-frames, for which* $(\mathsf{U_q})$ *holds.*

Actually, the following, more general result can be proved, which will be useful in subsequent sections.

Proposition 3.10 *Let Λ be a regular modal logic and I be a cΛ-theory. Then,*

$$(\forall\varphi \in \mathcal{L}_\Box)(\mathfrak{M}^{\Lambda,I} \Vdash \varphi \iff I \vdash_\Lambda \varphi)$$

PROOF. $(\Rightarrow)$
Suppose that $I \nvdash_\Lambda \varphi$. If $\{\neg\varphi\}$ were $Iinc\Lambda$, then, by definition, $I \vdash_\Lambda \neg\varphi \supset \bot$, which is a contradiction, so $\{\neg\varphi\}$ is a $Ic\Lambda$-theory, and, by Lindenbaum's lemma (3.5), there is a $\mathsf{r} \in W^{\Lambda,I}$ s.t. $\neg\varphi \in \mathsf{r}$. Hence, by Lem.3.8, $\mathfrak{M}^{\Lambda,I}, \mathsf{r} \Vdash \neg\varphi$, so, $\mathfrak{M}^{\Lambda,I} \nVdash \varphi$.
$(\Leftarrow)$
Suppose that $I \vdash_\Lambda \varphi$. Then, by Lem.3.3(iii), $(\forall \mathsf{r} \in W^{\Lambda,I})\ \varphi \in \mathsf{r}$. Hence, by Lem.3.8, $(\forall \mathsf{r} \in W^{\Lambda,I})\ \mathfrak{M}^{\Lambda,I}, \mathsf{r} \Vdash \varphi$, so, $\mathfrak{M}^{\Lambda,I} \Vdash \varphi$. ∎

Classical modal logics Analogously to regular modal logics, the notion of strong provabillity from premises in a classical modal logic is defined.

Definition 3.11 *If $\{A_0, \ldots, A_n\} \subseteq \mathcal{L}_\Box$ is a set of axioms of classical modal logic Λ (i.e. $\Lambda = \mathbf{A_0} \ldots \mathbf{A_{nC}}$ is the smallest classical modal logic containing $A_0, \ldots, A_n$) and $I \subseteq \mathcal{L}_\Box$ is a set of premises, then for any formula φ we say that there is an* RE-proof of φ from premises I in Λ $(I \vdash_\Lambda \varphi)$ *iff there is a Hilbert-style proof, where each step of the proof is either a formula in* $\mathbf{PC} \cup US(A_0) \cup \ldots \cup US(A_n) \cup I$ *or a result of applying* **MP** *or* **RE** *to formulas of previous steps and the last formula in this proof is φ.*

Definition of a theory, which is consistent with a classical modal logic, or I-consistent with a classical modal logic, or maximal I-consistent with a classical modal logic, is exactly as for regular modal logics. In fact, an observation of the proofs of lemmata 3.3, 3.4 and 3.5 (Lindenbaum) reveals that the only requirement for Λ is to be a modal logic. So, they are true for classical modal logics too. Now, let us construct the following model, for which it will be proved afterwards, that its theory contains exactly all formulae, which can be proved (by an RE-proof) from premises in a classical modal logic Λ.

Definition 3.12 *Let Λ be a classical modal logic and I be a cΛ-theory. The* canonical model $\mathfrak{N}^{\Lambda,I}$ *for Λ and I is the n-model, which is defined as the triple $\langle W^{\Lambda,I}, E^{\Lambda,I}, V^{\Lambda,I}\rangle$, where:*
(i) $W^{\Lambda,I} = \{\mathsf{r} \subseteq \mathcal{L}_\Box \mid \mathsf{r} : mIc\Lambda\}$
(ii) $(\forall \mathsf{r} \in W^{\Lambda,I})(\forall\varphi \in \mathcal{L}_\Box)(|\varphi|_{\Lambda,I} \in E^{\Lambda,I}(\mathsf{r}) \iff \Box\varphi \in \mathsf{r})$
where $|\varphi|_{\Lambda,I} = \{\mathsf{r} \in W^{\Lambda,I} \mid \varphi \in \mathsf{r}\}$
(iii) $(\forall p \in \Phi)(V^{\Lambda,I}(p) = \{\mathsf{r} \in W^{\Lambda,I} \mid p \in \mathsf{r}\})$

Again, Lemmata 3.4 and 3.5 guarantee that $W^{\Lambda,I} \neq \varnothing$, but it must be proved that $E^{\Lambda,I}$ in (ii) is well-defined, i.e. that for any $mIc\Lambda$-theory Γ and $(\forall \varphi, \psi \in \mathcal{L}_\Box)$, if $|\varphi|_{\Lambda,I} = |\psi|_{\Lambda,I}$, then $\Box\varphi \in \Gamma \iff \Box\psi \in \Gamma$. This will be established by proving following Lemma.

Lemma 3.13 $|\varphi|_{\Lambda,I} \subseteq |\psi|_{\Lambda,I} \Rightarrow I \vdash_\Lambda \varphi \supset \psi$

Proof. Suppose that $|\varphi|_{\Lambda,I} \subseteq |\psi|_{\Lambda,I}$. Then, for any $mIc\Lambda$-theory Γ, $\Gamma \in |\varphi|_{\Lambda,I} \Rightarrow \Gamma \in |\psi|_{\Lambda,I}$, i.e., by Def.3.12(ii), $\varphi \in \Gamma \Rightarrow \psi \in \Gamma$, hence, $\varphi \notin \Gamma$ or $\psi \in \Gamma$, so, by Lem.3.3(ii), $\neg\varphi \in \Gamma$ or $\psi \in \Gamma$, therefore, by Lem.3.3(ii),(iv), $\neg\varphi \vee \psi \in \Gamma$, hence, $\varphi \supset \psi \in \Gamma$. Now, if $\{\neg(\varphi \supset \psi)\}$ were an $Ic\Lambda$-theory, then, by Lindenbaum's lemma, it would exist a $mIc\Lambda$-theory Γ s.t. $\neg(\varphi \supset \psi) \in \Gamma$, hence, since Γ is consistent, $\varphi \supset \psi \notin \Gamma$, which is a contradiction. Therefore, $\{\neg(\varphi \supset \psi)\}$ is an $Iinc\Lambda$-theory, i.e. $I \vdash_\Lambda \varphi \supset \psi$. ∎

So, if $|\varphi|_{\Lambda,I} = |\psi|_{\Lambda,I}$, then, by the previous Lemma, $I \vdash_\Lambda \varphi \supset \psi$ and $I \vdash_\Lambda \psi \supset \varphi$, hence, $I \vdash_\Lambda \varphi \equiv \psi$, and, with an **RE**-step in the proof, $I \vdash_\Lambda \Box\varphi \equiv \Box\psi$, so, by Lem.3.3(iii), $\Box\varphi \equiv \Box\psi \in \Gamma$, hence, by Lem.3.3(i), $\Box\varphi \in \Gamma \iff \Box\psi \in \Gamma$. So, the well-definition of $E^{\Lambda,I}$ is proved.

Now, a 'Truth Lemma' for canonical models of classical logics can be easily proved.

Lemma 3.14 (Truth Lemma) *Let Λ be a classical modal logic and I be a $c\Lambda$-theory. Then,* $(\forall \varphi \in \mathcal{L}_\Box)(\forall \Gamma \in W^{\Lambda,I})(\mathfrak{M}^{\Lambda,I}, \Gamma \Vdash \varphi \iff \varphi \in \Gamma)$

Proof. By induction on the complexity of φ. Ind.Base follows from Def.3.6(iv) and $\varphi \supset \psi$ – part of Ind.Step follows immediatelly from Ind.Hypothesis using (i) to (iv) of Lem.3.3. Now, to the $\Box\varphi$ – case.

Firstly, let Δ be any $mIc\Lambda$-theory. Then, $\Delta \in \overline{V}(\varphi) \iff \mathfrak{M}^{\Lambda,I}, \Delta \Vdash \varphi \overset{\text{Ind.Hyp.}}{\iff} \varphi \in \Delta \iff \Delta \in |\varphi|_{\Lambda,I}$. Hence, $\overline{V}(\varphi) = |\varphi|_{\Lambda,I}$ (1)

So, $\Box\varphi \in \Gamma \iff |\varphi|_{\Lambda,I} \in E^{\Lambda,I}(\Gamma) \overset{(1)}{\iff} \overline{V}(\varphi) \in E^{\Lambda,I}(\Gamma) \iff \mathfrak{M}^{\Lambda,I}, \Gamma \Vdash \Box\varphi$. ∎

Exactly as in the proof of Prop.3.10, but using Truth Lemma 3.14 instead of Truth Lemma 3.8, we come up with the following result (for classical modal logics this time).

Proposition 3.15 *Let Λ be a classical modal logic and I be a $c\Lambda$-theory. Then,*

$$(\forall \varphi \in \mathcal{L}_\Box)(\mathfrak{M}^{\Lambda,I} \Vdash \varphi \iff I \vdash_\Lambda \varphi)$$

4 RM-stable theories

Having set the appropriate background, we proceed to define our first variant of a stable belief set by taking the most obvious road: substituting $\mathbf{RM_c}$ for $\mathbf{RN_c}$ in Stalnaker's definition.

Definition 4.1 *A theory $S \subseteq \mathcal{L}_\Box$ is called* RM-stable *iff*

(i) $\mathbf{PC} \subseteq S$ *and S is closed under* **MP**

(ii) *S is closed under rule* $\mathbf{RM_c}$. $\frac{\varphi \supset \psi \,\in\, S}{\Box\varphi \supset \Box\psi \,\in\, S}$

(iii) *S is closed under rule* $\mathbf{NI_c}$. $\frac{\varphi \,\notin\, S}{\neg\Box\varphi \,\in\, S}$

The first observation is that the axiom $\Box\top$ plays here a role similar to the one encountered in non-normal modal logics, where $\Box\top$ eliminates queer worlds and leads to the realm of normal modal logics. Addition of $\Box\top$ to an RM-stable set leads to the classical Stalnaker notion.

Fact 4.2 *A theory S is a Stalnaker stable set iff it is an RM-stable set containing $\Box\top$.*

From the proof-theoretic viewpoint, the following result shows that RM-stable sets stand to the regular logic $\mathbf{S5'_R}$, as Stalnaker (RN-)stable sets stand to $\mathbf{S5}$. The following Theorem should be compared to Theor.2.3(i).

Theorem 4.3 *Let S be an RM-stable set.*

(i) $\mathbf{K}$, $\mathbf{T}$, $\mathbf{5}_\top$ *are contained in S.*

(ii) *S is closed under strong $\mathbf{S5'_R}$ provability, i.e.* $S = \{\varphi \in \mathcal{L}_\Box \mid S \vdash_{\mathbf{S5'_R}} \varphi\}$.

(iii) *If S is consistent, then it is a consistent with $\mathbf{S5'_R}$ theory (c$\mathbf{S5'_R}$-theory).*

Proof.

(i) Consider any $\varphi, \psi \in \mathcal{L}_\Box$.

- If $\neg\Box(\varphi \supset \psi) \in S$, then $\Box(\varphi \supset \psi) \supset (\Box\varphi \supset \Box\psi) \in S$.
  If $\neg\Box(\varphi \supset \psi) \notin S$, then, by $\mathbf{NI_c}$, $\varphi \supset \psi \in S$, and, by $\mathbf{RM_c}$, $\Box\varphi \supset \Box\psi \in S$, so again, $\Box(\varphi \supset \psi) \supset (\Box\varphi \supset \Box\psi) \in S$.
- If $\neg\Box\varphi \in S$, then, by Def.4.1(i), $\Box\varphi \supset \varphi \in S$.
  If $\neg\Box\varphi \notin S$, then, by $\mathbf{NI_c}$, $\varphi \in S$, and again, by Def.4.1(i), $\Box\varphi \supset \varphi \in S$.
- If $\Box\top \supset \Box\varphi \in S$, then, by Def.4.1(i), $(\Box\top \supset \Box\varphi) \vee (\Box\top \supset \Box\neg\Box\varphi) \in S$.
  If $\Box\top \supset \Box\varphi \notin S$, then, by $\mathbf{RM_c}$, $\top \supset \varphi \notin S$, hence, by Def.4.1(i), $\varphi \notin S$, so, by $\mathbf{NI_c}$, $\neg\Box\varphi \in S$, and again by Def.4.1(i), $\top \supset \neg\Box\varphi \in S$, consequently, by $\mathbf{RM_c}$, $\Box\top \supset \Box\neg\Box\varphi \in S$, and finally, by Def.4.1(i), $(\Box\top \supset \Box\varphi) \vee (\Box\top \supset \Box\neg\Box\varphi) \in S$. Therefore, in any case, $(\Box\top \supset \Box\varphi) \vee (\Box\top \supset \Box\neg\Box\varphi) \in S$. But it is easy to see that $(\Box\top \supset \Box\varphi) \vee (\Box\top \supset \Box\neg\Box\varphi) \equiv (\neg\Box\varphi \wedge \Box\top \supset \Box\neg\Box\varphi) \in \mathbf{PC}$. Therefore, by Def.4.1(i), $\neg\Box\varphi \wedge \Box\top \supset \Box\neg\Box\varphi \in S$.

(ii)
It is obvious that, if $\varphi \in S$, then $S \vdash_{\mathbf{S5'_R}} \varphi$. Conversely, suppose that $S \vdash_{\mathbf{S5'_R}} \varphi$. Then, since $\mathbf{S5'_R} = \mathbf{KT5}_\top$ (see Lem.A.2), there is a Hilbert-style proof, in which every step is a formula of $\mathbf{PC} \cup \mathbf{K} \cup \mathbf{T} \cup \mathbf{5}_\top \cup S$ or a result of applying **MP** or **RM** to formulas

of previous steps. It will be proved by induction on the proof's length, that $\varphi \in S$. For Ind.Basis, if $\varphi \in \mathbf{PC}$, then, by Def.4.1(i), $\varphi \in S$; if $\varphi \in \mathbf{K} \cup \mathbf{T} \cup \mathbf{5}_\top$, then, by (i), $\varphi \in S$. For Ind.Step, if ψ and $\psi \supset \varphi$ are formulas of the proof in previous steps, then, by Ind.Hypothesis, $\psi, \psi \supset \varphi \in S$ and so, by Def.4.1(i), $\varphi \in S$; if $\varphi = \Box\psi \supset \Box\chi$ and $\psi \supset \chi$ is a formula of the proof in a previous step, then, by Ind.Hypothesis, $\psi \supset \chi \in S$ and so, by $\mathbf{RM_c}$, $\varphi \in S$.
(iii)
Suppose that S is an $inc\mathbf{S5'_R}$-theory. Then $S \vdash_{\mathbf{S5'_R}} \bot$, hence, by (ii), $\bot \in S$, and so, because $\bot \supset \bot \in \mathbf{PC}$, by definition, S is inconsistent. ∎

Representation theory for RM-stable sets. We can provide *model-theoretic characterizations* of RM-stable theories in terms of q-models and n-models. We can set RM-stable theories in an one-to-one-correspondence to theories of q-models consisting of a cluster of normal worlds 'seeing' every non-normal world (if any). We can also characterize RM-stable sets as the set of beliefs held within a normal world in such a q-model.

Theorem 4.4 *Let $S \subseteq \mathcal{L}_\Box$ be a consistent theory. S is RM-stable iff there is a q-model $\mathfrak{M} = \langle W, N, R, V\rangle$ satisfying property $(\mathsf{U_q})$ s.t. $Th(\mathfrak{M}) = S$.*

PROOF. ($\Rightarrow$) Since S is RM-stable and consistent, by Theor.4.3(iii), S is a $c\mathbf{S5'}_R$-theory. So, model $\mathfrak{M}^{\mathbf{S5'}_R,S}$ does exist and, by Prop.3.10, $Th(\mathfrak{M}^{\mathbf{S5'}_R,S}) = \{\varphi \in \mathcal{L}_\Box \mid S \vdash_{\mathbf{S5'}_R} \varphi\}$. Consequently, by Theor.4.3(ii), $Th(\mathfrak{M}^{\mathbf{S5'}_R,S}) = S$.
Now, consider any $\Gamma \in N^{\mathsf{S5'}_R,S}$ and $\Delta \in W^{\mathsf{S5'}_R,S}$. For any $\psi \in \mathcal{L}_\Box$ s.t. $\Box\psi \in \Gamma$, since Γ is $Sc\mathsf{S5'}_R$, $\neg\Box\psi \notin \Gamma$. Suppose now that $\neg\Box\psi$ were in S. Then, $S \vdash_{\mathbf{S5'}_R} \neg\Box\psi$, hence, by Lem.3.3(iii), $\neg\Box\psi \in \Gamma$, which is a contradiction. So $\neg\Box\psi \notin S$. But, S is RM-stable, so, by $\mathbf{NI_c}$, $\psi \in S$, hence, $S \vdash_{\mathbf{S5'}_R} \psi$, consequently, again by Lem.3.3(iii), $\psi \in \Delta$. So, by Def.3.6, $\Gamma R^{\mathbf{S5'}_R,S} \Delta$.
($\Leftarrow$)
For Def.4.1(i). $Th(\mathfrak{M})$ contains every tautology in $\mathcal{L}_\Box$ and is closed under **MP**.
For Def.4.1(ii)($\mathbf{RM_c}$). Let $\varphi, \psi \in \mathcal{L}_\Box$ s.t. $\varphi \supset \psi \in Th(\mathfrak{M})$ and $w \in W$ s.t. $\mathfrak{M}, w \Vdash \Box\varphi$. Then, $w \in N$ and $(\forall v \in W)\ wRv \Rightarrow \mathfrak{M}, v \Vdash \varphi$. Therefore, since $\varphi \supset \psi \in Th(\mathfrak{M})$, $\mathfrak{M}, v \Vdash \psi$, hence, $\mathfrak{M}, w \Vdash \Box\psi$. So, $\Box\varphi \supset \Box\psi \in Th(\mathfrak{M})$.
For Def.4.1(iii)($\mathbf{NI_c}$). Let $\varphi \in \mathcal{L}_\Box$ s.t. $\varphi \notin Th(\mathfrak{M})$ i.e. there is $v \in W$ s.t. $\mathfrak{M}, v \nVdash \varphi$. Let now be any $w \in W$. If $w \in W \setminus N$, then, by definition of q-models, $\mathfrak{M}, w \Vdash \neg\Box\varphi$. If $w \in N$, then again, since wRv and $\mathfrak{M}, v \nVdash \varphi$, $\mathfrak{M}, w \Vdash \neg\Box\varphi$.
So, $\neg\Box\varphi \in Th(\mathfrak{M})$. ∎

The following characterization is the parallel to the characterization of Stalnaker stable sets in terms of beliefs held 'inside' a **KD45** situation, and as such, seems amenable to generalization in multi-agent situations (as argued convincingly in [Hal97b]).

Proposition 4.5 *Let $S \subseteq \mathcal{L}_\Box$ be a consistent theory. S is RM-stable iff there is a q-model $\mathfrak{M} = \langle W, N, R, V\rangle$ and $u \in N$ s.t. $S = \{\varphi \in \mathcal{L}_\Box \mid \mathfrak{M}, u \Vdash \Box\varphi\}$ and $(\forall w \in N)(\forall v \in W \setminus \{u\})wRv$.*

PROOF. Firstly, it will be proved that, if $\mathfrak{M}' = \langle W', N', R', V'\rangle$ is a q-model s.t. $(\forall w \in N')(\forall v \in W')wR'v$ and $\mathfrak{M} = \langle W, N, R, V\rangle$ is another q-model s.t. $W = W' \cup \{u\}$ (where $u \notin W'$), $N = N' \cup \{u\}$, $R = R' \cup (\{u\} \times W')$ and $V = V'$, then, $Th(\mathfrak{M}') = \{\varphi \in \mathcal{L}_\Box \mid \mathfrak{M}, u \Vdash \Box\varphi\}$.
Proof: Since $(\forall w \in W')\neg wRu$, it can be proved (by a trivial induction on φ) that $(\forall w \in W')\mathfrak{M}', w \Vdash \varphi$ iff $(\forall w \in W')\mathfrak{M}, w \Vdash \varphi$, hence, $\varphi \in Th(\mathfrak{M}')$ iff $\mathfrak{M}, u \Vdash \Box\varphi$.
Now, Theor.4.4 is applicable (on $\mathfrak{M}'$), and the proof is complete. ∎

By using again Theor.4.4, we obtain a representation for RM-stable sets, in terms of neighborhood semantics.

Proposition 4.6 *Let $S \subseteq \mathcal{L}_\Box$ be a consistent theory. S is RM-stable iff there is an n-model $\mathfrak{N} = \langle W, E, V\rangle$ s.t. $Th(\mathfrak{N}) = S$ and $(\forall w \in W)(E(w) = \varnothing$ or $E(w) = \{W\})$.*

PROOF. $(\Rightarrow)$ By Theor.4.4, there is a q-model $\mathfrak{M} = \langle W, N, R, V\rangle$ s.t. $Th(\mathfrak{M}) = S$ and $(\forall w \in N)(\forall v \in W)wRv$. Consider now $\mathfrak{N}_\mathfrak{M} = \langle W, E, V\rangle$, the equivalent n-model produced by $\mathfrak{M}$ (see Def.2.1). By Prop.2.2 follows immediately that $Th(\mathfrak{N}_\mathfrak{M}) = Th(\mathfrak{M}) = S$. Furthermore, if $w \in W \setminus N$, then $E(w) = \varnothing$ and if $w \in N$, then $E(w) = \{X \subseteq W \mid R_w \subseteq X\} = \{W\}$, since $(\forall v \in W)wRv$.
$(\Leftarrow)$
For Def.4.1(i). $Th(\mathfrak{N})$ contains every tautology in $\mathcal{L}_\Box$ and is closed under (MP).
For Def.4.1(ii)($\mathbf{RM_c}$). Let $\varphi, \psi \in \mathcal{L}_\Box$ s.t. $\varphi \supset \psi \in Th(\mathfrak{N})$ and $w \in W$ s.t. $\mathfrak{N}, w \Vdash \Box\varphi$. Then, $\overline{V}(\varphi) \in E(w)$, hence, $E(w) = \{W\}$ and $\overline{V}(\varphi) = W$. Therefore, since $\varphi \supset \psi \in Th(\mathfrak{N})$, $(\forall w \in W)\mathfrak{N}, w \Vdash \psi$, i.e. $\overline{V}(\psi) = W$, so, $\overline{V}(\psi) \in E(w)$, hence, $\mathfrak{N}, w \Vdash \Box\psi$. So, $\Box\varphi \supset \Box\psi \in Th(\mathfrak{N})$.
For Def.4.1(iii)($\mathbf{NI_c}$). Let $\varphi \in \mathcal{L}_\Box$ s.t. $\varphi \notin Th(\mathfrak{N})$ i.e. $\overline{V}(\varphi) \neq W$. Let now be any $w \in W$. $E(w) = \varnothing$ or $E(w) = \{W\}$, so in both cases, $\overline{V}(\varphi) \notin E(w)$. Hence, $\mathfrak{N}, w \Vdash \neg\Box\varphi$. So, $\neg\Box\varphi \in Th(\mathfrak{N})$. ∎

Furthermore, Theor.2.3(ii) comes now as an immediate consequence of Theor.4.4.

Corollary 4.7 *A consistent theory is stable iff it is a theory of a standard Kripke model (without impossible worlds), equipped with a universal relation.*

PROOF. $(\Rightarrow)$
Let S be a consistent and stable theory. By Fact.4.2, it is RM-stable and contains $\Box\top$, so, $S \vdash_{\mathbf{S5'}_R} \Box\top$, hence, by Lem.3.3(iii), for any $mSc\mathbf{S5'}_R$-theory Γ, $\Box\top \in \Gamma$, so, $N^{\mathbf{S5'}_R,S} = W^{\mathbf{S5'}_R,S}$. Consequently, by Theor.4.4, $Th(\mathfrak{M}^{\mathbf{S5'}_R,S}) = S$ and $\mathfrak{M}^{\mathbf{S5'}_R,S}$ has a universal relation.
$(\Leftarrow)$
Let $\mathfrak{M} = \langle W, R, V\rangle$ be a universal, standard Kripke model. Then, $\mathfrak{M}^q = \langle W, W, R, V\rangle$ is a q-model, and by definition of truth in q-models, $Th(\mathfrak{M}) = Th(\mathfrak{M}^q)$. But, by Theor.4.4 (applied for $\mathfrak{M}^q$), $Th(\mathfrak{M}^q)$ is RM-stable. Furthermore, since $\mathfrak{M}$ is standard, $\Box\top \in Th(\mathfrak{M})$, hence, by Fact.4.2, $Th(\mathfrak{M})$ is stable. ∎

Analogously, Theor.2.3(iii) comes as an immediate consequence of Prop.4.5. Finally, as a result of Prop.4.6, we obtain immediately the following representation of Stalnaker stable sets, in terms of n-models, given for the first time.

Proposition 4.8 *Let $S \subseteq \mathcal{L}_\Box$ be a consistent theory. S is stable iff there is an n-model $\mathfrak{N} = \langle W, E, V \rangle$ s.t. $Th(\mathfrak{N}) = S$ and $(\forall w \in W) E(w) = \{W\}$.*

5 RE-stable theories

Following a typical route, it is tempting to attempt weakening further the positive introspection condition. Rule $\mathbf{RE_c}$ seems the obvious candidate, but we have soon to face the obvious problem that the introspective reasoner should be able to distinguish tautologies as equivalent formulae. We have then to consider the addition of $\Box\top$ and this leads us to the following generic notion:

Definition 5.1 *A theory $S \subseteq \mathcal{L}_\Box$ is called* RE-stable *iff*

(i) $\mathbf{PC} \subseteq S$ *and S is closed under* $\mathbf{MP}$

(ii) $\Box\top \in S$

(iii) *S is closed under rule* $\mathbf{RE_c}$. $\frac{\varphi \equiv \psi \ \in \ S}{\Box\varphi \equiv \Box\psi \ \in \ S}$

With proofs identical to Theorem's 4.3(ii) and (iii), we can conclude that RE-stable theories are consistent with strong provability in classical modal logics.

Proposition 5.2 *Let S be an RE-stable set containing every instance of axiomatic schemes $\mathbf{A_0}, \ldots, \mathbf{A_n}$.*

(i) S is closed under strong $\mathbf{A_0} \ldots \mathbf{A_{nC}}$ provability, i.e. $S = \{\varphi \in \mathcal{L}_\Box \mid S \vdash_{\mathbf{A_0} \ldots \mathbf{A_{1C}}} \varphi\}$.

(ii) If S is consistent, then it is a consistent with $\mathbf{A_1} \ldots \mathbf{A_{nC}}$ theory (c$\mathbf{A_1} \ldots \mathbf{A_{nC}}$-theory)

But, it comes that by adding $\Box\top$, we get nothing less than $\mathbf{RN_c}$, as in the original definition.

Lemma 5.3 *Any RE-stable theory is closed under $\mathbf{RN_c}$.*

Proof. Let S be an RE-stable theory and $\varphi \in S$. Since $\varphi \supset (\top \supset \varphi) \in S$, by Def.5.1(i), $\top \supset \varphi \in S$. Furthermore, $\varphi \supset \top \in S$, so, by Def.5.1(i), $\top \equiv \varphi \in S$, hence, by $\mathbf{RE_c}$, $\Box\top \equiv \Box\varphi \in S$, and, by Def.5.1(i), $\Box\top \supset \Box\varphi \in S$, and finally, by Def.5.1(ii) and (i), $\Box\varphi \in S$. ∎

This means we have to proceed to different notions of negative introspection and by doing so, we obtain two different notions of RE-stable sets.

5.1 REw-stable theories

We introduce the following context rule for negative introspection:

$$\mathbf{NI_{c-w}}.\frac{\neg\varphi \notin S}{\Box\varphi \in S \;\vee\; \neg\Box\varphi \in S}$$

which 'says' that *if φ is consistent with what is believed, something is known about it.*

Definition 5.4 *An RE-stable theory S is called* REw-stable *iff it is closed under* $\mathbf{NI_{c-w}}$.

We readily prove the presence of axiom **w5** and then, we can obtain a representation theorem for REw-stable theories in terms of n-models.

Lemma 5.5 *Every instance of axiomatic scheme* **w5** *is contained in any REw-stable theory.*

Proof. Let S be an REw-stable theory and $\varphi \in \mathcal{L}_\Box$.
If $\neg\varphi \in S$ or $\Box\varphi \in S$, then, by Def.5.1(i), $(\varphi \wedge \neg\Box\varphi) \supset \Box\neg\Box\varphi \in S$.
If $\neg\varphi \notin S$ and $\Box\varphi \notin S$, then, by $\mathbf{NI_{c-w}}$, $\neg\Box\varphi \in S$, and , by Lem.5.3, $\Box\neg\Box\varphi \in S$, hence again, $(\varphi \wedge \neg\Box\varphi) \supset \Box\neg\Box\varphi \in S$. ∎

Theorem 5.6 *Let $S \subseteq \mathcal{L}_\Box$ be a consistent theory. S is REw-stable iff there is an n-model $\mathfrak{N} = \langle W, E, V\rangle$ s.t. $Th(\mathfrak{N}) = S$ and*

$$(\forall w \in W) W \in E(w) \quad (1) \qquad \textit{and} \qquad (\forall v \in W)(E(v) \setminus E(w) \subseteq \{\varnothing\}) \quad (2)$$

Proof. (⇒) Since S is REw-stable, by Lem.5.5, S contains **w5**, hence, since S is RE-stable and consistent, by Prop.5.2(ii), S is a $c\mathbf{w5}_C$-theory. So, model $\mathfrak{N}^{\mathbf{w5}_C,S}$ does exist. For simplicity, let us denote $\mathfrak{N}^{\mathbf{w5}_C,S}$ as $\mathfrak{N} = \langle W, E, V\rangle$. Then, by Prop.3.15, $Th(\mathfrak{N}) = \{\varphi \in \mathcal{L}_\Box \mid S \vdash_{\mathbf{w5}_C} \varphi\}$. Consequently, by Prop.5.2(i), $Th(\mathfrak{N}) = S$. Now, fix any $\Gamma \in W$.
(1) By Def.5.1(i), $\top \in S$, so, by Lem.3.3(iii), $(\forall \Delta \in W)\top \in \Delta$, hence, since every Δ is a $mSc\mathbf{w5}_C$-theory, $|\top|_{\mathsf{w5}_C,S} = W$. But, by Def.5.1(ii), $\Box\top \in S$, i.e., by Lem.3.3(iii), $\Box\top \in \Gamma$, hence, by Def.3.12(ii), $|\top|_{\mathsf{w5}_C,S} \in E(\Gamma)$. Consequently, $W \in E(\Gamma)$.
(2) Consider any $\Delta \in W$ and let $Y \subseteq W$ s.t. $Y \in E(\Delta)$ but $Y \notin E(\Gamma)$. Then, by Def.3.12(ii), there must be a $\varphi \in \mathcal{L}_\Box$ s.t. $Y = |\varphi|_{\mathbf{w5}_C,S}$ and $\Box\varphi \in \Delta$ $\quad (I)$
But, since $Y \notin E(\Gamma)$, $\Box\varphi \notin \Gamma$, hence, by Lem.3.3(iii), $\Box\varphi \notin S$ $\quad (II)$
Suppose now, for the sake of contradiction, that $Y \neq \varnothing$. Then, there is a $\Xi \in Y$. Since $Y = |\varphi|_{\mathbf{w5}_C,S}$, $\varphi \in \Xi$, and since Ξ is consistent, $\neg\varphi \notin \Xi$, so, by Lem.3.3(iii), $\neg\varphi \notin S$ $\quad (III)$
Now, (II) and (III) imply by $\mathbf{NI_{c-w}}$, $\neg\Box\varphi \in S$, therefore, again by Lem.3.3(iii), $\neg\Box\varphi \in \Delta$, hence, by (I), Δ is inconsistent, which is a contradiction. So, $Y = \varnothing$.
(⇐)
For Def.5.1(i). $Th(\mathfrak{N})$ contains every tautology in $\mathcal{L}_\Box$ and is closed under (MP).
For Def.5.1(ii). Since $\overline{V}(\top) = W$ and, by (1), $(\forall w \in W) W \in E(w)$, $\Box\top \in Th(\mathfrak{N})$.

For Def.5.1(iii)($\mathbf{RE_c}$). Let $\varphi, \psi \in \mathcal{L}_\Box$ s.t. $\varphi \equiv \psi \in Th(\mathfrak{N})$. Then, $\overline{V}(\varphi) = \overline{V}(\psi)$, hence, $(\forall w \in W)$ $(\overline{V}(\varphi) \in E(w) \iff \overline{V}(\psi) \in E(w))$, consequently, $\Box\varphi \equiv \Box\psi \in Th(\mathfrak{N})$.
For Def.5.4($\mathbf{NI_{c-w}}$). Let $\varphi \in \mathcal{L}_\Box$ s.t. $\neg\varphi \notin Th(\mathfrak{N})$ and $\Box\varphi \notin Th(\mathfrak{N})$. Then, $\overline{V}(\neg\varphi) \neq W$ and $(\exists w \in W)$ $\mathfrak{N}, w \nVdash \Box\varphi$, i.e. $\overline{V}(\varphi) \neq \varnothing$ and $(\exists w \in W)\overline{V}(\varphi) \notin E(w)$. Now, suppose for the sake of contradiction, that there is a $v \in W$ s.t. $\overline{V}(\varphi) \in E(v)$. Then, $\overline{V}(\varphi) \in E(v) \setminus E(w)$, hence, by (2), $\overline{V}(\varphi) = \varnothing$, which is a contradiction. So, $(\forall v \in W)$ $\overline{V}(\varphi) \notin E(v)$, i.e. $(\forall v \in W)$ $\mathfrak{N}, v \Vdash \neg\Box\varphi$, hence $\neg\Box\varphi \in Th(\mathfrak{N})$. ∎

5.2 REp-stable theories

We can alternatively consider the following rule for negative introspection:

$$\mathbf{NI_{c-p}}.\frac{\varphi \notin S \ \wedge \ \neg\varphi \notin S}{\neg\Box\varphi \in S}$$

which 'says' that *if nothing is known to hold about φ, then it is known that φ is not known.*

Definition 5.7 *An RE-stable theory S is called* REp-stable *iff it is closed under* $\mathbf{NI_{c-p}}$.

This notion is stronger than the previous one and contains every instance of axiom **p5**, introduced in [KZ09].

If S is an REp-stable theory, then $\Box\varphi \notin S$ implies, by Lem.5.3, $\varphi \notin S$, hence, $\neg\varphi \notin S$ and $\Box\varphi \notin S$ imply $\neg\varphi \notin S$ and $\varphi \notin S$, so, by $\mathbf{NI_{c-p}}$, $\neg\varphi \notin S$ and $\Box\varphi \notin S$ imply $\neg\Box\varphi \in S$. This proves the following.

Fact 5.8 *Every REp-stable theory is REw-stable.*

Lemma 5.9 *Every instance of axiomatic scheme* **p5** *is contained in any REp-stable theory.*

Proof. Let S be an REp-stable theory and $\varphi \in \mathcal{L}_\Box$.
If $\Box\varphi \in S$ or $\Box\neg\varphi \in S$, then, by Def.5.1(i), $(\neg\Box\varphi \wedge \neg\Box\neg\varphi) \supset \Box\neg\Box\varphi \in S$.
If $\Box\varphi \notin S$ and $\Box\neg\varphi \notin S$, then, by Lem.5.3, $\varphi \notin S$ and $\neg\varphi \notin S$, so, by $\mathbf{NI_{c-p}}$, $\neg\Box\varphi \in S$, and, by Lem.5.3, $\Box\neg\Box\varphi \in S$, hence again, $(\neg\Box\varphi \wedge \neg\Box\neg\varphi) \supset \Box\neg\Box\varphi \in S$. ∎

Furthermore, we can prove a representation theorem for REp-stable sets.

Theorem 5.10 *Let $S \subseteq \mathcal{L}_\Box$ be a consistent theory. S is REp-stable iff there is an n-model $\mathfrak{N} = \langle W, E, V\rangle$ s.t. $Th(\mathfrak{N}) = S$ and $(\forall w \in W)(E(w) = \{W\}$ or $E(w) = \{\varnothing, W\})$.*

Proof. ($\Rightarrow$) Since S is REp-stable, by Lem.5.9, S contains **p5**, hence, since S is RE-stable and consistent, by Prop.5.2(ii), S is a $c\mathbf{p5}_C$-theory. So, model $\mathfrak{N}^{\mathbf{p5}_C, S}$ does exist. For simplicity, let us denote $\mathfrak{N}^{\mathbf{p5}_C, S}$ as $\mathfrak{N} = \langle W, E, V\rangle$. Then, by Prop.3.15, $Th(\mathfrak{N}) =$

$\{\varphi \in \mathcal{L}_\Box \mid S \vdash_{\mathbf{p5}_C} \varphi\}$. Consequently, by Prop.5.2(i), $Th(\mathfrak{N}) = S$.
Now, let $\mathsf{r} \in W$. Exactly as in Theor.5.6(1), one can prove that $W \in E(\mathsf{r})$.
Consider now any $Y \in E(\mathsf{r})$ s.t. $Y \neq W$. Then, by Def.3.12(ii), there must be a $\varphi \in \mathcal{L}_\Box$ s.t. $Y = |\varphi|_{\mathbf{p5}_C,S}$ and $\Box\varphi \in \mathsf{r}$ $\quad (I)$
But, since $|\varphi|_{\mathbf{p5}_C,S} \subset W$, there is a $mSc\mathbf{p5}_C$-theory Δ s.t. $\Delta \notin |\varphi|_{\mathbf{p5}_C,S}$, hence, $\varphi \notin \Delta$, consequently, by Lem.3.3(iii), $\varphi \notin S$ $\quad (II)$
Suppose now, for the sake of contradiction, that $Y \neq \varnothing$. Then, there is a $\Xi \in Y$. Since $Y = |\varphi|_{\mathbf{p5}_C,S}$, $\varphi \in \Xi$, and since Ξ is consistent, $\neg\varphi \notin \Xi$, so, by Lem.3.3(iii), $\neg\varphi \notin S$ $\quad (III)$
Now, (II) and (III) imply by $\mathbf{NI_{c-p}}$, $\neg\Box\varphi \in S$, therefore, again by Lem.3.3(iii), $\neg\Box\varphi \in \mathsf{r}$, hence, by (I), r is inconsistent, which is a contradiction. So, $Y = \varnothing$.
$(\Leftarrow)$
Properties (i), (ii) and (iii)($\mathbf{RE_c}$) in Def.5.1 can be proved exactly as in Theor.5.6. So, let us prove property $\mathbf{NI_{c-p}}$ (of Def.5.7). Let $\varphi \in \mathcal{L}_\Box$ s.t. $\varphi \notin Th(\mathfrak{N})$ and $\neg\varphi \notin Th(\mathfrak{N})$. Then, $\overline{V}(\varphi) \neq W$ and $\overline{V}(\varphi) \neq \varnothing$, hence, for any $w \in W$, since $E(w) = \{W\}$ or $E(w) = \{\varnothing, W\}$, $\overline{V}(\varphi) \notin E(w)$, consequently, $(\forall w \in W)\ \mathfrak{N}, w \Vdash \neg\Box\varphi$, hence $\neg\Box\varphi \in Th(\mathfrak{N})$. ∎

Remark 5.11 If $(\forall w \in W)(E(w) = \{W\}$ or $E(w) = \{\varnothing, W\})$, then E satisfies properties (1) and (2) of Theor.5.6. So, using Theor.5.10 and Theor.5.6, we see again that every REp-stable theory is REw-stable.

Theorem 5.10 and Fact 5.8 allow us to prove that REp-stable (and hence, REw-stable) theories do not suffer from the presence of all known epistemic axioms.

Corollary 5.12 *There is an REp-stable theory (which is also REw-stable), which does not contain an instance of* **K**, *of* **T**, *of* **4** *and of* **5**.

Proof. Consider the n-model $\mathfrak{N} = \langle W, E, V\rangle$ where $W = \{w, v\}$, $E(w) = \{\varnothing, W\}$, $E(v) = \{W\}$, $V(p) = \varnothing$ and $V(q) = \{w\}$. Then, by Theor.5.10, $Th(\mathfrak{N})$ is REp-stable. Furthermore,

- $\overline{V}(p \supset q) = W \in E(w)$, $V(p) \in E(w)$ but $V(q) \notin E(w)$, hence $\mathfrak{N}, w \Vdash \Box(p \supset q) \wedge\Box p \wedge \neg\Box q$, therefore $(\Box p \wedge \Box(p \supset q)) \supset \Box q \notin Th(\mathfrak{N})$.
- $V(p) \in E(w)$ but $w \notin V(p)$, hence $\mathfrak{N}, w \Vdash \Box p \wedge \neg p$, therefore $\Box p \supset p \notin Th(\mathfrak{N})$.
- $V(p) \in E(w)$ but $\{w\} \notin E(w)$, hence, $\{u \in W \mid V(p) \in E(u)\} \notin E(w)$, so, $\overline{V}(\Box p) \notin E(w)$, i.e. $\mathfrak{N}, w \Vdash \Box p \wedge \neg\Box\Box p$, therefore $\Box p \supset \Box\Box p \notin Th(\mathfrak{N})$.
- $V(p) \notin E(v)$ but $\{v\} \notin E(v)$, hence, $W \setminus \{u \in W \mid V(p) \in E(u)\} \notin E(v)$, so, $\overline{V}(\neg\Box p) \notin E(v)$, i.e. $\mathfrak{N}, w \Vdash \neg\Box p \wedge \neg\Box\neg\Box p$, therefore $\neg\Box p \supset \Box\neg\Box p \notin Th(\mathfrak{N})$.

∎

6 Related Work - Future Research

The notion of a stable belief set has been very useful in modal nonmonotonic reasoning. Investigations on stable sets have mainly focused on identifying their technical properties and representing them with the aid of model-theoretic constructions known from classical modal logic. It seems natural however to investigate, both from the logician's and the KR engineer's viewpoint, what can be obtained by loosening the conditions in the original definition of R. Stalnaker. To the best of our knowledge, it is the first time that notions of stable sets are investigated by varying the positive and negative introspection closure conditions. Up to now, there have been approaches which build belief sets by changing classical logic in condition (i) to a weaker one (intuitionistic logic) [ACP97] or generalizing the notion of stability in a way somewhat related to the second question of our introduction [Jas91].

The basic motivation of the research reported in our paper, is to define more plausible notions of an epistemic state and the ultimate goal is to employ these notions in new mechanisms for nonmonotonic modal logics, à la McDermott and Doyle. The latter goal is the first step in the roads of future research, along with the investigation on the assessment of epistemic states which emerge if we adopt even weaker notions of positive introspection, for instance by employing a context-dependent version of *Oscar Becker's rule* which has been employed in the study of modal systems which go some way towards solving the logical omniscience problem [Fit93].

A Appendix

A.1 Regular Modal Logic $\mathbf{S5'_R}$

Firstly, let us point out that although any regular modal logic Λ is closed under uniform substitution (**US**) and every proof in Λ does not contain any **US**-step, one can prove (see final part of this Appendix) that

Lemma A.1 $(\forall\varphi \in \mathcal{L}_\Box)(\vdash_\Lambda \varphi \iff \varphi \in \Lambda)$

We remind following definitions: $\mathbf{5}_\top = US(\neg\Box p \wedge \Box\top \supset \Box\neg\Box p)$ and $\mathbf{S5'_R} = \mathbf{KT4}_\top\mathbf{B}_\top\mathbf{R}$. Then,

Lemma A.2 $\mathbf{S5'_R} = \mathbf{KT5}_\top\mathbf{R}$

Proof. By Lem.A.1, we can work with syntactical proofs.
$(\subseteq)$

Following proof shows that $\vdash_{\mathbf{KT5}_\top\mathbf{R}} \mathbf{B}_\top$

1.	$\varphi \supset \neg\Box\neg\varphi$	(**T**)
2.	$\varphi \wedge \Box\top \supset \neg\Box\neg\varphi \wedge \Box\top$	(1. **PC**)
3.	$\neg\Box\neg\varphi \wedge \Box\top \supset \Box\neg\Box\neg\varphi$	($\mathbf{5}_\top$)
4.	$\varphi \wedge \Box\top \supset \Box\neg\Box\neg\varphi$	(2. 3. **PC**)

Next proof shows that $\vdash_{\mathbf{KT5_{\top}R}} \mathbf{4}_{\top}$

1.	$\varphi \supset \top$	(**PC**)
2.	$\Box\varphi \supset \Box\top$	(1. **RM**)
3.	$\Box\varphi \supset \Box\varphi \land \Box\top$	(2. **PC**)
4.	$\Box\varphi \supset \neg\Box\neg\Box\varphi$	(**T**)
5.	$\Box\varphi \land \Box\top \supset \neg\Box\neg\Box\varphi \land \Box\top$	(4. **PC**)
6.	$\neg\Box\neg\Box\varphi \land \Box\top \supset \Box\neg\Box\neg\Box\varphi$	($\mathbf{5}_{\top}$)
7.	$\Box\varphi \land \Box\top \supset \Box\neg\Box\neg\Box\varphi$	(5. 6. **PC**)
8.	$\Box\varphi \supset \Box\neg\Box\neg\Box\varphi$	(3. 7. **PC**)
9.	$\neg\Box\neg\Box\varphi \supset (\Box\top \supset \Box\varphi)$	($\mathbf{5}_{\top}$)
10.	$\Box\neg\Box\neg\Box\varphi \supset \Box(\Box\top \supset \Box\varphi)$	(9. **RM**)
11.	$\Box\varphi \supset \Box(\Box\top \supset \Box\varphi)$	(8. 10. **PC**)

($\supseteq$)

Following proof shows that $\vdash_{\mathbf{KT4_{\top}B_{\top}R}} \mathbf{5}_{\top}$

1.	$(\Box\top \supset \Box\varphi) \supset \neg\neg(\Box\top \supset \Box\varphi)$	(**PC**)
2.	$\Box(\Box\top \supset \Box\varphi) \supset \Box\neg\neg(\Box\top \supset \Box\varphi)$	(1. **RM**)
3.	$\neg\Box\neg\neg(\Box\top \supset \Box\varphi) \supset \neg\Box(\Box\top \supset \Box\varphi)$	(2. **PC**)
4.	$\Box\neg\Box\neg\neg(\Box\top \supset \Box\varphi) \supset \Box\neg\Box(\Box\top \supset \Box\varphi)$	(3. **RM**)
5.	$\neg(\Box\top \supset \Box\varphi) \land \Box\top \supset \Box\neg\Box\neg\neg(\Box\top \supset \Box\varphi)$	($\mathbf{B}_{\top}$)
6.	$\neg(\Box\top \supset \Box\varphi) \land \Box\top \supset \Box\neg\Box(\Box\top \supset \Box\varphi)$	(5. 4. **PC**)
7.	$\neg\Box\varphi \land \Box\top \supset \neg(\Box\top \supset \Box\varphi) \land \Box\top$	(**PC**)
8.	$\neg\Box\varphi \land \Box\top \supset \Box\neg\Box(\Box\top \supset \Box\varphi)$	(7. 6. **PC**)
9.	$\neg\Box(\Box\top \supset \Box\varphi) \supset \neg\Box\varphi$	($\mathbf{4}_{\top}$)
10.	$\Box\neg\Box(\Box\top \supset \Box\varphi) \supset \Box\neg\Box\varphi$	(9. **RM**)
11.	$\neg\Box\varphi \land \Box\top \supset \Box\neg\Box\varphi$	(8. 10. **PC**)

■

Furthermore, for a q-frame $\mathfrak{F} = \langle W, N, R\rangle$, we employ following properties:

($\mathsf{E_q}$) $(\forall w, v \in N)(\forall u \in W)(wRv \land wRu \Rightarrow vRu)$

($\mathsf{ER_q}$) (R is an equivalence relation in N and
$(\forall w \in N)(\forall u \in W \setminus N)(wRu \Rightarrow (\forall v \in [w]_R)vRu)$ [3]

Then, following correspondence results can be proved.

Proposition A.3

(i) $\mathfrak{F} \Vdash \mathbf{5}_{\top} \iff$ ($\mathsf{E_q}$) *holds for* $\mathfrak{F}$

(ii) $\mathfrak{F} \Vdash \mathbf{T} \iff \mathfrak{F}$ *is reflexive in* N

Proof.
(i)($\Rightarrow$)
The contrapositive will be proved. Suppose that $(\exists w, v \in N)(\exists u \in W)(wRv \land wRu \land \neg vRu)$. Now, let V be a valuation s.t. $V(p) = \{s \in W \mid vRs\}$. Then, $\langle\mathfrak{F}, V\rangle, v \Vdash \Box p$,

[3] $[w]_R$ is the equivalence class of w, i.e. $[w]_R = \{v \in N \mid wRv\}$.

hence, since wRv, $\langle\mathfrak{F},V\rangle,w \Vdash \neg\Box\neg\Box p$. Furthermore, since $\neg vRu$, $\langle\mathfrak{F},V\rangle,u \Vdash \neg p$, so, since wRu, $\langle\mathfrak{F},V\rangle,w \Vdash \neg\Box p$. But $w \in N$, so, $\langle\mathfrak{F},V\rangle,w \Vdash \Box\top$. Putting all together: $\langle\mathfrak{F},V\rangle,w \Vdash \neg\Box p \wedge \Box\top \wedge \neg\Box\neg\Box p$.
($\Leftarrow$)
Let φ be a formula, V a valuation and w a world s.t. $\langle\mathfrak{F},V\rangle,w \Vdash \Box\top \wedge \neg\Box\neg\Box\varphi$. Then, $w \in N$ and there is a $v \in W$ s.t. wRv and $\langle\mathfrak{F},V\rangle,v \Vdash \Box\varphi$. But then, $v \in N$. Consider now any $u \in W$ s.t. wRu. Since wRv and $w,v \in N$, by ($\mathsf{E_q}$), vRu, therefore, since $\langle\mathfrak{F},V\rangle,v \Vdash \Box\varphi$, $\langle\mathfrak{F},V\rangle,u \Vdash \varphi$. Hence, $\langle\mathfrak{F},V\rangle,w \Vdash \Box\varphi$, i.e. $\langle\mathfrak{F},V\rangle,w \Vdash \Box\top \wedge \neg\Box\neg\Box\varphi \supset \Box\varphi$.
(ii)($\Rightarrow$)
The contrapositive will be proved. Suppose that $(\exists w \in N)\neg wRw$. Now, let V be a valuation s.t. $V(p) = W \setminus \{w\}$. Then of course, $\langle\mathfrak{F},V\rangle,w \Vdash \neg p$. Consider now any $v \in W$ s.t. wRv. If $\langle\mathfrak{F},V\rangle,v \Vdash \neg p$, then $v = w$, hence wRw, which is a contradiction. So, $\langle\mathfrak{F},V\rangle,v \Vdash p$, and since $w \in N$, $\langle\mathfrak{F},V\rangle,w \Vdash \Box p$.
($\Leftarrow$)
Let φ be a formula, V a valuation and w a world s.t. $\langle\mathfrak{F},V\rangle,w \Vdash \Box\varphi$. Then, $w \in N$ and since $\mathfrak{F}$ is reflexive in N, wRw, hence $\langle\mathfrak{F},V\rangle,w \Vdash \varphi$, i.e. $\langle\mathfrak{F},V\rangle,w \Vdash \Box\varphi \supset \varphi$. ∎

Corollary A.4 $\mathfrak{F} \Vdash \mathbf{T} \wedge \mathbf{5}_\top \iff$ ($\mathsf{ER_q}$) *holds for* $\mathfrak{F}$

Proof. By Prop.A.3, it suffices to show

$$(\mathsf{E_q}) \text{ holds for } \mathfrak{F} \text{ and } \mathfrak{F} \text{ is reflexive in } N \iff (\mathsf{ER_q}) \text{ holds for } \mathfrak{F}$$

($\Rightarrow$)
Reflexivity in N is guaranteed. For symmetry in N, consider any $w,v \in N$ s.t. wRv. Since wRw, by ($\mathsf{E_q}$), vRw. For transitivity in N, consider any $w,v,u \in N$ s.t. wRv and vRu. Then, by symmetry, vRw, and by ($\mathsf{E_q}$), wRu. Hence, R is an equivalence relation in N.
Let now w be a normal world and u be a non-normal world s.t. wRu. Furthermore, consider any $v \in [w]_R$, i.e. $v \in N$ and wRv. Then, by ($\mathsf{E_q}$), vRu.
($\Leftarrow$)
Since R is an equivalence relation in N, $\mathfrak{F}$ is reflexive in N. Consider now any $w,v \in N$ and $u \in W$ s.t. wRv and wRu. If $u \in N$, then, since wRv, by symmetry, vRw, and since wRu, by transitivity, vRu. If $u \in W \setminus N$, then, since $v \in [w]_R$, by ($\mathsf{ER_q}$), again vRu. Hence, ($\mathsf{E_q}$) holds for $\mathfrak{F}$. ∎

Next two lemmas will be helpful for proving the completeness result for $\mathbf{S5'_R}$. Fix any regular modal logic Λ. Then,

Lemma A.5

(i) If Λ is consistent, then $\varnothing$ is a $c\Lambda$-theory.

(ii) $I \cup \{\neg\varphi\} : \varnothing inc\Lambda \;\Rightarrow\; I \vdash_\Lambda \varphi \;\Rightarrow\; I \cup \{\neg\varphi\} : inc\Lambda$

(iii) If I is a $\varnothing c\Lambda$-theory, then it is a consistent theory.

PROOF.
(i)
If $\varnothing$ is an $inc\Lambda$-theory, then, $\vdash_\Lambda \bot$, consequently, by Lem.A.1, $\bot \in \Lambda$, i.e., since $\bot \supset \bot \in \mathbf{PC}$, Λ is inconsistent.
(ii)
Supposed that $I \cup \{\neg\varphi\}$ is $\varnothing inc\Lambda$, there are $n > 0$ and $\varphi_1, \ldots, \varphi_n \in I$ s.t. $\vdash_\Lambda \varphi_1 \wedge \ldots \wedge \varphi_n \wedge \neg\varphi \supset \bot$ or $\vdash_\Lambda \varphi_1 \wedge \ldots \wedge \varphi_n \supset \bot$ or $\vdash_\Lambda \neg\varphi \supset \bot$. Hence, $I \vdash_\Lambda \varphi_1 \wedge \ldots \wedge \varphi_n \supset \varphi$ or $I \vdash_\Lambda \varphi$. Now, by adding to the first proof, formulas $\varphi_1, \ldots, \varphi_n$ and by applying **MP** n times, we get again a proof of φ from I in Λ ($I \vdash_\Lambda \varphi$).
Furthermore, $I \vdash_\Lambda \varphi$ implies that $I \cup \{\neg\varphi\} \vdash_\Lambda \varphi$, and since, $I \cup \{\neg\varphi\} \vdash_\Lambda \neg\varphi$, it follows that $I \cup \{\neg\varphi\} \vdash_\Lambda \bot$, i.e. $I \cup \{\neg\varphi\} : inc\Lambda$.
(iii)
If I is an inconsistent theory, then there are $n > 0$ and $\varphi_1, \ldots, \varphi_n \in I$ s.t. $\varphi_1 \wedge \ldots \wedge \varphi_n \supset \bot \in \mathbf{PC}$, hence, $\vdash_\Lambda \varphi_1 \wedge \ldots \wedge \varphi_n \supset \bot$, i.e., I is $\varnothing inc\Lambda$. ∎

Lemma A.6 *Let S be any class of structures (frames or models) and suppose that for every $\varnothing c\Lambda$-theory I, there is a $\mathfrak{S} \in \mathsf{S}$, in which I is satisfiable. Then, Λ is strongly complete with respect to the class of structures S.* [4]

PROOF. The contrapositive will be proved. So, assume that Λ is not strongly complete with respect to S, i.e. there are $I \subseteq \mathcal{L}_\Box$, $\varphi \in \mathcal{L}_\Box$ s.t. $I \Vdash_{\mathsf{S}} \varphi$ and $I \nvdash_\Lambda \varphi$. Then, by Lem.A.5(ii), $I \cup \{\neg\varphi\}$ is a $\varnothing c\Lambda$-theory. Furthermore, let $\mathfrak{S}$ be any structure from S and suppose, for the sake of contradiction, that there is a world w in $\mathfrak{S}$ s.t. $\mathfrak{S}, w \Vdash I \cup \{\neg\varphi\}$. Hence, $\mathfrak{S}, w \Vdash \neg\varphi$ and $\mathfrak{S}, w \Vdash I$, but, since $I \Vdash_{\mathsf{S}} \varphi$, $\mathfrak{S}, w \Vdash \varphi$, which is a contradiction. Consequently, in all structures of S, $I \cup \{\neg\varphi\}$ is not satisfiable. ∎

And now we come to the main results.

Theorem A.7 (Soundness)

$$(\forall \Gamma \cup \{\varphi\} \subseteq \mathcal{L}_\Box)(\Gamma \vdash_{\mathbf{S5'}_R} \varphi \Rightarrow \Gamma \Vdash^g_{\mathsf{U_q}} \varphi)$$ [5]

PROOF. By Lem.A.2, it suffices to show (since the rest is nearly obvious) that **T** and $\mathbf{5}_\top$ are valid in any q-frame $\mathfrak{F} = \langle W, N, R \rangle$ s.t. $(\forall w \in N)$ $(\forall v \in W) wRv$. But, such a frame is reflexiv in N and property $(\mathsf{E_q})$ holds, so, by Prop.A.3, **T** and $\mathbf{5}_\top$ are valid. ∎

Proposition A.8 *Let Λ be a consistent regular modal logic. If $\mathfrak{M}^{\Lambda,\varnothing}$ belongs to a class S of structures, then Λ is strongly complete with respect to S.*

PROOF. By Lem.A.5(i), $\varnothing$ is $c\Lambda$, so, $\mathfrak{M}^{\Lambda,\varnothing}$ does exist. Let I be a $\varnothing c\Lambda$-theory. Then, by Lindenbaum's lemma, there is a $m\varnothing c\Lambda$-theory r s.t. $I \subseteq \mathsf{r}$. Hence, by Lem.3.8, $\mathfrak{M}^{\Lambda,\varnothing}, \mathsf{r} \Vdash I$, so, since $\mathfrak{M}^{\Lambda,\varnothing}$ belongs to S, by Lem.A.6, Λ is strongly complete with respect to S. ∎

[4] i.e. $(\forall I \subseteq \mathcal{L}_\Box)(\forall \varphi \in \mathcal{L}_\Box)(I \Vdash_{\mathsf{S}} \varphi \Rightarrow I \vdash_\Lambda \varphi)$. $I \Vdash_{\mathsf{S}} \varphi$ means local semantic consequence, i.e. $(\forall \mathfrak{M} = \langle W, R, V \rangle \in \mathsf{S})(\forall w \in W)(\mathfrak{M}, w \Vdash I \Rightarrow \mathfrak{M}, w \Vdash \varphi)$.

[5] If S is a class of frames, $\Gamma \Vdash^g_{\mathsf{S}} \varphi$ means global semantic consequence, i.e. $(\forall \mathfrak{F} \in \mathsf{S})(\mathfrak{F} \Vdash \Gamma \Rightarrow \mathfrak{F} \Vdash \varphi)$.

Theorem A.9 $\mathbf{S5'}_R$ *is strongly complete with respect to all q-frames, for which* $(\mathsf{ER_q})$ *holds.*

Proof. By Prop.A.8, it suffices to show that $(\mathsf{ER_q})$ holds for canonical frame $\mathfrak{F}^{\mathbf{S5'}_R,\varnothing}$. Hence, by Prop.A.3 and Corol.A.4, it suffices to show that $\mathfrak{F}^{\mathbf{S5'}_R,\varnothing}$ is reflexiv in $N^{\mathbf{S5'}_R,\varnothing}$ and that property $(\mathsf{E_q})$ holds for $\mathfrak{F}^{\mathbf{S5'}_R,\varnothing}$. For simplicity, let us denote as $\mathfrak{F}' = \langle W', N', R' \rangle$ the canonical frame $\mathfrak{F}^{\mathbf{S5'}_R,\varnothing}$.
For reflexivity.
Let Γ be a $m\varnothing c\mathbf{S5'}_R$-theory and $\varphi \in \mathcal{L}_\Box$ s.t. $\Box\varphi \in \Gamma$. But, $\varnothing \vdash_{\mathbf{S5'}_R} (T)$, hence, by Lem.3.3(iii), $\Box\varphi \supset \varphi \in \Gamma$, and, by Lem.3.3(i), $\varphi \in \Gamma$. So, by Def.3.6(iii), $\Gamma R' \Gamma$.
For $(\mathsf{E_q})$
Let Γ, Δ, Θ be $m\varnothing c\mathbf{S5'}_R$-theories s.t. $\Box\top \in \Gamma, \Delta$ and $\Gamma R' \Delta$, $\Gamma R' \Theta$. Let, furthermore, be any $\varphi \in \mathcal{L}_\Box$ s.t. $\Box\varphi \in \Delta$. Suppose that $\Box\neg\Box\varphi \in \Gamma$. Then, since $\Gamma R' \Delta$, $\neg\Box\varphi \in \Delta$, which is a contradiction, since Δ is, by Lem.A.5(iii), consistent. So, $\Box\neg\Box\varphi \notin \Gamma$, hence, by Lem.3.3(ii), $\neg\Box\neg\Box\varphi \in \Gamma$. But, by Lem.A.2, $\vdash_{\mathbf{S5'}_R} (5')$, hence, by Lem.3.3(iii), $\Box\top \wedge \neg\Box\neg\Box\varphi \supset \Box\varphi \in \Gamma$. Furthermore, $\Box\top \in \Gamma$, so, by Lem.3.3(i), $\Box\varphi \in \Gamma$. Finally, since $\Gamma R' \Theta$, $\varphi \in \Theta$. Hence, it has been proved that, if $\Box\varphi \in \Delta$, then $\varphi \in \Theta$, so, by Def.3.6(iii), $\Delta R' \Theta$. ∎

The result in the previous theorem can be proved for another, simpler class of q-frames, by introducing and using generated q-submodels. They are defined in the obvious way, but by ommiting R-edges starting from impossible worlds.

Definition A.10 *Let* $\mathfrak{M} = \langle W, N, R, V \rangle$, $\mathfrak{M}' = \langle W', N', R', V' \rangle$ *be two q-models.* $\mathfrak{M}'$ *is called a* generated q-submodel *of* $\mathfrak{M}$ *(in symbols:* $\mathfrak{M}' \rightarrowtail \mathfrak{M}$*) iff*

- $W' \subseteq W$
- $N' = N \cap W'$
- $R' = R \cap (N' \times W')$
- $(\forall p \in \Phi) V'(p) = V(p) \cap W'$
- $(\forall w \in N')(\forall v \in W)(wRv \Rightarrow v \in W')$

If $D \subseteq W$*, then the smallest generated q-submodel of* $\mathfrak{M}$ *containing* D *is called* the q-submodel of $\mathfrak{M}$ generated by D.

The expected fact about modal satisfaction invariance under generated q-submodels, can be easily proved.

Proposition A.11 *If* $\mathfrak{M}' \rightarrowtail \mathfrak{M}$*, then*

$$(\forall \varphi \in \mathcal{L}_\Box)(\forall w \in W')(\mathfrak{M}, w \Vdash \varphi \iff \mathfrak{M}', w \Vdash \varphi)$$

Finally, using Theor.A.9 and Prop.A.11 one can prove the following result.

Corollary A.12 (Completeness)
$\mathbf{S5'}_R$ *is strongly complete with respect to all q-frames, for which* $(\mathsf{U_q})$ *holds, i.e.*

$$(\forall \Gamma \cup \{\varphi\} \subseteq \mathcal{L}_\Box)(\Gamma \Vdash_{\mathsf{U_q}} \varphi \Rightarrow \Gamma \vdash_{\mathbf{S5'}_R} \varphi)$$

Proof. Firstly, let us denote as $\mathsf{S_U}$ the class of all q-frames, for which $(\mathsf{U_q})$ holds and as $\mathsf{S_{ER}}$ the class of all q-frames, for which $(\mathsf{ER_q})$ holds. Now, let $\Gamma \subseteq \mathcal{L}_\Box$ and $\varphi \in \mathcal{L}_\Box$ s.t. $\Gamma \Vdash_{\mathsf{S_U}} \varphi$. Furthermore, assume any $\mathfrak{F} = \langle W, N, R\rangle \in \mathsf{S_{ER}}$, any $V : \Phi \to \mathcal{P}(W)$ and any $w \in W$ s.t. $\langle \mathfrak{F}, V\rangle, w \Vdash \Gamma$. Let now $\mathfrak{M}' = \langle W', N', R', V'\rangle$ be the q-submodel of $\langle \mathfrak{F}, V\rangle$ generated by $\{w\}$. If $w \notin N$, then, by Def.A.10, $W' = \{w\}$ and $N' = \varnothing$, so, $\langle W', N', R'\rangle \in \mathsf{S_U}$. If $w \in N$, then $N' = [w]_R$ and since $\mathfrak{M}'$ is the smallest q-submodel containing $\{w\}$, $(\forall v \in W' \setminus N')(\exists u \in N')uR'v$. So again, since $(\mathsf{ER_q})$ holds for $\mathfrak{F}$, $\langle W', N', R'\rangle \in \mathsf{S_U}$.
But, by Prop.A.11, $\mathfrak{M}', w \Vdash \Gamma$. Hence, since $\langle W', N', R'\rangle \in \mathsf{S_U}$ and $\Gamma \Vdash_{\mathsf{S_U}} \varphi$, $\mathfrak{M}', w \Vdash \varphi$. Consequently, again by Prop.A.11, $\langle \mathfrak{F}, V\rangle, w \Vdash \varphi$.
Hence, it has been proved that $\Gamma \Vdash_{\mathsf{S_{ER}}} \varphi$. So, by Theor.A.9, $\Gamma \vdash_{\mathbf{S5'_R}} \varphi$. ∎

Proof of Lemma A.1
Firstly, recall that regular modal logic is any set of formulae, containing all propositional tautologies (**Taut**) and axiom K (i.e. the formula $\Box p \wedge \Box(p \supset q) \supset \Box q$), and which is closed under Modus Ponens (**MP**), uniform substitution (**US**) and rule **RM**. Furthermore, given formulae (axioms) $A_1, \ldots, A_n$, the set

$$\bigcap\{\Lambda \subseteq \mathcal{L}_\Box \mid \Lambda : \text{ regular modal logic and } A_1, \ldots, A_n \in \Lambda\}$$

is the smallest regular modal logic containing $A_1, \ldots, A_n$, and it is denoted as $\mathbf{KA_1 \ldots A_{nR}}$. Recall also, that $\vdash_\Lambda \varphi$ means that there is a Hilbert-style proof, where each step of the proof is either a member of $US(K) \cup US(A_1) \cup \ldots \cup US(A_n) \cup \mathbf{PC}$ or a result of applying **MP** or **RM** to formulae of previous steps and where the last formula in this proof is φ.

Now, let us define recursively the following sequence of sets

$$\begin{aligned}
\Lambda_0 &= \{K, A_1, \ldots, A_n\} \cup \mathbf{Taut} \\
\Lambda_{n+1} &= \Lambda_n \cup \Lambda_{n+1}^{\mathbf{MP}} \cup \Lambda_{n+1}^{\mathbf{US}} \cup \Lambda_{n+1}^{\mathbf{RM}}, \text{ where} \\
&\quad \Lambda_{n+1}^{\mathbf{MP}} = \{\varphi \in \mathcal{L}_\Box \mid \psi, \psi \supset \varphi \in \Lambda_n\}, \\
&\quad \Lambda_{n+1}^{\mathbf{US}} = \{\varphi[\varphi_0/p_0, \ldots, \varphi_k/p_k] \in \mathcal{L}_\Box \mid \\
&\qquad\qquad \varphi \in \Lambda_n, k \in \mathbb{N}, \varphi_0, \ldots, \varphi_k \in \mathcal{L}_\Box, p_0, \ldots, p_k \in \Phi\}, \\
&\quad \Lambda_{n+1}^{\mathbf{RM}} = \{\Box\varphi \supset \Box\psi \in \mathcal{L}_\Box \mid \varphi \supset \psi \in \Lambda_n\} \qquad (n \in \mathbb{N})
\end{aligned}$$

and set $\Lambda = \bigcup_{n \in \mathbb{N}} \Lambda_n$. Then it follows, by a trivial induction, that $\Lambda \subseteq \mathbf{KA_1 \ldots A_{nR}}$, and by observing that Λ is a regular modal logic containing $A_1, \ldots, A_n$, that $\mathbf{KA_1 \ldots A_{nR}} \subseteq \Lambda$. Therefore, $\mathbf{KA_1 \ldots A_{nR}} = \Lambda$. Hence, to prove Lemma A.1, it suffices to show that

$$(\forall \varphi \in \mathcal{L}_\Box)(\varphi \in \Lambda \iff \vdash_\Lambda \varphi)$$

Proof. $(\Rightarrow)$
We will firstly show, by induction, that

$$(\forall k \in \mathbb{N})(\forall \varphi \in \mathcal{L}_\Box)((\varphi \in \Lambda_k \wedge k = \min\{n \in \mathbb{N} \mid \varphi \in \Lambda_n\}) \Rightarrow \vdash_\Lambda \varphi) \qquad (*)$$

Ind.Base is trivial, since $\varphi \in \Lambda_0$ implies $\vdash_\Lambda \varphi$. Supposed the statement is true $\forall i \leq k$, we continue with Ind.Step. Let $\varphi \in \mathcal{L}_\Box$ s.t. $\varphi \in \Lambda_{k+1}$ and $(\forall i \leq k)\varphi \notin \Lambda_i$. Since $\varphi \notin \Lambda_k$, there are three cases left:

- If $\varphi \in \Lambda_{k+1}^{\mathbf{MP}}$, then $\psi, \psi \supset \varphi \in \Lambda_k$. Now, let us define $i = \min\{n \in \mathbb{N} \mid \psi \in \Lambda_n\}$ and $j = \min\{n \in \mathbb{N} \mid \psi \supset \varphi \in \Lambda_n\}$. Then, $i, j \leq k$, hence by Ind.Hypothesis, $\vdash_\Lambda \psi$ and $\vdash_\Lambda \psi \supset \varphi$, hence, $\vdash_\Lambda \varphi$.

- If $\varphi \in \Lambda_{k+1}^{\mathbf{RM}}$, then $\varphi = \Box\psi \supset \Box\chi$ and $\psi \supset \chi \in \Lambda_k$. Now, let us define $i = \min\{n \in \mathbb{N} \mid \psi \supset \chi \in \Lambda_n\}$. Then, $i \leq k$, hence, by Ind.Hypothesis, $\vdash_\Lambda \psi \supset \chi$, so, $\vdash_\Lambda \Box\psi \supset \Box\chi$, i.e. $\vdash_\Lambda \varphi$.

- If $\varphi \in \Lambda_{k+1}^{\mathbf{US}}$, then $\varphi = \psi[\varphi_0/p_0, \ldots, \varphi_n/p_n]$, where $\psi \in \Lambda_k$ (and $\varphi_0, \ldots, \varphi_n \in \mathcal{L}_\Box$, $p_0, \ldots, p_n \in \Phi$). Let us define $i = \min\{n \in \mathbb{N} \mid \psi \in \Lambda_n\}$. Then, $i \leq k$.
  - If $i = 0$, then $\psi \in \Lambda_0$, hence, $\varphi \in US(K) \cup US(A_1) \cup \ldots \cup US(A_n) \cup \mathbf{PC}$, so, $\vdash_\Lambda \varphi$.
  - If $i > 0$, then $\psi \notin \Lambda_{i-1}$ and if, ad absurdum, $\psi \in \Lambda_i^{\mathbf{US}}$, then $\psi = \chi[\psi_0/q_0, \ldots, \psi_m/q_m]$, where $\chi \in \Lambda_{i-1}$ (and $\psi_0, \ldots, \psi_m \in \mathcal{L}_\Box$, $q_0, \ldots, q_m \in \Phi$). But then, $\varphi = \chi[\varphi_0/p_0, \ldots, \varphi_n/p_n, \psi_0/q_0, \ldots, \psi_m/q_m]$, hence, $\varphi \in \Lambda_i^{\mathbf{US}}$, i.e. $\varphi \in \Lambda_i$, which is a contradiction, since $i \leq k$. So, there are only two cases left:
    - If $\psi \in \Lambda_i^{\mathbf{MP}}$, then $\chi, \chi \supset \psi \in \Lambda_{i-1}$, hence, $\chi[\varphi_0/p_0, \ldots, \varphi_n/p_n] \in \Lambda_i^{\mathbf{US}}$ and $(\chi \supset \psi)[\varphi_0/p_0, \ldots, \varphi_n/p_n] \in \Lambda_i^{\mathbf{US}}$, therefore, $\chi[\varphi_0/p_0, \ldots, \varphi_n/p_n] \in \Lambda_i$ and $(\chi \supset \psi)[\varphi_0/p_0, \ldots, \varphi_n/p_n] \in \Lambda_i$. Now, let $s = \min\{n \in \mathbb{N} \mid \chi[\varphi_0/p_0, \ldots, \varphi_n/p_n] \in \Lambda_n\}$ and $t = \min\{n \in \mathbb{N} \mid (\chi \supset \psi)[\varphi_0/p_0, \ldots, \varphi_n/p_n] \in \Lambda_n\}$. Then, $s, t \leq i \leq k$, hence, by Ind.Hypothesis, $\vdash_\Lambda \chi[\varphi_0/p_0, \ldots, \varphi_n/p_n]$ and $\vdash_\Lambda (\chi \supset \psi)[\varphi_0/p_0, \ldots, \varphi_n/p_n]$, therefore, since, $(\chi \supset \psi)[\varphi_0/p_0, \ldots, \varphi_n/p_n] = \chi[\varphi_0/p_0, \ldots, \varphi_n/p_n] \supset \psi[\varphi_0/p_0, \ldots, \varphi_n/p_n]$, by **MP**, $\vdash_\Lambda \psi[\varphi_0/p_0, \ldots, \varphi_n/p_n]$, i.e. $\vdash_\Lambda \varphi$.
    - If $\psi \in \Lambda_i^{\mathbf{RM}}$, then $\psi = \Box\chi_0 \supset \Box\chi_1$ and $\chi_0 \supset \chi_1 \in \Lambda_{i-1}$, hence, $(\chi_0 \supset \chi_1)[\varphi_0/p_0, \ldots, \varphi_n/p_n] \in \Lambda_i^{\mathbf{US}}$, therefore, $(\chi_0 \supset \chi_1)[\varphi_0/p_0, \ldots, \varphi_n/p_n] \in \Lambda_i$. Now, let $s = \min\{n \in \mathbb{N} \mid (\chi_0 \supset \chi_1)[\varphi_0/p_0, \ldots, \varphi_n/p_n] \in \Lambda_n\}$. Then, $s \leq i \leq k$, hence, by Ind.Hypothesis, $\vdash_\Lambda (\chi_0 \supset \chi_1)[\varphi_0/p_0, \ldots, \varphi_n/p_n]$, i.e. $\vdash_\Lambda \chi_0[\varphi_0/p_0, \ldots, \varphi_n/p_n] \supset \chi_1[\varphi_0/p_0, \ldots, \varphi_n/p_n]$, therefore, by **RM**, $\vdash_\Lambda \Box\chi_0[\varphi_0/p_0, \ldots, \varphi_n/p_n] \supset \Box\chi_1[\varphi_0/p_0, \ldots, \varphi_n/p_n]$, so, $\vdash_\Lambda (\Box\chi_0 \supset \Box\chi_1)[\varphi_0/p_0, \ldots, \varphi_n/p_n]$, hence, $\vdash_\Lambda \psi[\varphi_0/p_0, \ldots, \varphi_n/p_n]$, i.e. $\vdash_\Lambda \varphi$.

The inductive proof of $(*)$ is complete. Assume now any $\varphi \in \Lambda$. Then, $(\exists n \in \mathbb{N})\varphi \in \Lambda_n$ and for $k = \min\{n \in \mathbb{N} \mid \varphi \in \Lambda_n\}$ result $(*)$ is applicable, hence, $\vdash_\Lambda \varphi$.
$(\Leftarrow)$
We will show, by induction on the length of proof, that

$$(\forall k \in \mathbb{N})(\forall \varphi \in \mathcal{L}_\Box)(\vdash_\Lambda^k \varphi \Rightarrow \varphi \in \Lambda)^6 \qquad (\star)$$

[6] $\vdash_\Lambda^k$ means an RM-proof in Λ with at most k steps.

For the Ind.Base, if $\vdash^0_\Lambda \varphi$, then $\varphi \in US(K) \cup US(A_1) \cup \ldots \cup US(A_n) \cup \mathbf{PC}$, hence, $\varphi \in \Lambda^{\mathbf{US}}_1$, i.e. $\varphi \in \Lambda$. For Ind.Step, let $\vdash^{k+1}_\Lambda \varphi$.

- If $\varphi \in US(K) \cup US(A_1) \cup \ldots \cup US(A_n) \cup \mathbf{PC}$, then exactly as in Ind.Base, $\varphi \in \Lambda$.
- If $(k+1)$-th step is an application of **MP**, then there are $\psi, \psi \supset \varphi \in \mathcal{L}_\Box$ s.t. $\vdash^k_\Lambda \psi$ and $\vdash^k_\Lambda \psi \supset \varphi$, hence, by Ind.Hypothesis, $\psi, \psi \supset \varphi \in \Lambda$, so, since Λ is a modal logic, $\varphi \in \Lambda$.
- If $(k+1)$-th step is an application of **RM**, then $\varphi = \Box\psi \supset \Box\chi$ and $\vdash^k_\Lambda \psi \supset \chi$, consequently, by Ind.Hypothesis, $\psi \supset \chi \in \Lambda$, hence, since Λ is a regular modal logic, $\Box\psi \supset \Box\chi \in \Lambda$, i.e. $\varphi \in \Lambda$.

The inductive proof of $(\star)$ is complete. Now, for any $\varphi \in \mathcal{L}_\Box$, if $\vdash_\Lambda \varphi$, then there is a $k \in \mathbb{N}$ s.t. $\vdash^k_\Lambda \varphi$, hence, by $(\star)$, $\varphi \in \Lambda$. ■

References

[ACP97] G. Amati, L. Carlucci Aiello, and F. Pirri. Intuitionistic autoepistemic logic. *Studia Logica*, 59, 1997.

[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. [GHR93], 1993.

[GHR93] D. M. Gabbay, C. J. Hogger, and J. A. Robinson, editors. *Handbook of Logic in Artificial Intelligence and Logic Programming.* Oxford University Press, 1993.

[Hal97a] J. Halpern. A critical reexamination of default logic, autoepistemic logic and only-knowing. *Computational Intelligence*, 13(1):144–163, 1997. A preliminary version appears in *Proceedings of the Kurt Gödel Symposium, LNCS-713, Springer-Verlag,1993.*

[Hal97b] Joseph Y. 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.

[HM92] J. Y. Halpern and Y. Moses. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54(2):319–379, 1992.

[Jas91] J. Jaspars. A generalization of stability and its application to circumscription of positive introspective knowledge. In E. Börger, H. K. Büning, M. M. Richter, and W. Schönfeld, editors, *Computer Science Logic, 4th Workshop, CSL '90*, volume 533 of *Lecture Notes in Computer Science*, pages 289–299. Springer-Verlag, 1991.

[KZ09] C. Koutras and Y. Zikos. On a modal epistemic axiom emerging from McDermott-Doyle logics. *Fundamenta Informaticae*, 96(1-2):111–125, 2009.

[Len78] W. Lenzen. *Recent Work in Epistemic Logic*. North-Holland, 1978.

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

[MD80] D. McDermott and J. Doyle. Non-monotonic logic I. *Artificial Intelligence*, 13:41–72, 1980.

[Moo85] R. C. Moore. Semantical considerations on non-monotonic logics. *Artificial Intelligence*, 25:75–94, 1985.

[MST93] V. W. Marek, G. F. Schwarz, and M. Truszczyński. Modal non-monotonic logics: Ranges, characterization, computation. *Journal of the ACM*, 40:963–990, 1993.

[MT93] V. W. Marek and M. Truszczyński. *Non-Monotonic 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.