Noname manuscript No.
(will be inserted by the editor)

The Logic of Secrets and the Interpolation Rule

Zuojun Xiong · Thomas Ågotnes

Received: date / Accepted: date

Abstract In this article we formalise the notion of *knowing a secret* as a modality, by combining standard notions of knowledge and ignorance from modal epistemic logic. Roughly speaking, Ann knows a secreet if and only if she knows it and she knows that everyone else does not know it. The main aim is to study the properties of these secretly knowing modalities. It turns out that the modalities are non-normal, and are characterised by a derivation rule we call *Interpolation* that is stronger than *Equivalence* but weaker than *Monotonicity.* We study the Interpolation rule and position it in the landscape of non-normal modal logics. We show that it, in combination with basic axioms, gives us a complete characterisation of the properties of the secretly knowing modalities under weak assumptions about the properties of individual knowledge, in the form of a sound and complete axiomatisation. This characterisation gives us the most basic and fundamental principles of secretly knowing.

1 Introduction

In this article we formalise the notion of *knowing a secret*, in epistemic logic – a standard framework for formal reasoning about information in multi-agent systems. The notion of secrets is fundamental in areas such as safety and security, in particular in cryptography, authentication and access control. What is

Z. Xiong
Institute of Logic and Intelligence, Southwest University
Chongqing, China
E-mail: zuojunxiong@swu.edu.cn

T. Ågotnes
Institute of Logic and Intelligence, Southwest University
Chongqing, China
Department of Information Science and Media Studies, University of Bergen
Bergen, Norway
E-mail: thomas.agotnes@uib.no

a secret? While dictionary definitions of the noun varies somewhat – “a piece of knowledge that is hidden and intended to be kept hidden” (Wiktionary); “a piece of information that is only known by one person or a few people and should not be told to others” (Cambridge Dictionary); “something that is kept or meant to be kept unknown or unseen by others” (Oxford English Dictionary); “something kept from the knowledge of others” (Merriam-Webster) – it is clear that secrets are fundamentally about *knowledge* and *ignorance* (the lack of knowledge). For example, when we say that “Ann keeps her pin code secret” or “Bill has a secret girlfriend” we mean that there is something (Ann’s pin code or the identity of Bill’s girlfriend) that is (1) known by someone (Ann or Bill) and (2) not known by others.

In this paper we formalise the notion of knowing a secret in the standard framework for formalising knowledge, namely epistemic logic (Fagin et al., 1995). We introduce a modality S_a, such that $S_a\varphi$ is intended to mean that φ is a secret that agent a knows/has. We henceforth refer to these modalities as *secretly knowing* modalities, and say that a secretly knows φ[1]. Our goal is to study the properties of secretly knowing that follow from the basic definition based on knowledge and ignorance.

We focus here on the *epistemic* properties of secrets. As discussed above these are quite fundamental, but it should be mentioned that there are other aspects of secrets, such as *intentionality* that we don’t model explicitly. Furthermore, in this paper we focus on formalising a basic notion of secretly knowing: we assume that the secret is exclusively known by a single person. As seen above definitions of secrets also allow for the secret to be known by a (small) number of people. We focus here on the simplest case in order to clarify the basic principles of secretly knowing as much as possible.

Conditions for “agent a secretly knows φ” includes (1) that a knows φ and (2) that any other agent b does not know φ. We argue, however, that the important thing here is not that b actually should not know, but that (2’) *a should know (or believe) that b doesn’t know*. This property of secrets is not explicitly mentioned in the definitions cited above, but it is often implicitly assumed, e.g., in formulations such as *“...intended to be kept hidden”*. Indeed, if Bill believes that other people know who his girlfriend is, or if he merely doesn’t know that they don’t know who his girlfriend is, the identity of his girlfriend wouldn’t typically be called a secret from Bill’s perspective.

We can now use the language of epistemic logic (Fagin et al., 1995), where $K_a\varphi$ intuitively means that agent a knows φ, to express the fact that “agent

[1] This term can be somewhat ambiguous without a precise definition. In this article we use it to mean “knowing” with the added condition that the knowledge must be a secret (not known by others): we say that a secretly knows φ, or has secret knowlege of φ, whenever a knows φ and (she knows that) no one else knows φ. Another interpretation of “a secretly knows φ” would be that others don’t know that a knows φ, but that they themselves might actually know φ – which is not what we have in mind.

a secretly knows φ" as

$$K_a\varphi \wedge K_a\left(\bigwedge_{b\neq a} \neg K_b\varphi\right), \tag{SKs}$$

combining (1) and (2'), or, equivalently in epistemic logic,

$$K_a\left(\varphi \wedge \bigwedge_{b\neq a} \neg K_b\varphi\right). \tag{SKs'}$$

If *veridicality* (if $K_a\varphi$ is true then φ is true) is assumed, as if often is in epistemic logic, then this implies $\bigwedge_{b\neq a} \neg K_b\varphi$ (2) as well. Under the veridicality assumption, (SKs) can be seen as an "objective" definition of secretly-knowing: a knows that φ and that b does not know that φ and both those things are actually true. We are particularly interested in this case. However, (SKs) makes sense also without this assumption, in a "subjective" sense of secretly knowing: a knows (or rather in this case, believes[2]) that φ and that b does not know φ, but neither might actually be the case. In fact, this subjective view is alluded to in some of the definitions above, in formulations like "something that is kept *or meant to be kept* unknown" (our emphasis) – it is not required that the secret *actually is* unknown by others.

Alternatively, it could perhaps be argued that "b not knowing *that* φ" should be replaced with "b does not know *whether* φ (Hart et al., 1996; Fan et al., 2015): $K_a\varphi \wedge K_a\left(\bigwedge_{b\neq a}(\neg K_b\varphi \wedge \neg K_b\neg\varphi)\right)$. Under the veridicality assumption this is actually equivalent to (SKs).

The remainder of the article is organised as follows. In the next section we recall the standard framework of epistemic logic and extend it with new "secretly knowing" modalities S_a such that $S_a\varphi$ captures the meaning discussed above. In Section 3 we study properties of secretly knowing by investigating valid (and non-valid) formulas and (non-) validity-preserving derivation rules, under different assumptions about the properties of knowledge. We look at both interaction properties between standard knowledge modalities K_a and secretly knowing modalities (Section 3.1) and between secretly knowing modalities of different agents (Section 3.2), but perhaps most interesting are properties that involve secretly knowing modalities of only a single agent (Section 3.3) such as distribution over conjunction or preservation of positive introspection – these are the core properties of the secretly knowing modalities. It turns out that the modalities are non-normal, and in Section 4 we attempt to position them in the landscape of non-normal modal logics. In particular, we show that they are characterised by a derivation rule which we call *Interpolation* which

[2] Non-veridical knowledge is sometimes called "belief". We will, however, like in, e.g., the standard text book in epistemic logic Fagin et al. (1995) be somewhat agnostic about the properties of knowledge and use "knowledge" also in the cases that veridicality is not assumed and will henceforth thus not use the word "belief". The properties of knowledge we assume (if any) will be clear from the context.

is stronger than Equivalence but weaker than Monotonicity. A key question is what a *complete* characterisation of the properties of the secretly knowing modalities might be, and in Section 5 we answer that question for a language with only a single secretly knowing modality and the cases that no assumptions about the properties of individual knowledge is assumed (general Kripke models) and that veridicality (reflexive Kripke models) is assumed. These settings represent the most basic and fundamental properties of secretly knowing. We give sound and complete axiomatisations of the logic in these cases, via a translation to an alternative semantics. In Section 6 we conclude and discuss related and future work.

2 Language and Semantics

In order to formally study the logical properties of secretly knowing, we introduce new modalities S_a, such that $S_a\varphi$ means that agent a secretly knows φ in the precise sense defined above. $S_a\varphi$ is, of course, definable in terms of the standard epistemic modalities K_a and K_b, but we introduce it as a primary operator because we are interested not only in the interaction properties between secrets and knowledge, but also in the core principles of secretly knowing. Thus we define the formal language $\mathcal{L}_{SK}$, parameterised by a non-empty set $\texttt{Prop}$ of propositional letters and a finite set $\texttt{Agt}$ of at least two agents as follows:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_a\varphi \mid S_a\varphi$$

where $p \in \texttt{Prop}, a \in \texttt{Agt}$. We use the usual derived propositional connectives, as well as $\widehat{K_a}\varphi$ for $\neg K_a\neg\varphi$ and $\langle S_a\rangle\varphi$ for $\neg S_a\neg\varphi$. We let $\bot$ stand for $p\wedge\neg p$ for some arbitrary $p \in \texttt{Prop}$, and $\top$ stand for $\neg\bot$. The fragment of the language without any S_a operators is called the (purely) *epistemic language*.

For the semantics we use standard Kripke models, as usual in epistemic logic (Fagin et al., 1995). A *(Kripke) model* $M = (W, R, V)$ consists of a set of states W, a function $R : \texttt{Agt} \to 2^{W\times W}$ mapping each agent $a \in \texttt{Agt}$ to a binary relation $R(a)$ on W called a's accessibility relation, and a valuation function $V : W \to 2^{\texttt{Prop}}$. We usually write R_a for $R(a)$. Satisfaction of a formula $\varphi \in \mathcal{L}_{SK}$ in a state w of a model $M = (W, R, V)$ is defined as follows:

$M, w \models p$ iff $w \in V(p)$.
$M, w \models \neg\varphi$ iff $M, w \not\models \varphi$.
$M, w \models \varphi \wedge \psi$ iff $M, w \models \varphi$ and $M, w \models \psi$.
$M, w \models K_a\varphi$ iff $\forall w' \in W$, if wR_aw' then $M, w' \models \varphi$.
$M, w \models S_a\varphi$ iff $\forall w' \in W$, if wR_aw' then $M, w' \models \varphi$ and $\forall b \neq a, \exists u \in W$ s.t. $w'R_bu$ and $M, u \models \neg\varphi$.

The semantics of the K_a modalities is standard. The semantics of $S_a\varphi$ says that it is true if and only if in all states a considers possible not only is φ true, but it is also true that b does not know that. This expresses exactly the formula (SKs) from the introduction, as we make precise in the next section.

It is easy to see that the semantics of the dual of the secretly knowing operator is as follows:

(Prop)		Instances of propositional tautologies	
(K)		$K_a(\varphi \to \psi) \to (K_a\varphi \to K_a\psi)$	Distribution
(MP)		From $(\varphi \to \psi)$ and φ, infer ψ	Modus Ponens
(Nec)		From φ, infer $K_a\psi$	Necessitation
(D)	seriality	$K_a\varphi \to \neg K_a\neg\varphi$	Consistency
(T)	reflexivity	$K_a\varphi \to \varphi$	Veridicality
(4)	transitivity	$K_a\varphi \to K_aK_a\varphi$	Positive Introspection
(5)	euclidicity	$\neg K_a\varphi \to K_a\neg K_a\varphi$	Negative Introspection

Table 1: Axiomatisation in the purely epistemic language of the class of all models (top). Axiomatisation of the class of models with any combination of the properties at the bottom is obtained by adding the corresponding axioms.

$$\begin{array}{lll} M, w \models \langle S_a\rangle\varphi & \text{iff} & \exists w' \in W \text{ such that } wR_aw' \text{ and either } M, w' \models \varphi \\ & & \text{or } \exists b \neq a\ \forall u \in W \text{ if } w'R_bu \text{ then } M, u \models \neg\varphi. \end{array}$$

We will particularly be interested in formulas φ that are *valid*, denoted $\models \varphi$, i.e., formulas that are satisfied in any state in any model. Those are the universally valid principles of (secret) knowledge. Tab. 1 (top) shows a sound and complete axiomatisation of the class of all models in the purely epistemic language. In the definition of that class of models (the standard definition in epistemic logic), there is no restriction on the accessibility relations. In epistemic logic various properties of knowledge are often assumed, such as $K_a\varphi \to K_aK_a\varphi$ (*positive introspection*). There is a well-known correspondence between the most commonly considered properties and properties of the accessibility relations (Fagin et al., 1995), as also shown in Tab. 1.

In this paper we will be particularly interested in two special cases: the class of all reflexive models (T) and the class of all models where the accessibility relations are equivalence relations (S5). The first corresponds to assuming that knowledge is veridical as discussed in the introduction, the second that also all of the other properties in Tab. 1 hold. We write $\models_T \varphi$ and $\models_{S5} \varphi$ to denote that φ is valid in these model classes respectively.

In the next section we look at some valid properties as well as some properties that are not valid.

3 The Properties of Secret Knowledge

In this section we study the properties of the secretly knowing operators in terms of valid formulas and preservation of properties (e.g., positive introspection) assumed of standard knowledge. Of course, since those operators are definable from the knowledge operators, their properties can strictly speaking be completely characterised by a single "axiom", namely

$$S_a\varphi \leftrightarrow \left(K_a\varphi \wedge K_a \bigwedge_{b\in\mathtt{Agt}\setminus\{a\}} \neg K_b\varphi\right),$$

in the sense that all properties of the S_a operators follow from this formula together with the properties of the K_a operators. However, while technically correct that characterisation does not shed much light on intuitively interesting potential properties of secret knowledge such as veridicality or positive introspection.

We now take a closer look at such properties under different assumptions about the properties of the standard knowledge modalities (accessibility relations), and more generally at how secret knowledge operators interact with other operators. First, in Section 3.1 we look at interaction properties between secret knowledge and knowledge, in Section 3.2 we look at interaction properties between secret knowledge operators for two different agents, and in Section 3.3 we look at properties involving secret knowledge operators for only a single agent. The properties studied in this last section are arguably the most fundamental and represent the basic principles of secretly knowing.

In the following we look at formula schemata typically including some otherwise unspecified sub-formulas φ or ψ and agents a or b with implicit universal quantification: when we say that such a schemata is valid, we implicitly mean for *all* such formulas and agents, and correspondingly when we say that it is not valid we mean for *some*. Most of the following propositions follow directly from the semantic definitions and we leave out many of the proofs.

3.1 Secrets and Knowledge

Directly from the semantic definition:

Proposition 1 (Reducibility) $\models S_a\varphi \leftrightarrow \left(K_a\varphi \land K_a\bigwedge_{b\in\mathtt{Agt}\setminus\{a\}} \neg K_b\varphi\right)$

If something is secretly known by a then a knows it and other agents don't, if we assume veridicality:

Proposition 2 (Secret privacy) $\models_T S_a\varphi \to (K_a\varphi \land \neg K_b\varphi)$*, when* $a \neq b$

In fact, assuming veridicality other agents cannot possibly know that a secretly knows φ:

Proposition 3 (Secret unknowability) $\models_T \neg K_b S_a\varphi$*, when* $a \neq b$

Let us now assume that knowledge has all the S5 properties. Conversely to Prop. 3, an agent cannot possibly secretly know that another agent knows φ – other agents' knowledge cannot be a secret.

Proposition 4 (Knowledge no secret) $\models_{S5} \neg S_a K_b\varphi$*, when* $a \neq b$

Similarly, secrets also cannot be about other agents' ignorance:

Proposition 5 (Ignorance no secret) $\models_{S5} \neg S_a \neg K_b\varphi$*, when* $a \neq b$

But an agent can have knowledge about her own secrets. In fact, an agent always has complete knowledge about her own secrets, in the sense that for any formula, she knows whether or not she secretly knows it, as the following shows.

Proposition 6 (Secret negation completeness) $\models_{S5} K_a S_a \varphi \vee K_a \neg S_a \varphi$

A corollary of secret negation completeness is the following negative introspection property of the combination of secret knowledge and knowledge: if you don't secretly know something, then you know that you don't.

Corollary 1 (Negative secret knowledge introspection) $\models_{S5} \neg S_a \varphi \to K_a \neg S_a \varphi$

We also have the corresponding positive introspection property.

Proposition 7 (Positive secret knowledge introspection) $\models_{S5} S_a \varphi \to K_a S_a \varphi$

3.2 Secrets and Others' Secrets

Let us move on to interaction properties between S_a and S_b for two different agents a and b. First, two agents can't share the same secret, assuming veridicality.

Proposition 8 (Secret Exclusivity) $\models_T S_a \varphi \to \neg S_b \varphi$, *when* $a \neq b$

Just like other agents' knowledge nor their ignorance can be a secret assuming S5 properties, neither can other agents' secrets or non-secrets.

Proposition 9 (No secret secrets) $\models_{S5} \neg S_a S_b \varphi$, *when* $a \neq b$

Proposition 10 (No secret non-secrets) $\models_{S5} \neg S_a \neg S_b \varphi$, *when* $a \neq b$

3.3 Secrets of a Single Agent

Finally, we look at properties involving secretly knowing operators for a single agent.

3.3.1 T and S5

Assuming veridicality of knowledge, secretly-knowing satisfies consistency (but not the converse) and in fact also veridicality[3] :

Proposition 11 (Consistency) $\models_T S_a \varphi \to \langle S_a \rangle \varphi$ *but* $\not\models_{S5} \langle S_a \rangle \varphi \to S_a \varphi$

Proof $\models_T S_a \varphi \to \neg S_a \neg \varphi$ follows trivially from $\models S_a \neg \varphi \to K_a \neg \varphi$, $\models_T K_a \neg \varphi \to \neg K_a \varphi$, $\models \neg K_a \varphi \to \neg S_a \varphi$. For $\not\models \neg S_a \neg \varphi \to S_a \varphi$, we offer a counterexample in Figure 1: $M, w \models \neg S_a \neg p \wedge \neg S_a p$. □

[3] Note that $\not\models_{S5} \varphi$ implies $\not\models_T \varphi$ and $\not\models \varphi$. We generally state the non-validities for the *strongest* possible of the cases we consider, and validities for the *weakest* possible cases.

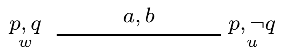


Fig. 1: S5 model M (reflexive edges omitted).

Proposition 12 (Secret veridicality) $\models_T S_a\varphi \to \varphi$

It immediately follows that contradictions cannot be secretly known – since they can't be known.

Corollary 2 (No secret contradictions) $\models_T \neg S_a\bot$

In fact, not even tautologies can be secretly known (assuming veridicality):

Proposition 13 (No secret tautologies) $\models_T \neg S_a\top$

Let us move on to the case that knowledge has the all the S5 properties. In Section 3.1 we saw that in this case, secretly-knowing also satisfies a weak version of positive and negative introspection: if something is secretly (not) known then that fact is known. What about the stronger version, in other words, do the (4) and (5) axioms hold for the S_a modalities? For positive introspection: if something is secretly known, is it then secretly known that it is secretly known? The answer in this case is yes.

Proposition 14 (Positive Secret introspection) $\models_{S5} S_a\varphi \to S_aS_a\varphi$

Positive secret introspection shows that it is possible to secretly know that one secretly knows something – in fact in S5 this holds exactly when one secretly knows *that something* ($S_aS_a\varphi$ holds iff $S_a\varphi$ holds). Is it possible to secretly know that one does *not* secretly know something? The following shows that the answer is yes.

Proposition 15 (Non-secrets can be secrets) $\not\models_{S5} \neg S_a\neg S_a\varphi$

Proof Proof by counterexample. See Figure 2. Let $\varphi = p$. We have $M, w \models \neg K_ap$ and $M, w \models \neg S_ap$. Similarly, we have $M, w' \models \neg S_ap$ from $M, w' \models \neg K_ap$. Therefore, $M, w \models K_a\neg S_ap$. Also $M, w_1 \models \neg S_ap$ from $M, w_1 \models \neg K_ap$. Let $b \neq a$. $M, w_2 \models S_ap$, as $M, w_2 \models K_ap \wedge K_a\neg K_bp$. Thus, $M, w \models \neg K_b\neg S_ap$. Similarly, we also have $M, w' \models \neg K_b\neg S_ap$. Therefore, $M, w \models K_a\neg K_b\neg S_ap$ for any $b \neq a$. Together with $M, w \models K_a\neg S_ap$, we get that $M, w \models S_a\neg S_ap$. $\square$

Negative introspection, however, does not carry over from knowing to secretly-knowing:

Proposition 16 (No negative secret introspection) $\not\models_{S5} \neg S_a\varphi \to S_a\neg S_a\varphi$ *and* $\not\models_{S5} \neg S_a\varphi \to \neg S_a\neg S_a\varphi$

Proof Let $M = (W, R, V)$ and $N = (W', R', V')$ be the two S5 models defined as in Figure 3. Let $\varphi = p$.

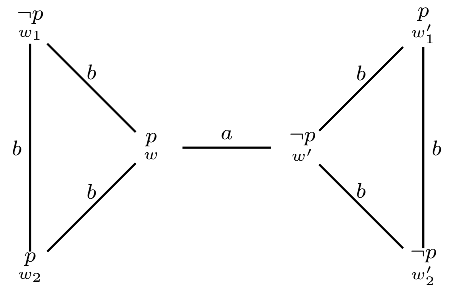


Fig. 2: S5 model M (reflexive edges omitted). Accessibility for all agents different from a is the same and labelled with b.

(a) $M = (W, R, V)$ (b) $N = (W', R', V')$

Fig. 3: S5 models M and N (reflexive edges omitted). Accessibility for all agents different from a is the same and labelled with b.

- Counterexample for $\neg S_a p \to S_a \neg S_a p$. From $M, w \models \neg K_a p$, $M, w \models \neg S_a p$. From $M, u \models \neg K_a p$, $M, u \models \neg S_a p$. That gives us $M, u \models K_a \neg S_a p$. Let $b \neq a$. Since $M \models \neg S_a p$, we have $M \models K_b \neg S_a p$, $M \models K_a K_b \neg S_a p$. That means $M, w \not\models K_a \neg K_b \neg S_a p$, and $M, w \models \neg S_a \neg S_a p$ by semantics.
- Counterexample for $\neg S_a p \to \neg S_a \neg S_a p$. We have $N, w' \models \neg S_a p$, since $N, w' \models \neg K_a p$. Also, $N, u' \models S_a p$ since $N, u' \models K_a p \wedge K_a \neg K_b p$ for any $b \neq a$. It follows that $N, w' \models \neg K_b \neg S_a p$ for any $b \neq a$ and $N, w' \models K_a \neg S_a p$. From $N, w' \models \neg K_b \neg S_a p$, we have $N, w' \models K_a \neg K_b \neg S_a p$. By semantics, $N, w' \models S_a \neg S_a p$. Together with $N, w' \models \neg S_a p$, we have $N, w' \not\models \neg S_a p \to \neg S_a \neg S_a p$.

□

To sum up the case that accessibility relations are equivalence relations we have seen that the S_a modalities satisfy most of the mentioned properties of S5 knowledge: consistency, veridicality, positive introspection – but not negative introspection. In the terminology of Ågotnes and Wáng (2021a): veridicality and positive introspection are *preserved* when going from knowledge to secret knowledge, while negative introspection is not. This is similar to the case of *general (group) knowledge* (everybody-knows), for which negative introspection is also not preserved (Ågotnes and Wáng, 2021a). It is also worth pointing out again that we do however have the weaker form of negative introspection in Corollary 1.

3.3.2 Basic properties: K

Let us move on to basic properties, validities that hold on the class of *all* models. First, like K_a, S_a distributes over implication:

Proposition 17 (Secret distribution) $\models S_a(\varphi \to \psi) \to (S_a\varphi \to S_a\psi)$

Proof Let $M, w \models S_a(\varphi \to \psi)$ and $M, w \models S_a\varphi$. We show that $M, w \models S_a\psi$. Proof by contradiction. Assume that $M, w \models \neg S_a\psi$, then there is $w' \in W$ such that wR_aw'

(1) but $M, w' \models \neg\psi$, or
(2) there is a $b \neq a \in \texttt{Agt}$ such that for all $u \in W$ if $w'R_bu$, then $M, u \models \psi$.

If (1) is the case, then from $M, w \models S_a\varphi$, we have $M, w' \models \varphi$. It follows that $M, w' \models \neg(\varphi \to \psi)$, and $M, w \not\models S_a(\varphi \to \psi)$, a contradiction. If (2) is the case, then (3) $M, u \models \varphi \to \psi$ for any $u \in W$ with $w'R_bu$. From $M, w \models S_a(\varphi \to \psi)$ and wR_aw', i.e., (for any $b \neq a$) there exists $u \in W$ such that $w'R_bu$ and $M, u \models \neg(\varphi \to \psi)$, contradicting (3). □

The secretly knowing modalities also distribute over conjunction in one direction, the so-called (C) axiom is valid:

Proposition 18 (Secret combination) $\models (S_a\varphi \wedge S_a\psi) \to S_a(\varphi \wedge \psi)$

Proof Proof by contradiction. Let $M, w \models S_a\varphi \wedge S_a\psi$ but $M, w \not\models S_a(\varphi \wedge \psi)$. Then there exists a $w' \in W$ such that wR_aw'

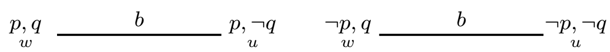
(1) but $M, w' \models \neg(\varphi \wedge \psi)$, or
(2) there is a $b \neq a \in \texttt{Agt}$ such that for all $u \in W$ if $w'R_bu$, then $M, u \models \varphi \wedge \psi$.

Clearly, (1) is not possible, since $M, w \models S_a\varphi \wedge S_a\psi$ means that $M, w' \models (\varphi \wedge \psi)$ by wR_aw'. Let b be as in (2). From $M, w \models S_a\varphi \wedge S_a\psi$, we know that for some $u \in W$ $w'R_bu$ and $M, u \models \neg\varphi$, and for some $u' \in W$ $w'R_bu'$ and $M, u' \models \neg\psi$. A contradiction. □

What about the other direction? If you secretly know $p \wedge q$, do you secretly know p? The answer is in fact "no" – the monotonicity axiom (M) does not hold:

Proposition 19 (Secrets are not monotonic) $\not\models_{S5} S_a(\varphi \wedge \psi) \to S_a\varphi$

Proof See Figure 4(a):



(a) $M = (W, R, V)$ (b) $N = (W, R, V)$

Fig. 4: S5 models M and N (reflexive arrows omitted).

$M, w \models S_a(p \wedge q)$, since $M, w \models K_a(p \wedge q)$ and $M \models \neg K_b(p \wedge q)$. But by $M, w \models K_bp$, $M, w \not\models S_ap$. □

Intuitively, monotonicity fails because secretly knowing requiries a combination of knowledge (of a) and ignorance (of b) and while the former is monotonic, the latter is not: if a secretly knows $p \wedge q$, it might be the case that she knows p but not *secretly* – b might know p even though she doesn't know $p \wedge q$.

Thus, the secretly-knowing modalities are not *normal* modalities. We also immediately see that they also don't distribute over disjunction: $\not\models S_a q \rightarrow S_a(p \vee q)$ (Figure 4(a): $M, w \models S_a q \wedge \neg S_a(p \vee q)$) and $\not\models S_a(p \vee q) \rightarrow S_a p$ (Figure 4(b): $N, w \models S_a(p \vee q) \wedge \neg S_a p$).

We saw that in the T (and S5) case there is at least one "new" validity that is not an instance of a T validity, namely $\neg S_a \top$. This is not guaranteed to hold if veridicality is not assumed, because then it can be that an agent has *no* accessible states in some state in which case $S_a \top$ will hold. But in that case also $S_a \bot$ will hold, so we have the following new general validity which we in the following will refer to as (S).

Proposition 20 (S)

$$\models S_a \top \leftrightarrow S_a \bot \tag{S}$$

Proof $M, w \models S_a \top$, iff w is a dead-end for a, iff $M, w \models S_a \bot$. □

While $\neg S_a \top$ is not valid in the general case, it is of course *satisfiable* – $S_a \top$ is not valid. In fact, we do actually have that $\not\models S_a \varphi$ for *any* φ – there are *no* tautological secrets.

Proposition 21 (No tautological secrets) *For any φ, $\not\models_{S5} S_a \varphi$*

Proof Let φ be an arbitrary formula and assume that $\models S_a \varphi$. Let M, s be an arbitrary pointed S5 model. $M, s \models S_a \varphi$. By reflexivity of R_a, $M, s \models \varphi$ and there is a state t and $b \neq a$ such that $sR_b t$ and $M, t \models \neg \varphi$. But since $\models S_a \varphi$, $M, t \models \varphi$, a contradiction. □

In particular, the (normal) Necessitation rule (from φ infer $S_a \varphi$) does not hold. In Section 4 we look more closely at which sub-normal rules the S_a modalities do and do not satisfy. But let us first consider another pertinent issue.

3.3.3 Is it a box or a diamond?

We argued above that the S_a modalities are not normal – not normal *box* modalities that is: they don't have the properties of normal box modalities (see, e.g., Blackburn et al. (2001)). However, given the exists-forall flavour of the semantics, it is not obvious that they should be viewed as boxes and not as diamonds. And indeed, if we instead view $\langle S_a \rangle$ as the box (despite our notation), then necessitation actually holds. The same is true even for $\neg S_a$.

Proposition 22 (Negative necessitation and diamond necessitation) *If $\models \varphi$ then $\models \neg S_a \varphi$. If $\models \varphi$ then $\models \langle S_a \rangle \varphi$.*

Proof Let $\models \varphi$. It follows that $\models K_b\varphi$ for any b, from which it immediately follows that $S_a\varphi$ is unsatisfiable. It also follows that $\models \neg K_a\neg\varphi$, from which it follows that $\neg S_a\neg\varphi$. □

So are we just viewing the modalities the wrong way, could it be that actually $\langle S_a\rangle$ is a normal box? The answer is "no", but this time for a different reason: neither of the two mentioned variants satisfy the (K) axiom.

Proposition 23 (No K-axiom of $\langle S_a\rangle$) $\not\models_{S5} \langle S_a\rangle(\varphi \to \psi) \to (\langle S_a\rangle\varphi \to \langle S_a\rangle\psi)$

Proof Let $\varphi = p, \psi = q$. See Figure 5. We have $M, w \models \neg S_a\neg(p \to q)$ since $M, w \models \neg K_a\neg(p \to q)$. Also, $M, w \models \neg S_a\neg p$, since $M, w \models \neg K_a\neg p$. But we have $M, w \models S_a\neg q$ since $M, w \models K_a\neg q$ and $M \models \neg K_b\neg q$ for any $b \neq a$. Thus, $M, w \not\models \neg S_a\neg q$, which means $\not\models_{S5} \neg S_a\neg(p \to q) \to (\neg S_a\neg p \to \neg S_a\neg q)$. □

$$\begin{tikzcd} \underset{w'}{p,q} \arrow[r, dash, "b"] & \underset{w}{p,\neg q} \arrow[r, dash, "a"] & \underset{u}{\neg p,\neg q} \arrow[r, dash, "b"] & \underset{u'}{p,q} \end{tikzcd}$$

Fig. 5: S5 model M (reflexive arrows omitted). Accessibility for all agents different from a is the same and labelled with b.

Proposition 24 (No K-axiom of $\neg S_a$) $\not\models_{S5} \neg S_a(\varphi \to \psi) \to (\neg S_a\varphi \to \neg S_a\psi)$

Proof Let $\varphi = p, \psi = q$. See Figure 6. We have $M, w \models \neg S_a(p \to q)$ since $M, w \models \widehat{K_a}K_b(p \to q)$ from $M, u \models K_b(p \to q)$. Also, $M, w \models \neg S_a p$, since $M, w \models \neg K_a p$. But we have $M, w \models S_a q$ since $M, w \models K_a q$ and $M \models \neg K_b q$ for any $b \neq a$. Thus, $M, w \not\models \neg S_a q$, which means $\not\models_{S5} \neg S_a(p \to q) \to (\neg S_a p \to \neg S_a q)$. □

$$\begin{tikzcd} \underset{w'}{p,\neg q} \arrow[r, dash, "b"] & \underset{w}{p,q} \arrow[r, dash, "a"] & \underset{u}{\neg p,q} \arrow[r, dash, "b"] & \underset{u'}{\neg p,\neg q} \end{tikzcd}$$

Fig. 6: S5 model M (reflexive arrows omitted). Accessibility for all agents different from a is the same and labelled with b.

4 Between Monotonicity and Equivalence: the Interpolation Rule

It is natural to ask, then, what about other, weaker, rules, known from other non-normal modal logics? Let us first consider the *monotonicity rule* (*Rm*): from $\varphi \to \psi$ derive $S_a\varphi \to S_a\psi$. This does not hold for S_a (which should come as little surprise given Proposition 19), neither does it hold for $\neg S_a$, $S_a\neg$ or $\langle S_a\rangle$.

Proposition 25 (Non-monotonicity)

- *There are* φ *and* ψ *such that* $\models \varphi \to \psi$, *but* $\not\models_{S5} S_a\varphi \to S_a\psi$
- *There are* φ *and* ψ *such that* $\models \varphi \to \psi$, *but* $\not\models_{S5} \neg S_a\varphi \to \neg S_a\psi$
- *There are* φ *and* ψ *such that* $\models \varphi \to \psi$, *but* $\not\models_{S5} S_a\neg\varphi \to S_a\neg\psi$
- *There are* φ *and* ψ *such that* $\models \varphi \to \psi$, *but* $\not\models_{S5} \neg S_a\neg\varphi \to \neg S_a\neg\psi$

Proof Proof by counterexamples. For the first, let $\varphi = p$ and $\psi = (q \vee \neg q)$, then clearly $\models p \to (q \vee \neg q)$, $\not\models_{S5} S_a p \to S_a(q \vee \neg q)$ (since $S_a p \wedge \neg S_a(q \vee \neg q)$ are satisfiable). For the second, let $\varphi = (p \wedge q)$, but $\psi = p$, then we have $M, w \not\models \neg S_a(p \wedge q) \to \neg S_a p$ from Figure 7. For the third, let $\varphi = q$ and $\psi = (p \vee \neg p)$, then $\models q \to (p \vee \neg p)$, and from Figure 5, we have $M, w \not\models S_a\neg q \to S_a\neg(p \vee \neg p)$. For the last, let $\varphi = (p \wedge \neg p)$, $\psi = q$, then $\models (p \wedge \neg p) \to q$, and from Figure 5, we have $M, w \not\models \neg S_a\neg(p \wedge \neg p) \to \neg S_a\neg q$. $\square$

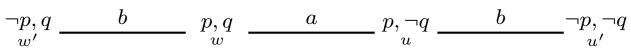


Fig. 7: S5 model M (reflexive edges omitted).

The *equivalence rule* (Re) does, however, hold:

Proposition 26 (Equivalence) *If* $\models \varphi \leftrightarrow \psi$ *then* $\models S_a\varphi \leftrightarrow S_a\psi$.

Proof Let $\models \varphi \leftrightarrow \psi$ and assume that $M, w \models S_a\varphi$ but $M, w \not\models S_a\psi$. From the latter we get that there is a w' such that wR_aw' and either (1) $M, w' \not\models \psi$ or (2) for all u, if $w'R_bu$ then $M, u \models \psi$. From the former we get that $M, w' \models \varphi$ and that there is a u' such that $w'R_bu'$ and $M, u' \models \neg\varphi$. If (1) was the case then $M, w' \models \neg\varphi$ from $\models \varphi \leftrightarrow \psi$, a contradiction. If (2) was the case then $M, u' \models \psi$ and thus $M, u' \models \varphi$ from the assumption that $\models \varphi \leftrightarrow \psi$, also a contradiction. $\square$

We note that the Equivalence rule also preserves validity on T and on S5, which strictly speaking does not follow directly (although the same proof works).

That the equivalence rule holds for S_a modality is of some significance: it means (together with the fact that the logic extends propositional logic) that the logic of S_a extends the weakest non-normal modal logic **E** that has *neighbourhood semantics* (see, e.g., Pacuit (2017)). Given the “forall-exists” Kripke semantics of the secretly knowing modalities, this should perhaps not come as a big surprise.

Let us sum up thus far by positioning the S_a modalities in the landscape of non-normal modal logics: they are *classical* (satisfy the principles of the basic system **E**, namely (Re) in addition to propositional logic) and adjunctive (satisfy (C)) but not *monotonic* or *regular* (do not satisfy (M)). In addition they satisfy (C) and (K) and thus satisfy all the properties of the well-known **ECK** system. In addition, we have seen exactly one additional validity (on the class of all models), namely (S) from Prop. 20. Thus, these modalities seem to be **ECKS** modalities.

In the S5 case (i.e., when accessibility relations are equivalence relations), we also have the (D), (T), (4), ($\top$) and $\neg S_a\bot$, but not the (5), axioms. Both (D) and $\neg S_a\bot$ are derivable using (T)[4], so in this case the modalities seem to be **ECKT4**$\top$ modalities.

The question then is, do the S_a modalities satisfy any other principles? The answer seems to be "yes". Take the following formula:

$$\gamma = (S_a(p \wedge q) \wedge S_a(q \vee r)) \to S_a q$$

It is easy to see that γ is valid. However, it does not seem to be derivable in **ECKS**. Note, in particular, that (see Prop. 19 and related discussion)

$$\not\models S_a(p \wedge q) \to S_a q \qquad \not\models S_a(q \vee r) \to S_a q.$$

Inspecting γ we can see that $\psi = q$ acts as an "interpolant" between $\varphi = p \wedge q$ and $\chi = q \vee r$:

$$\models p \wedge q \to q \qquad \models q \to q \vee r,$$

and that secret knowledge of both φ and χ implies secret knowledge of ψ.

We now arrive at a perhaps surprising key result: *any* such interpolant is secretly known.

Proposition 27 (Interpolation rule)

– *If* $\models \varphi \to \psi$ *and* $\models \psi \to \chi$, *then* $\models (S_a\varphi \wedge S_a\chi) \to S_a\psi$.

Proof Assume that $\models \varphi \to \psi$ and $\models \psi \to \chi$, we need to show that $\models (S_a\varphi \wedge S_a\chi) \to S_a\psi$. Let M, w be an arbitrary pointed model such that $M, w \models S_a\varphi \wedge S_a\chi$, but $M, w \not\models S_a\psi$. Then $M, w \models \langle S_a\rangle\neg\psi$ by semantics, i.e., there exists u such that $wR_a u$ and

(1) $M, u \models \neg\psi$, or
(2) there is a $b \neq a$ such that for any v, if $uR_b v$ then $M, v \models \psi$.

First consider case (2), let b be such an agent. From $wR_a u$ and $M, w \models S_a\varphi \wedge S_a\chi$, we have $M, u \models \varphi \wedge \chi$, and there exists v_1 and v_2 such that $uR_b v_1$, $uR_b v_2$, $M, v_1 \models \neg\varphi$, and $M, v_2 \models \neg\chi$. From $\models \psi \to \chi$, we have $M, v_2 \models \neg\psi$ – a contradiction. Second, consider case (1). We have that $M, u \models \varphi \wedge \chi$ and since $\models \varphi \to \psi$, then get that $M, u \models \psi$ – again a contradiction. Thus, $M, w \models (S_a\varphi \wedge S_a\chi) \to S_a\psi$. $\square$

We call the corresponding inference rule, from $\varphi \to \psi$ and $\psi \to \chi$ infer $(S_a\varphi \wedge S_a\chi) \to S_a\psi$, the *Interpolation rule*[5] (I).

[4] (D): from the (T) instance $S_a\varphi \to \varphi$ we get $\neg\varphi \to \neg S_a\varphi$ which combined with a second (T) instance $S_a\neg\varphi \to \neg\varphi$ gives us $\neg\varphi \to \neg S_a\varphi$ which is the contrapositive of (D). $\neg S_a\bot$: by (T) we have that $S_a\bot \to \bot$ and thus $\neg\bot \to \neg S_a\bot$, and we have $\top = \neg\bot$.

[5] Note that unlike in Craig's interpolation theorem (Craig, 1957), the probably most well-known use of the term "interpolation" in formal logic, there is no assumption about common vocabulary among the involved formulas.

(Prop)	Instances of propositional tautologies
(MP)	From $\varphi \to \psi$ and φ infer ψ
(M)	$\Box(\varphi \wedge \psi) \to (\Box\varphi \wedge \Box\psi)$
(C)	$(\Box\varphi \wedge \Box\psi) \to \Box(\varphi \wedge \psi)$
(K)	$\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$
(N)	$\Box\top$
(Rm)	From $\varphi \to \psi$ infer $\Box\varphi \to \Box\psi$
(Nec)	From φ infer $\Box\varphi$
(Re)	From $\varphi \leftrightarrow \psi$ infer $\Box\varphi \leftrightarrow \Box\psi$
(S)	$\Box\top \leftrightarrow \Box\bot$
(I)	From $\varphi \to \psi$ and $\psi \to \chi$ infer $(\Box\varphi \wedge \Box\chi) \to \Box\psi$

Table 2: Axioms and rules.

.

4.1 Interpolation In the Landscape of Non-Normal Modal Logics

Let us have a closer look at the Interpolation rule (I), and how it relates to other principles commonly considered in the study of non-normal modal logics. We also consider the (S) axiom (see p. 11). In this section we use the standard modal language:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box\varphi$$

where $p \in \texttt{Prop}$. The most commonly used axioms and rules for non-normal modal logics (Pacuit, 2017), most of them already mentioned, together with (I) and the axiom (S), are shown in Table 2.

We follow the standard convention of naming axiomatic systems (see Pacuit (2017)). All the systems we consider are implicity assumed to have (Prop) as axioms and (MP) as a rule. **E** extends this basic system with the (Re) rule, **I** extends the basic system with the (I) rule. **EK** extends **E** with axiom (K), and so on. We will abuse notation and sometimes use **L** for the smallest set of formulas that contains the axioms of **L** and is closed under the rules of **L**. We write $\vdash_{\mathbf{L}} \varphi$ for $\varphi \in \mathbf{L}$ and say that φ is derivable in **L**. When (A) is an axiom schema, $\vdash_{\mathbf{L}} (A)$ means that all instances of (A) are derivable in **L**. We say that a rule is *admissible* in a system **L** if **L** is closed under the rule.

We first show that (Re) is derivable from (I).

Proposition 28 (Re) *is admissible in any system containing* (I).

Proof By taking $\chi = \varphi$, (I) gives us $\Box\varphi \to \Box\psi$ from $\varphi \to \psi$ and $\psi \to \varphi$. Swapping φ and ψ gives us $\Box\psi \to \Box\varphi$. □

One consequence of having the (Re) rule is that uniform substitution holds: from $\psi \leftrightarrow \psi'$ we can derive $\varphi \leftrightarrow \varphi[\psi/\psi']$ where $\varphi[\psi/\psi']$ is the result of replacing some occurrences of ψ with ψ' (Pacuit, 2017).

Next, we show that (I) is derivable from (Rm).

Proposition 29 (I) *is admissible in any system containing* (Rm).

Proof From $\varphi \to \psi$ Monotonicity gives us $\Box\varphi \to \Box\psi$ which implies $(\Box\varphi \wedge \Box\chi) \to \Box\psi$. □

Thus we have, for example, that **ECK** $\subseteq$ **ICK** $\subseteq$ **RmCK**. We now show that both inclusions are strict.

Proposition 30 (I) *is not admissible in* ***EK*** *or* ***ECK***.

Proof We are going to use the fact that **EK** (**ECK**) is sound with respect to standard interpretation in neighbourhood models property (k) (and (c)) as defined below (Lewis, 1974; Surendonk, 1997; Pacuit, 2017; Van De Putte and McNamara, 2021). A neighbourhood model $M = (W, N, V)$ where W and V are as in a (Kripke) model and N maps any state in W to a set of subsets of W. $M, w \models \Box\varphi$ iff $\varphi^M \in N(w)$, where $\varphi^M = \{v \in W : M, v \models \varphi\}$. The mentioned properties are ($\overline{X} = W \setminus X$):

(k) if $X \in N(w)$ and $\overline{X} \cup Y \in N(w)$ then $Y \in N(w)$,
(c) if $X \in N(w)$ and $Y \in N(w)$ then $X \cap Y \in N(w)$.

Consider the following formula:

$$\gamma = (\Box(p \wedge q) \wedge \Box(q \vee r)) \to \Box q$$

which is derivable from $(p \wedge q) \to q$ and $q \to (q \vee r)$ by (I). We are now going to show that there is a neighbourhood model M, having both the (k) and (c) properties, that falsifies γ. That means that γ is not valid on the class of such models, and thus, by soundness it cannot be that it is derivable in either **EK** or **ECK**.

The model $M = (W, N, V)$ is defined as follows. $W = \{w, u, v, s\}$, $N(w) = \{\{w\}, \{w, u, v\}\}$ ($N(u)$, $N(v)$ and $N(s)$ are arbitrary such that (k) and (c) are satisfied). $V(p) = \{w\}$, $V(q) = \{w, u\}$, and $V(r) = \{v\}$. For (c), we have $\{w\} \cap \{w, u, v\} = \{w\} \in N(w)$. Now we show that M has the (k) property. If $X = \{w\}$, then $\overline{X} = \{u, v, s\}$ and $\{u, v, s\} \cup Y \notin N(w)$ for any $Y \subseteq W$, (k) holds trivially. If $X = \{w, u, v\}$, then $\overline{X} = \{s\}$ and $\{s\} \cup Y \notin N(w)$ for any $Y \subseteq W$, (k) holds trivially. We have $(p \wedge q)^M = \{w\} \in N(w)$, $(q \vee r)^M = \{w, u, v\} \in N(w)$, that's $M, w \models \Box(p \wedge q) \wedge \Box(q \vee r)$, but $q^M = \{w, u\} \notin N(w)$, that's $M, w \not\models \Box q$. $\Box$

Proposition 31 (Rm) *is not admissible in* **IC** *or* **ICS**.

Proof Assume that (Rm) is admissible in **ICS** or in **IC**. In both cases all instances of (M) are derivable in **ICS** (see (Pacuit, 2017, Lemma 2.39)). **ICS** is sound with respect to all Kripke models (we prove this formally in the next section), but there are instances of (M) that are not valid on those models (Prop. 19) – a contradiction. $\Box$

Thus, the Interpolation rule is at the same time a strengthening of the Equivalence rule and a weaking of the Monotonicity rule. One difference between Equivalence and Monotonicity, and one that will be important for us, is that in the presence of the (C) axiom the latter can derive the (K) axiom while the former cannot. As Interpolation is "in between", a natural question is on which side it falls on. The following answers that question.

Proposition 32 $\vdash_{\mathbf{IC}} \Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$

Proof

1. $((\varphi \to \psi) \wedge \varphi) \to \psi$	(Prop)
2. $\psi \to (\varphi \to \psi)$	(Prop)
3. $(\Box((\varphi \to \psi) \wedge \varphi) \wedge \Box(\varphi \to \psi)) \to \Box\psi$	(I), $1 + 2$
4. $\Box((\varphi \to \psi) \wedge \varphi) \to (\Box(\varphi \to \psi) \to \Box\psi)$	(Prop), (MP), 3
5. $(\Box(\varphi \to \psi) \wedge \Box\varphi) \to \Box((\varphi \to \psi) \wedge \varphi)$	(C)
6. $(\Box(\varphi \to \psi) \wedge \Box\varphi) \to (\Box(\varphi \to \psi) \to \Box\psi)$	(MP), $4 + 5$
7. $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$	(Prop), (MP), 6

□

Not surprisingly, Prop. 30 holds also if we add the (S) axiom:

Proposition 33 (I) *is not admissible in **ECKS**.*

Proof **ECKS** is sound w.r.t. all neighbourhood models with the (k) and (c) properties (see the proof of Prop. 30) as well as the following property:

(s) $W \in N(w)$ iff $\emptyset \in N(w)$

This follows immediately from the facts that (C) and (K) and (S) are valid on that model class, and that (Re) preserves validity on the model class. Consider again the formula γ and the model M in the proof of Prop. 30 (slightly modified to require that also (s) holds for $N(u)$, $N(v)$ and $N(s)$). Observe that M also has the (s) property. Thus, γ is not valid on the class of (k), (c) and (s) models, so it cannot be derivable. □

Let us now sum up what we know about the Interpolation rule (and the (S) axiom), and position it in the landscape of non-normal modal logics:

$$\begin{array}{ccccccccccccc}
\mathbf{EK} & \subset^{(1)} & \mathbf{ECK} & \subset^{(2)} & \mathbf{ICK} & =^{(3)} & \mathbf{IC} & \subset^{(4)} & \mathbf{RmCK} & =^{(5)} & \mathbf{EMCK} & =^{(6)} & \mathbf{EMC} \\
\subset^{(7)} & & \subset^{(7)} & & \subset^{(7)} & & \subset^{(7)} & & \subset^{(7)} & & \subset^{(7)} & & \subset^{(7)} \\
\mathbf{EKS} & \subseteq & \mathbf{ECKS} & \subset^{(8)} & \mathbf{ICKS} & =^{(3)} & \mathbf{ICS} & \subset^{(4)} & \mathbf{RmCKS} & =^{(5)} & \mathbf{EMCKS} & =^{(6)} & \mathbf{EMCS}
\end{array}$$

(1) Strictness: (Pacuit, 2017, Observation 2.43).
(2) Inclusion: Prop. 28. Strictness: Prop. 30.
(3) Prop. 32.
(4) Inclusion: Prop. 29. Strictness: Prop. 31.
(5) (Pacuit, 2017, Lemma 2.39).
(6) (Pacuit, 2017, Lemma 2.41).
(7) Strictness: all the logics mentioned in the first line (i.e., without explicit mention of the (S) axiom) are included in (the smallest normal logic) **K**. (S) is not derivable in **K** (by soundness w.r.t. Kripke models).
(8) Inclusion: Prop. 28. Strictness: Prop. 33.

In the next section we show that the logic **ICS** is a sound and complete axiomatisation of core properties of the secretly knowing modalities.

5 Completeness of Core Properties

In the previous sections we looked at some properties of the S_a modalities in the form of valid formulas and validity preserving rules, both interaction properties relating an S_a modality to other modalities as well as basic properties involving only S_a modalities for one agent a. The latter are the most interesting in the sense that they represent the basic properties of these modalities. We tried to place these modalities in the landscape of non-normal modalities, in particular we discovered the Interpolation rule and studied it in some detail as a general rule for sub-normal modal logics. The obvious question is: are there any other valid properties? Or do all valid properties follow from the principles (basic properties and rules) we have discovered so far? In this section we show that the answer to the first question in a natural sense is "no", and we prove that by showing that a set of basic principles is *complete*.

Of course, if we consider the full language there is a simple complete characterisation of all valid properties of the S_a modalities: the Reducibility property (Prop. 1) together with the characterisation of the knowledge modalities (Table 1). However, this does not shed much light on interesting properties. Thus, in order to focus on the most basic and fundamental properties of the S_a in this section we will consider a language *without* the individual knowledge modalities K_a, and with only a *single* S_a modality for a fixed agent a (in other words, the basic modal language considered in Section 4.1 where the box is interpreted as a-secretly-knowing). On the semantic side we consider two cases: the class of all models and the class of all reflexive models, leaving more challenging cases such as S5 for future work (see the discussion in Section 6). All valid formulas for this restricted language and this generalised semantics are of course also valid formulas in the general language and under S5 semantics. Indeed, we argue that they represent the *most basic principles of secretly knowing*. Below, we give a complete characterisation of these properties in terms of axioms and rules already discussed in the previous sections.

One final restriction is that we consider the case of only two agents. Given that we only have a single secretly knowing modality, for a single agent, this is not a significant restriction – it doesn't matter how many "other" agents b there are as long as there is at least one. However, the restriction to only one "other" agents makes the technical details less cluttered.

The main results are that the logic **ICS** from Section 4.1 is in fact sound and complete with respect to all models, and its extension **ICST** with the (T) axiom is sound and complete with respect to reflexive models. The rest of this section is organised as follows. We first formally define the restricted language and semantics in Section 5.1. We then define the basic axiomatic system **ICS** in Section 5.2, argue that it is sound and discuss some derivable theorems. In Section 5.3 we introduce an alternative semantics and show that it is equivalent to the Kripke semantics. The main result for **ICS** is found in Section 5.4: completeness with respect to Kripke semantics (via the alternative semantics). Finally, in Section 5.5 we adapt these results to **ICST**.

(Prop)	Instances of propositional tautologies
(C)	$\vdash (S\varphi \wedge S\psi) \to S(\varphi \wedge \psi)$
(S)	$\vdash S\top \leftrightarrow S\bot$
(MP)	From $\vdash (\varphi \to \psi)$ and $\vdash \varphi$, infer $\vdash \psi$
(I)	From $\vdash \varphi \to \psi$ and $\vdash \psi \to \chi$, infer $\vdash (S\varphi \wedge S\chi) \to S\psi$

Fig. 8: The axiomatisation **ICS**, for the $\mathcal{L}_S$ language.

5.1 Language and Semantics

In this section we assume that $\texttt{Agt} = \{a, b\}$. Let the language $\mathcal{L}_S$ be defined by the following grammar:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid S\varphi$$

where $p \in \texttt{Prop}$. Intuitively, $S\varphi$ means the same as $S_a\varphi$: we have a single secretly knowing modality for a single agent, more specifically the agent a. Agent b is "the other" agent. We write $\langle S\rangle$ for $\neg S \neg$. Models are like before; since there are only two agents we write $M = (W, R_a, R_b, V)$. The interpretation of a formula in a state of a model is defined exactly as for the full language, replacing S_a with S.

5.2 ICS

The system **ICS** over the language $\mathcal{L}_S$ is defined (again) in Figure 8. In addition to propositional logic it contains the "secret combination" axiom (C) and the (S) axiom discussed in Section 3.3, and the Interpolation rule. Henceforth, $\vdash \varphi$ means that φ is derivable in the system **ICS**.

Theorem 1 (Soundness) *The axiomatisation* **ICS** *is sound.*

Proof (C) is valid (Prop. 18), (S) is valid (Prop. 20), (I) preserves validity (Prop. 27). The (Prop) and (MP) cases are immediate.

In Section 4.1 we showed several rules and theorems that are admissible and derivable in **ICS**, most notably (K) and (Re). They will be used without further comment in the following. We also consider the following rule and theorems, which will be of technical use in the completeness proof.

Proposition 34 *The following rule* (Es) *is admissible in* **ICS***.*

– *from* $\vdash \varphi$*, infer* $\vdash S\varphi \leftrightarrow S\neg\varphi$*.*

Proof Let $\vdash \varphi$. Then $\vdash \varphi \leftrightarrow \top$, by (Re), $\vdash S\varphi \leftrightarrow S\top$. From (S) and propositional logic, $\vdash S\varphi \leftrightarrow S\bot$. Also, we have $\vdash S\neg\varphi \leftrightarrow S\bot$ by applying (Re) rule to $\vdash \neg\varphi \leftrightarrow \bot$. Then by propositional logic, we have $\vdash S\varphi \leftrightarrow S\neg\varphi$. □

Proposition 35 (Derivable theorems) *The following theorems are derivable in* **ICS***:*

1. $\vdash S(\varphi \wedge \neg\varphi) \to (S\varphi \wedge S\neg\varphi)$
2. $\vdash S(\varphi \vee \neg\varphi) \to (S\varphi \wedge S\neg\varphi)$

Proof For (1) : we have $\vdash (\bot \to \varphi) \leftrightarrow \top$. By (Re), $\vdash S(\bot \to \varphi) \leftrightarrow S\top$, and it follows that $\vdash S\top \to S(\bot \to \varphi)$. By (K), $\vdash S(\bot \to \varphi) \to (S\bot \to S\varphi)$. Then by propositional logic, $\vdash S\top \to (S\bot \to S\varphi)$, i.e., $\vdash S\bot \to (S\top \to S\varphi)$. Therefore, we have $\vdash (S\bot \to S\top) \to (S\bot \to S\varphi)$ by propositional logic. From the (S) axiom, we have $\vdash (S\bot \to S\top)$, then by (MP), $\vdash S\bot \to S\varphi$. It follows that $\vdash S(\varphi \wedge \neg\varphi) \to S\varphi$. Similarly we can derive $\vdash S(\varphi \wedge \neg\varphi) \to S\neg\varphi$ from $\vdash (\bot \to \neg\varphi) \to \top$. It follows that $S(\varphi \wedge \neg\varphi) \to (S\varphi \wedge S\neg\varphi)$ by prop. logic.

For (2): we have $\vdash (\bot \to \varphi) \leftrightarrow \top$. By (Re), $\vdash S(\bot \to \varphi) \leftrightarrow S\top$, so $\vdash S\top \to S(\bot \to \varphi)$. By (K), $\vdash S(\bot \to \varphi) \to (S\bot \to S\varphi)$. Then by propositional logic, $\vdash S\top \to (S\bot \to S\varphi)$. Therefore, we have $\vdash (S\top \to S\bot) \to (S\top \to S\varphi)$ by propositional logic. From the (S) axiom, we have $\vdash (S\top \to S\bot)$, then by (MP), $\vdash S\top \to S\varphi$. It follows that $\vdash S(\varphi \vee \neg\varphi) \to S\varphi$. Similarly, we can derive $\vdash S(\varphi \vee \neg\varphi) \to S\neg\varphi$ from $\vdash (\bot \to \neg\varphi) \to \top$. It follows that $S(\varphi \vee \neg\varphi) \to (S\varphi \wedge S\neg\varphi)$ by propositional logic. □

The following weakening of the (I) rule by the (C) axiom will come in handy (note that the antecedent is equivalent to $(\bigwedge_k S\delta_k \wedge S\delta) \to S\neg\chi$):

Proposition 36 *If* $\vdash \bigwedge_k \delta_k \to \neg\chi$ *and* $\vdash \neg\chi \to \delta$ *then* $\vdash (\bigwedge_k S\delta_k \wedge \langle S\rangle\chi) \to \neg S\delta$.

Proof From rule (I) we have $\vdash (S(\bigwedge_k \delta_k) \wedge S\delta) \to S\neg\chi$, that's (1) $\vdash S(\bigwedge_k \delta_k) \to (S\delta \to S\neg\chi)$. By (C), we have (2) $\vdash (\bigwedge_k S\delta_k) \to S(\bigwedge_k \delta_k)$. (1) and (2) by propositional logic, $\vdash (\bigwedge_k S\delta_k) \to (S\delta \to S\neg\chi)$, then, $\vdash (\bigwedge_k S\delta_k) \to (\langle S\rangle\chi \to \neg S\delta)$, $\vdash (\bigwedge_k S\delta_k \wedge \langle S\rangle\chi) \to \neg S\delta$. □

5.3 Alternative Semantics

In this section we define an alternative semantics for the language, which is shown to be equivalent. This will be used in the completeness proof in the following section.

First, we show that *reflexivity for b* (or the lack of it) cannot be detected by the logical language.

Proposition 37 (b-reflexive-ignorance) *Let $M = (W, R_a, R_b, V)$ be a model, and $M^r = (W, R_a, {R_b}^r, V)$ be the R_b-reflexive model of M: ${R_b}^r = R_b \cup \{(w,w) \mid w \in W\}$. We claim that for any $\varphi \in \mathcal{L}_S$, $M, w \models \varphi$ iff $M^r, w \models \varphi$.*

Proof Induction on $\varphi \in \mathcal{L}_S$. The propositional cases are immediate. We discuss the case $\varphi = \langle S\rangle\psi$.
(Left-to-right). Let $M, w \models \langle S\rangle\psi$, by semantics, there exists $u \in W$ such that wR_au and we have the following two cases:

1. $M, u \models \psi$. It follows $M^r, u \models \psi$ by IH. From wR_au and the definition of M^r, $M^r, w \models \langle S\rangle\psi$.

2. If $M, u \not\models \psi$, then $M, u \models \neg\psi$. It follows that for all $v \in W$, $uR_b v$ implies $M, v \models \neg\psi$ by the semantics of $M, w \models \langle S\rangle\psi$. Also, by IH and reflexivity ($u{R_b}^r u$), $M^r, u \models \neg\psi$. Since ${R_b}^r(u) = R_b(u) \cup \{u\}$. Then for any $v \in W$ $u{R_b}^r v$ implies that either $uR_b v$ and $M, v \models \neg\psi$ (follows by the semantic consequence of the assumption), or $u = v$ and $M, v \models \neg\psi$ (follows by the assumption). It shows that for any $v \in W$, $u{R_b}^r v$ implies $M, v \models \neg\psi$. Then by semantics and the definition of M^r, $M^r, w \models \langle S\rangle\psi$.

(Right-to-left). Let $M^r, w \models \langle S\rangle\psi$, there exists $u \in W$ such that $wR_a u$ and we have the following two cases:

1. $M^r, u \models \psi$. Then we have $M, u \models \psi$ by IH. $M, w \models \langle S\rangle\psi$ by semantics.
2. If $M^r, u \not\models \psi$, then $M^r, u \models \neg\psi$. It follows that for all $v \in W$, if $u{R_b}^r v$, then $M^r, v \models \neg\psi$. Here, if $uR_b v$, we have $u{R_b}^r v$, then $M, v \models \neg\psi$ by IH. If not $uR_b v$, then we have $u = v$ from $u{R_b}^r v$ and $M^r, v \models \neg\psi$ from assumption, that's $M, v \models \neg\psi$ by IH. Either way, if $uR_b v$, then $M, v \models \neg\psi$. It follows that $M, w \models \langle S\rangle\psi$ by semantics.

Therefore, we have $M, w \models \langle S\rangle\psi$. □

Second, we show that satisfaction is also invariant if we remove any access for b in states not accessible from any other state by a.

Definition 1 (Refined model) Given a R_b-reflexive model $M^r = (W, R_a, {R_b}^r, V)$, the *refined model* (of M^r) is the model $M^{r|a} = (W, R_a, {R_b}^{r|a}, V)$ where

$${R_b}^{r|a} = {R_b}^r \setminus \{(w, u) \mid w{R_b}^r u \text{ and there is no } v \in W, \text{such that } vR_a w\}.$$

Proposition 38 (Refined model) *Let $M^{r|a} = (W, R_a, {R_b}^{r|a}, V)$ be the refined model of a R_b-reflexive model $M^r = (W, R_a, {R_b}^r, V)$. Then for any $\varphi \in \mathcal{L}_S$: $M^r, w \models \varphi$ iff $M^{r|a}, w \models \varphi$.*

Proof Induction on φ. Propositional cases are immediate. $\varphi = \langle S\rangle\psi$: $M^r, w \models \langle S\rangle\psi$; iff there exists $u \in W$ such that $wR_a u$ and if $M^r, u \models \neg\psi$ then for all $v \in W$, $uR_b v$ implies $M^r, v \models \neg\psi$; iff (by I.H.) there exists $u \in W$ such that $wR_a u$ and if $M^{r|a}, u \models \neg\psi$ then for all $v \in W$, $uR_b v$ implies $M^{r|a}, v \models \neg\psi$ (note that $uR_b v$ will not be deleted as $wR_a u$); iff $M^{r|a}, w \models \varphi$. □

Observe that refined models satisfy the following conditions:

- for any $w, u \in W$: $wR_a u$ implies $uR_b u$.
- for any $w, u \in W$: $wR_b u$ implies that there exists $v \in W$ such that $vR_a w$.

We now introduce the alternative semantics: interpretation in what we call *standard models*.

Definition 2 (Standard model) A standard model $M^o = (W^o, O, V^o)$ is defined as follows:

- W^o is a set of states.
- O is a ternary relation on W^o such that:

(i) for any $w, u \in W^o$: $O(w, u, u)$ iff there exists $v \in W^o$ such that $O(w, u, v)$;
(ii) for any $w, w', u, v, v' \in W^o$: $O(w, u, v)$ and $O(w', u, v')$ implies $O(w, u, v')$.

– V^o is a valuation function from `Prop` to the powerset of W^o.

A standard model "packs" the a and b relations into one ternary relation: intuitively $O(w, u, v)$ means that $wR_a u$ and $uR_b v$ (similar models already exist in the modal logic literature, see Section 6 for a discussion). Condition (ii) is thus a straightforward property of the composition of R_a and R_b. Condition (i) is b-reflexivity, which we without loss of generality (since it cannot be detected) can assume of our Kripke models.

The language is interpreted in standard models as follows. We use $\langle S\rangle$ here for simplicity, and the other cases are as in Kripke models. With the meaning mentioned above in mind, it is easy to see that this definition is equivalent to the corresponding definition for Kripke models.

$$\begin{array}{lll} M^o, w \models \langle S\rangle\varphi & \text{iff} & \exists u \in W^o \text{ such that } O(w,u,u) \text{ and } [M^o, u \models \varphi, \\ & & \text{or } \forall v \in W^o\text{: } O(w,u,v) \text{ implies } M^o, v \models \neg\varphi]. \end{array}$$

We can now define formal translations between models and standard models, making use of the notion of refined models introduced above. In particular, standard models are semantically equivalent to (binary) Kripke models as we will soon see it in Corollary 3.

Definition 3 (Translation) Given a standard moded $M^o = (W, O, V)$, the translated model $Tr(M^o) = (W, R_a, R_b, V)$ where:

– $wR_a u$ iff there exists $v \in W$ such that $O(w, u, v)$.
– $wR_b u$ iff there exists $v \in W$ such that $O(v, w, u)$.

We can now show that the translated model is equivalent to the standard model.

Theorem 2 *Let M^o be a standard model, and $Tr(M^o)$ be the translated model. We have for any $\varphi \in \mathcal{L}_S$:*

$$M^o, w \models \varphi \textit{ iff } Tr(M^o), w \models \varphi.$$

Proof Induction on φ. Boolean cases are immediate. We show the case $\varphi := \langle S\rangle\psi$. For simplicity, we denote $Tr(M^o)$ as M. Before we move to the diamond case, we prove the following two properties:

(*) for any $w, u \in W$: $wR_a u$ implies $uR_b u$.
(**) for any $w, u, v \in W$: from $wR_a u$ and $uR_b v$, we have $O(w, u, v)$.

For (*): Let $wR_a u$, then there exists $v \in W$ such that $O(w, u, v)$, and then by Definition 2(i), $O(w, u, u)$, and by Definition 3, $uR_b u$.
For (**): Let $wR_a u$ and $uR_b v$, by Definition 3 there exists $v', w' \in W$ such that $O(w, u, v')$ and $O(w', u, v)$. Then by Definition 2(ii), we have $O(w, u, v)$.

Now we move to the proof for $\varphi = \langle S\rangle\psi$.
(Left-to-right). Let $M^o, w \models \langle S\rangle\psi$, then by semantics, there exists $u \in W^o$ such that $O(w, u, u)$ and either $M^o, u \models \psi$, or $\forall v \in W^o$: if $O(w, u, v)$ then

$M^o, v \models \neg\psi$. Then there exists $u \in W$ such that $wR_a u$ (by Definition 3) and $uR_b u$ by (*) and either $M^o, u \models \psi$, or $\forall v \in W^o$: if $wR_a u$ and $uR_b v$ by (**), then $M^o, v \models \neg\psi$. By I.H., there exists $u \in W$ such that $wR_a u$ and $uR_b u$ and either $M, u \models \psi$, or $\forall v \in W$, if $wR_a u$ and $uR_b v$, then $M, v \models \neg\psi$. Then there exists $u \in W$ such that $wR_a u$ and either $M, u \models \psi$, or $\forall v \in W$, if $uR_b v$, then $M, v \models \neg\psi$. By semantics, $M, w \models \langle S \rangle \psi$.
(Right-to-left). Let $M, w \models \langle S \rangle \psi$, there exists $u \in W$ such that $wR_a u$ and either $M, u \models \psi$, or $\forall v \in W$, if $uR_b v$ then $M, v \models \neg\psi$. By (*), there exists $u \in W$ such that $wR_a u$ and $uR_b u$ and either $M, u \models \psi$, or $\forall v \in W$, if $wR_a u$ and $uR_b v$ then $M, v \models \neg\psi$. By Definition 3, Definition 2(ii), there exists $u \in W$ such that $O(w, u, u)$ and either $M, u \models \psi$, or $\forall v \in W$, if $O(w, u, v)$ then $M, v \models \neg\psi$. By I.H., there exists $u \in W$ such that $O(w, u, u)$ and either $M^o, u \models \psi$, or $\forall v \in W^o$, if $O(w, u, v)$ then $M^o, v \models \neg\psi$. $\square$

The last missing piece of the puzzle is to show that any (refined) model is the result of translating some standard model.

Theorem 3 *For any refined model M, there exists a standard model M^o such that $Tr(M^o) = M$.*

Proof Let $M = (W, R_a, R_b, V)$ be a refined model. $M^o = (W^o, O, V^o)$ is defined as follows: $W^o = W$, $V^o = V$ and (*) for any $w, u, v \in W^o$: $O(w, u, v)$ iff $wR_a u$ and $uR_b v$.

First we show that M^o has the standard model properties.

- For $O(w, u, u)$ iff $O(w, u, v)$ for some v. Left-to-right: trivial. Right-to-left: let $O(w, u, v)$, then $wR_a u$ and $uR_b v$ by (*), and thus $wR_a u$ and $uR_b u$ by the observations about refined model above, and then $O(w, u, u)$ by (*).
- For $O(w, u, v)$ and $O(w', u, v')$ implies $O(w, u, v')$. Assume that $O(w, u, v)$ and $O(w', u, v')$, by (*), we get $wR_a u$, $uR_b v$, $w'R_a u$, and $uR_b v'$, and then we get that $O(w, u, v')$ from $wR_a u$ and $uR_b v'$ by (*).

Now we show that $Tr(M^o) = M$. Let $Tr(M^o) = (W', {R_a}', {R_b}', V')$. Clearly $W' = W$ and $V' = V$. For ${R_a}'$ and ${R_b}'$:

- For any w, u: $w{R_a}'u$ iff $wR_a u$. If $w{R_a}'u$, by Definition 3 there exists v such that $O(w, u, v)$, and then by (*) we get that $wR_a u$. If $wR_a u$, by the refined model observations above, $uR_b u$, and then $O(w, u, u)$ by (*) and $w{R_a}'u$ by Definition 3.
- For any w, u: $w{R_b}'u$ iff $wR_b u$. If $w{R_b}'u$, by Definition 3 there exists v such that $O(v, w, u)$, then by (*), we infer $wR_b u$. If $wR_b u$, by the observations there exists v such that $vR_a w$, and then $O(v, w, u)$ by (*) and $w{R_b}'u$ by Definition 3.

Thus, M^o is indeed a standard model and $Tr(M^o) = M$. $\square$

Consequently, together with Proposition 38 and Proposition 37, any model is equivalent to some standard model.

Theorem 4 *For any model M, there is a standard model M^o such that for any $\varphi \in \mathcal{L}_S$: $M, w \models \varphi$ iff $M^o, w \models \varphi$.*

Proof Let M be a model, then by Proposition 37, we have $M, w \models \varphi$ iff $M^r, w \models \varphi$, and then by Proposition 38, we have $M^r, w \models \varphi$ iff $M^{r|a}, w \models \varphi$. We have some standard model M^o such that $M^{r|a} = Tr(M^o)$ by Theorem 3, which gives us $M^o, w \models \varphi$ iff $M^{r|a}, w \models \varphi$ by Theorem 2. □

Theorems 2 and 4 give us the following corollary.

Corollary 3 *Any $\varphi \in \mathcal{L}_S$ is valid on models iff it is valid on standard models.*

5.4 Completeness of ICS

We now prove completeness of **ICS** with respect to standard models. The translation in the previous section then immediately gives us completeness also with respect to Kripke models. The proof is based on the canonical model method, making use of the standard definition of maximal consistent sets of formulae. We are faced with two main challenges in applying this method, however. First, the model we will build is what we have called a standard model, instead of a Kripke model, making it necessary to come up with a (ternary) relation for standard models. Second, for reasons that will become clear soon, the conventional way of defining the state space as all maximal consistent sets does not give us enough states; we will need several "copies" of each maximal consistent set. Before we can define the canonical standard model we need a couple of intermediate definitions and observations, having to do with these challenges.

Definition 4 Let Δ be a maximal consistent set of formulae (mcs). We define the following abbreviations.

- $S(\Delta) := \{\varphi \mid S\varphi \in \Delta\}$,
- $N(\Delta) := \{\varphi \mid \langle S \rangle\varphi \in \Delta\}$,
- $E(\Delta) := \{\neg\chi \mid S(\Delta) \vdash \neg\chi$ and $\chi \in N(\Delta)\}$.

Intuitively, $E(\Delta)$ is the set of formulas $\neg\chi$ that (1) can be derived from secrets in Δ ($\{\varphi : S\varphi \in \Delta\} \vdash \neg\chi$) but (2) are not secrets in Δ ($\neg S\neg\chi \in \Delta$) (note that it is possible for (1) and (2) to hold at the same time due to the lack of monotonicity). In the construction of the canonical model that follows, we are going to have to make sure that for each of these formulas there is an accessible state for agent a that is a proper witness for $\langle S \rangle\chi$ being true. The following technical property of $S(\Delta)$ is needed for the truth lemma.

Proposition 39 *Let Δ be an mcs. If $S(\Delta)$ is consistent then for any $\delta \in S(\Delta)$ and $\neg\chi \in E(\Delta)$, the set $\{\neg\delta, \neg\chi\}$ is consistent.*

Proof Assume that $S(\Delta)$ is consistent, that $\delta \in S(\Delta), \neg\chi \in E(\Delta)$, and that $\{\neg\delta, \neg\chi\}$ is not consistent. Then $\vdash \neg\chi \to \delta$. Also, from the definition of $E(\Delta)$, $S(\Delta) \vdash \neg\chi$, there exists a sequence $\delta_1, \ldots, \delta_k \in S(\Delta)$ such that $\vdash \bigwedge\limits_k \delta_k \to \neg\chi$. From Proposition 36, we have $\vdash (\bigwedge\limits_k S\delta_k \wedge \langle S \rangle\chi) \to \neg S\delta$. As we have $\bigwedge\limits_k S\delta_k \in \Delta$

and $\langle S\rangle\chi \in \Delta$, then $\neg S\delta \in \Delta$ as well. Therefore, $S\delta \notin \Delta$ and thus $\delta \notin S(\Delta)$, contradicting the assumptions. $\Box$

Let $L(\varphi)$ be the set of mcss containing formula φ.

The following is a technical definition of two sets $Ra(\Delta)$ and $Rab(\Delta)$ that will be used to build the the standard canonical model, more precisely to define the states accessible by agent a, and by agent a followed by agent b, respectively.

Definition 5 ($Ra(\Delta), Rab(\Delta)$) Let Δ be an mcs. $Ra(\Delta)$ is defined as follows: if $S(\Delta)$ is not consistent, we let $Ra(\Delta) = \emptyset$, otherwise, $Ra(\Delta) = \{\Delta' \mid \Delta'$ is an mcs such that $S(\Delta) \subseteq \Delta'\}$. $Rab(\Delta)$ is defined as follows, only for the case that $S(\Delta)$ is consistent (undefined otherwise): if $E(\Delta) \neq \emptyset$ then $Rab(\Delta) = \{L(\neg\chi) \mid \neg\chi \in E(\Delta)\}$; if $E(\Delta) = \emptyset$ then $Rab(\Delta) = \{L(\neg\bot)\}$.

The following intermediate result will be used in the existence lemma.

Proposition 40 *Let Δ be an mcs. $S(\Delta)$ is consistent iff there exists $\varphi \in \mathcal{L}_S$ such that $\varphi \notin S(\Delta)$.*

Proof Let Δ be an mcs. The left-to-right direction is trivial. For the other direction, let $\varphi \notin S(\Delta)$ for some φ. Assume that $S(\Delta)$ is not consistent. Then for some $\delta_1, \ldots, \delta_n \in S(\Delta)$, $\vdash (\delta_1 \wedge \ldots \wedge \delta_n) \to \bot$. It follows that $\vdash (\delta_1 \wedge \ldots \wedge \delta_n) \leftrightarrow \bot$, and $\vdash S(\delta_1 \wedge \ldots \wedge \delta_n) \leftrightarrow S\bot$ by (Re). Similar, we have $\vdash S(\delta_1 \wedge \neg\delta_1) \leftrightarrow S\bot$ from $\vdash \bot \leftrightarrow (\delta_1 \wedge \neg\delta_1)$, it gives $\vdash S(\delta_1 \wedge \ldots \wedge \delta_n) \leftrightarrow S(\delta_1 \wedge \neg\delta_1)$. By applying Prop. 35, and we get $\vdash S(\delta_1 \wedge \ldots \wedge \delta_n) \to S\delta_1$. Repeat this process for each δ_i ($1 \leq i \leq n$), and we get $\vdash S(\delta_1 \wedge \ldots \wedge \delta_n) \to (S\delta_1 \wedge \ldots \wedge S\delta_n)$. By axiom (C) we get $\vdash S(\delta_1 \wedge \ldots \wedge \delta_n) \leftrightarrow (S\delta_1 \wedge \ldots \wedge S\delta_n)$, i.e., $\vdash S\bot \leftrightarrow (S\delta_1 \wedge \ldots \wedge S\delta_n)$ and $S\bot \in \Delta$, thus by Prop. 35 and definition of $\bot$ $\vdash S\bot \to S\varphi$ for any $\varphi \in \mathcal{L}_S$, and thus we have $S\varphi \in \Delta$ and $\varphi \in S(\Delta)$ for any φ, a contradiction to assumption. $\Box$

We can finally give the definition of the canonical standard model. One difference to conventional canonical model constructions, in addition to the fact that we employ standard models, is that the states will be what we called *labelled maximal consistent sets (lmcs)* instead of maximal consistent sets (mcs). An lmcs is of the form $\Gamma[\varphi]$, where Γ is an mcs and $\varphi \in \mathcal{L}_S$ is a formula. We write $\varphi \in \Delta[\chi]$ iff $\varphi \in \Delta$. Note that we can construct many different lmcs from *one* mcs by chosing different labels. The reason for the need to get more states by using labels is to get a sufficient number of "middle states" between a-b accessible states; we explain this in more detail after the definition.

Definition 6 (Canonical standard model) The canonical standard model $M^c = (W^c, O^c, V^c)$ is defined as follows.

- $W^c = \{\Delta[\varphi] \mid \Delta[\varphi]$ is a labelled mcs$\}$;
- If $S(\Delta)$ is consistent: for any $\neg\chi, \chi_1, \chi_2 \in \mathcal{L}_S$,
  - $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma'[\chi_2])$ iff $\Gamma \in Ra(\Delta)$, $L(\neg\chi) \in Rab(\Delta)$, and $\Gamma' \in L(\neg\chi)$.

– If $S(\Delta)$ is not consistent:
 – there is no mcs Γ, Γ', and $\neg\chi, \chi_1, \chi_2 \in \mathcal{L}_S$ s.t. $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma'[\chi_2])$.
– $V^c(p) = \{\Delta[\varphi] \mid p \in \Delta\}$ for any $p \in \texttt{Prop}$ and $\varphi \in \mathcal{L}_S$.

Note that the labels χ_1 and χ_2 play no role in the definition of O^c.

To understand the use of labels, consider the case that only used standard mcss. If $\langle S\rangle\chi \in \Delta$, it might be the case that we need all a-accessible states to contain $\neg\chi$. In that case, for any Γ a-accessible from Δ we need that for any b-accessible set from Γ, i.e., Γ', $\neg\chi \in \Gamma'$. For the same reason, we could also have $\langle S\rangle\delta \in \Delta$ such that $\neg\delta \in \Gamma$ and then we need for any b-accessible set from Γ, i.e., Γ', that $\neg\delta \in \Gamma'$. But that might not be possible: $\neg\chi$ and $\neg\delta$ might not even be consistent. That is why we need more "copies" of Γ labelled by, e.g., $\neg\chi$ or $\neg\delta$ – more "middle states".

Note that labels are only relevant for the lmcs in the middle place of the O^c relation: the b-accessible states are identified by the label. For instance, for $\Gamma[\neg\chi]$, the b-accessible lmcss are all mcss defined by $Rab(\Delta)$ that contains $\neg\chi$. And for $\Gamma[\neg\delta]$ (same mcs Γ), the b-accessible lmcss are all mcss defined by $Rab(\Delta)$ that contains $\neg\delta$. Also note that if $\neg\chi$ is inconsistent, then χ is a tautology, and we have $\chi \in \Gamma$ (it's b-accessible set is empty), and $\langle S\rangle\chi \in \Delta$ if Γ is a-accessible from Δ.

We now show that the standard canonical model indeed is a standard model.

Proposition 41 (Canonicity) *Let $\Delta[\chi_1]$ be any lmcs, Γ be any mcs, and $\neg\chi \in \mathcal{L}_S$ be any formula. we have the following properties:*

(1) $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma'[\chi_2])$ for some lmcs $\Gamma'[\chi_2]$ iff $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma[\neg\chi])$.
(2) For any lmcs $\Delta'[\chi_2]$, $\Omega[\chi_3]$, and $\Omega'[\chi_4]$:
from $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Omega[\chi_3])$ and $O^c(\Delta'[\chi_2], \Gamma[\neg\chi], \Omega'[\chi_4])$,
we have $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Omega'[\chi_4])$.

Proof Firstly, we show that (*) for any lmcs $\Delta[\chi_1], \Gamma[\neg\chi]$ and $\Gamma[\chi_2]$:

$$O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma[\chi_2]) \text{ iff } \Gamma \in Ra(\Delta) \text{ and } L(\neg\chi) \in Rab(\Delta).$$

Case 1: If $S(\Delta)$ is not consistent, then by the definition of $Ra(\Delta)$ (Def. 5), $Ra(\Delta) = \emptyset$. We have $\Gamma \notin Ra(\Delta)$ and not $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma[\chi_2])$ by Definition 6.

Case 2: If $S(\Delta)$ is consistent, we then consider by two directions.

(Left-to-right). Trivial, by Definition 6.

(Right-to-left). As we have $\Gamma \in Ra(\Delta)$ and $L(\neg\chi) \in Rab(\Delta)$, we only need to show that $\Gamma \in L(\neg\chi)$, by Definition 6. Based on the fact that $S(\Delta)$ is consistent, we consider the following two cases:

– If $E(\Delta) = \emptyset$, then $Rab(\Delta) = \{L(\neg\bot)\}$, and $\chi = \bot$ such that $L(\neg\bot)$ is the only element of $Rab(\Delta)$ by Definition 5. We have $\Gamma \in L(\neg\bot)$ by the definition of L, i.e., $\Gamma \in L(\neg\chi)$ from $\chi = \bot$.

– If $E(\Delta) \neq \emptyset$, then for any $L(\neg\chi) \in \mathit{Rab}(\Delta)$ we have $S(\Delta) \vdash \neg\chi$. As $\Gamma \in \mathit{Ra}(\Delta)$, we have $S(\Delta) \subseteq \Gamma$ by Definition 5, and by $S(\Delta) \vdash \neg\chi$, we have $\Gamma \vdash \neg\chi$. Then $\neg\chi \in \Gamma$ since Γ is an mcs, and it follows $\Gamma \in L(\neg\chi)$ by the definition of L.

Now we can move to the canonicity proof.
(1). (Right-to-left). Trivial. (Left-to-right). Let $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma'[\chi_2])$ for some lmcs $\Gamma'[\chi_2]$. By definition, $\Gamma \in \mathit{Ra}(\Delta)$ and $L(\neg\chi) \in \mathit{Rab}(\Delta)$, and by (*), $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma[\neg\chi])$.
(2). Let (a) $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Omega[\chi_3])$ and (b) $O^c(\Delta'[\chi_2], \Gamma[\neg\chi], \Omega'[\chi_4])$. Apply Definition 6. From (a), we have (i) $\Gamma \in \mathit{Ra}(\Delta)$ and (ii) $L(\neg\chi) \in \mathit{Rab}(\Delta)$; from (b), we have (iii) $\Omega' \in L(\neg\chi)$. By Definition 6, by (i), (ii), and (iii), we get $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Omega'[\chi_4])$. □

We have showed that the canonical standard model is a standard model. We proceeed by providing an existence lemma, and then the truth lemma.

Lemma 1 (Existence lemma) *If $\Delta[\chi_1]$ is an lmcs and $\langle S\rangle\varphi \in \Delta$, then there exists an lmcs $\Gamma[\neg\chi]$ such that $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma[\neg\chi])$ and*

– *either $\varphi \in \Gamma$,*
– *or for all $\Gamma'[\chi_2]$, $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma'[\chi_2])$ implies $\neg\varphi \in \Gamma'$.*

Proof Let $\Delta[\chi_1]$ be an lmcs and $\langle S\rangle\varphi \in \Delta$. Then $N(\Delta)$ is not empty, and $\neg\varphi \notin S(\Delta)$. Then $S(\Delta)$ is consistent from Proposition 40. Proof by cases:

– If $E(\Delta) = \emptyset$, then $\mathit{Rab}(\Delta) = \{L(\neg\bot)\}$ by definition, and we claim that $S(\Delta) \cup \{\varphi\}$ is consistent. Suppose not, we have $S(\Delta) \vdash \neg\varphi$, then $\neg\varphi \in E(\Delta)$ from $\varphi \in N(\Delta)$, a contradiction to the assumption that $E(\Delta) = \emptyset$. Then there is an mcs Γ containing $S(\Delta) \cup \{\varphi\}$ s.t. $O^c(\Delta[\chi_1], \Gamma[\neg\bot], \Gamma[\neg\bot])$ and $\varphi \in \Gamma$.
– If $E(\Delta) \neq \emptyset$, then $\mathit{Rab}(\Delta) = \{L(\neg\chi) \mid \neg\chi \in E(\Delta)\}$ by Definition 5. Based on $\varphi \in N(\Delta)$, we consider two cases:
  – If $\neg\varphi \in E(\Delta)$, then $L(\neg\varphi) \in \mathit{Rab}(\Delta)$ by Definition 5. We just let Γ be an mcs such that $S(\Delta) \subseteq \Gamma$ (clearly, $\Gamma \in \mathit{Ra}(\Delta)$ by Definition 5), by item (*) in the proof of Proposition 41 we have that $\Gamma[\neg\varphi]$ such that $O^c(\Delta[\chi_1], \Gamma[\neg\varphi], \Gamma[\neg\varphi])$, and for any $\Gamma'[\chi_2]$, $O^c(\Delta[\chi_1], \Gamma[\neg\varphi], \Gamma'[\chi_2])$ implies $\neg\varphi \in \Gamma'$, since $\Gamma' \in L(\neg\varphi)$ from $O^c(\Delta[\chi_1], \Gamma[\neg\varphi], \Gamma'[\chi_2])$ by Definition 6.
  – if $\neg\varphi \notin E(\Delta)$, from $\varphi \in N(\Delta)$ we have $S(\Delta) \nvdash \neg\varphi$, therefore $S(\Delta) \cup \{\varphi\}$ is consistent. Extend it into an mcs Γ (then $\Gamma \in \mathit{Ra}(\Delta)$ by Definition 5); we have for any $L(\neg\chi) \in \mathit{Rab}(\Delta)$, $O^c(\Delta, \Gamma[\neg\chi], \Gamma[\neg\chi])$ from item (*) in the proof of Proposition 41 and $\varphi \in \Gamma$ by construction.

□

In order to prove the truth lemma, and thus the desired result, we need the following proposition.

Proposition 42 *If there exists $\Gamma[\neg\chi] \in W^c$ such that $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma[\neg\chi])$ and [either $\psi \in \Gamma$, or for all $\Gamma'[\chi_2] \in W^c$: $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma'[\chi_2])$ implies $\neg\psi \in \Gamma'$], then we have $\langle S\rangle\psi \in \Delta$.*

Proof Proof by contraposition. Let $\langle S\rangle\psi \notin \Delta$ – then $S\neg\psi \in \Delta$ by the mcs property and duality. We consider two cases:

– if $S(\Delta)$ is not consistent, then $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma[\neg\chi])$ does not hold for any $\Gamma[\neg\chi] \in W^c$ by Definition 6, and we are done.
– if $S(\Delta)$ is consistent, we need to show that "for any $\Gamma[\neg\chi] \in W^c$, if $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma[\neg\chi])$ then $\neg\psi \in \Gamma$ and there exists $\Gamma'[\chi_2] \in W^c$: $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma'[\chi_2])$ and $\psi \in \Gamma'$". We consider two cases:
  – $E(\Delta) = \emptyset$: Assume that $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma[\neg\chi])$. Then $\Gamma \in Ra(\Delta)$ and $L(\neg\chi) \in Rab(\Delta)$ by definition of O^c (Def. 6). We immediately have $\neg\psi \in \Gamma$ from $\Gamma \in Ra(\Delta)$ and $S\neg\psi \in \Delta$. Since $E(\Delta) = \emptyset$, we have $Rab(\Delta) = \{L(\neg\bot)\}$ by definition of Rab (Def. 5), it follows that $L(\neg\chi) = L(\neg\bot)$ since $L(\neg\bot)$ is the only element of $Rab(\Delta)$. Now we claim that $\{\psi\}$ must be consistent. Suppose not, then $\vdash \psi \to \bot$, and $\vdash \neg\psi \leftrightarrow \neg\bot$ by propositional logic. Then by the (Re) rule, $\vdash S\neg\psi \leftrightarrow S\neg\bot$. From $S\neg\psi \in \Delta$, $S\neg\bot \in \Delta$ and $S\top \in \Delta$, then by the (S) axiom, $S\bot \in \Delta$, and it follows that $\bot \in S(\Delta)$ by Def. 4. That shows that $S(\Delta)$ is not consistent, a contradiction. Therefore, we have proved that $\{\psi\}$ is consistent. Then there is an mcs Γ' such that $\psi \in \Gamma'$ and clearly, $\Gamma' \in L(\neg\bot)$ by definition of L, and thus $\Gamma' \in L(\neg\chi)$ since $\chi = \bot$.
  – if $E(\Delta) \neq \emptyset$. Assume that $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma[\neg\chi])$, then $\Gamma \in Ra(\Delta)$, $L(\neg\chi) \in Rab(\Delta)$ by definition of O^c (Def. 6). We immediately have $\neg\psi \in \Gamma$ from $\Gamma \in Ra(\Delta)$ and $S\neg\psi \in \Delta$. Since $E(\Delta) \neq \emptyset$, we have $Rab(\Delta) = \{L(\neg\chi) \mid \neg\chi \in E(\Delta)\}$ by Def. 5. From Proposition 39, we have that $\{\neg\neg\psi, \neg\chi\}$ is consistent as $\neg\psi \in S(\Delta)$. Then for lmcs $\Gamma'[\chi_2]$ containing $\{\neg\neg\psi, \neg\chi\}$, we have $\Gamma' \in L(\neg\chi)$ and $\psi \in \Gamma'$.

□

Lemma 2 (Truth lemma) *Let $M^c = (W^c, O^c, V^c)$ be the canonical standard model. We have that $M^c, \Delta[\chi_1] \models \varphi$ iff $\varphi \in \Delta[\chi_1]$ for any formula $\varphi, \chi_1 \in \mathcal{L}_S$.*

Proof Let χ_1 be arbitrary. Propositional cases are immediate. Consider the $\varphi = \langle S\rangle\psi$ case.

(Left-to-right) If $M^c, \Delta[\chi_1] \models \langle S\rangle\psi$, then there exists an lmcs $\Gamma[\neg\chi] \in W^c$ such that $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma[\neg\chi])$ and if $M^c, \Gamma[\neg\chi] \not\models \neg\psi$, then for all $\Gamma'[\chi_2] \in W$: $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma'[\chi_2])$ implies $M, \Gamma'[\chi_2] \models \neg\psi$. By IH, there exists an lmcs $\Gamma[\neg\chi] \in W^c$ such that $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma[\neg\chi])$ and if $\psi \notin \Gamma$, then for all $\Gamma'[\chi_2] \in W$: $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma'[\chi_2])$ implies $\neg\psi \in \Gamma'$. We have $\langle S\rangle\psi \in \Delta$ from Proposition 42, and $\langle S\rangle\psi \in \Delta[\chi_1]$ follows by definition.

(Right-to-left) If $\langle S\rangle\psi \in \Delta[\chi_1]$ (that's $\langle S\rangle\psi \in \Delta$ by definition), then by Lemma 1, there exists an lmcs $\Gamma[\neg\chi]$ such that $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma[\neg\chi])$ and [either $\psi \in \Gamma$, or for all $\Gamma'[\chi_2]$, $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma'[\neg\chi_2])$ implies $\neg\psi \in \Gamma'$]. It follows that there exists an lmcs $\Gamma[\neg\chi]$ such that $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma[\neg\chi])$ and [either $M^c, \Gamma[\neg\chi] \models \psi$ (by IH), or for all Γ', $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma'[\chi_2])$ implies $M^c, \Gamma'[\chi_2] \models \neg\psi$ (by IH)]. Then by semantics, $M^c, \Delta[\chi_1] \models \langle S\rangle\psi$. □

Theorem 5 ICS *is sound and strongly complete with respect to the class of all standard models.*

Proof For soundness, assume that φ is not valid on all standard models, i.e., that $M^o, w \not\models \varphi$ for some M^o, w. By Theorem 2, $Tr(M^c)$ is the model such that $Tr(M^c), w \models \neg\varphi$. That means that φ is not valid on the class of all models, which by Theorem 1 means that $\not\vdash \varphi$. For completeness: proof by contraposition. Assume that $\Omega \not\vdash \varphi$, then $\Omega \cup \{\neg\varphi\}$ is consistent. Extend $\Omega \cup \{\neg\varphi\}$ into an lmcs $\Delta[\chi_1]$. We have $M^c, \Delta[\chi_1] \models \neg\varphi$ by Lemma 2, and it follows that $\Omega \not\models \varphi$ in standard models. $\square$

Theorem 6 (Completeness) ICS *is sound and strongly complete with respect to the class of all models.*

Proof Let $\Omega \not\vdash \varphi$. We let $\Omega \subseteq \Delta$ and $M^c, \Delta[\chi_1] \models \neg\varphi$ as in the proof of Theorem 5. By Theorem 2, for all φ $M^c, \Delta[\chi_1] \models \varphi$ iff $Tr(M^c), \Delta[\chi_1] \models \varphi$, and it follows that $\Omega \not\models \varphi$ on the class of models. $\square$

We mention two minor results that immediately follow: **ICS** is also sound and strongly complete with respect to (1) all R_b-reflexive models (Prop. 37, Theorem 2) as well as (2) all refined models (Prop. 38, Theorem 2).

5.5 ICST: Soundness and Completeness

Let **ICST** be **ICS** extended with the (T) axiom $S\varphi \rightarrow \varphi$. We will show that **ICST** is sound and complete with respect to *reflexive* models (models where both agents have reflexive accessibility relations).

Theorem 7 ***ICST*** *is sound with repect to reflexive models.*

Proof Follows immediately from soundness of **ICS**, and validity of (T) which follows immediately from reflexivity for agent a.

In the remainder of this section we adapt the completeness proof of **ICS** to **ICST**. In the rest of the section we use $\vdash$ to stand for $\vdash_{\mathbf{ICST}}$, by "consistent" we mean **ICST** consistent; by *mcs* we mean maximal **ICST**-consistent set, and similarly for *lmcs*s and so on. Definitions of L, Ra and Rab are as in the previous section. We define the canonical standard model for **ICST** as follows.

Definition 7 (Canonical standard model) The canonical standard model $M^c = (W^c, O^c, V^c)$ for **ICST** is defined as follows.

- $W^c = \{\Delta[\varphi] \mid \Delta[\varphi]$ is an lmcs such that $L(\varphi) \in Rab(\Delta)\}$;
- For any $\neg\chi \in \mathcal{L}_S$ such that $\Gamma[\neg\chi] \in W^c$ and for any $\Delta, \Gamma' \in W^c$,
  - $O^c(\Delta, \Gamma[\neg\chi], \Gamma')$ iff $\Gamma \in Ra(\Delta)$, $L(\neg\chi) \in Rab(\Delta)$ and $\Gamma' \in L(\neg\chi)$.
- $V^c(p) = \{\Delta \mid p \in \Delta\}$ for any $p \in \mathtt{Prop}$.

The following properties of the canonical standard model will be useful.

Corollary 4 *For any lmcs* $\Delta[\chi_1] \in W^c$, $S(\Delta)$ *is consistent and* $E(\Delta) \neq \emptyset$. *Moreover,* $\neg\bot \in E(\Delta)$, $Ra(\Delta) \neq \emptyset$, *and* $Rab(\Delta) = \{L(\neg\chi) \mid \neg\chi \in E(\Delta)\}$.

Proof Let $\Delta[\chi_1] \in W^c$ be an arbitrary lmcs. We have that $S(\Delta)$ is consistent by Proposition 40, since $\neg S\bot \in \Delta$. It follows that $Ra(\Delta) \neq \emptyset$ by Definition 5. Similarly, $\neg\bot \in E(\Delta)$ by Definition 4 from $\langle S\rangle\bot \in \Delta$ and $S(\Delta) \vdash \neg\bot$. It follows that $E(\Delta) \neq \emptyset$. Then by Definition 5, we have $Rab(\Delta) = \{L(\neg\chi) \mid \neg\chi \in E(\Delta)\}$. □

M^c is indeed a standard model. The following can be shown with the same proof strategy as in Proposition 41 (although the details differ a little).

Proposition 43 (Canonicity) *Let $\Delta[\chi_1], \Gamma[\neg\chi] \in W^c$ be any lmcs. We have the following properties:*

(1) $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma'[\chi_2])$ for some lmcs $\Gamma'[\chi_2] \in W^c$ iff $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma[\neg\chi])$.
(2) For any $\neg\chi \in \mathcal{L}_S$: from $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Omega[\chi_2])$ and $O^c(\Delta'[\chi_3], \Gamma[\neg\chi], \Omega'[\chi_4])$, we have $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Omega'[\chi_4])$.

We say that a standard model is *reflexive*, if $O(x, x, x)$ holds for any $x \in W$. The canonical standard model is reflexive:

Proposition 44 (Reflexivity) *The canonical standard model is reflexive.*

Proof Let $\Delta[\chi] \in W^c$. We will show that $O^c(\Delta[\chi], \Delta[\chi], \Delta[\chi])$. From $\Delta[\chi] \in W^c$, we have that $L(\chi) \in Rab(\Delta)$ by the definition of W^c (Definition 7). First, we prove that $\Delta \in Ra(\Delta)$. From $S(\Delta)$ being consistent and the (T) axiom, we have (*) $S(\Delta) \subseteq \Delta$ (if $\varphi \in S(\Delta)$, then $S\varphi \in \Delta$ by definition (Def. 4), and $\varphi \in \Delta$ by (T)), it then follows that $\Delta \in Ra(\Delta)$ by Definition 5.

Now, we show that $\Delta \in L(\chi)$. From $L(\chi) \in Rab(\Delta)$, we have (a) $S(\Delta) \vdash \chi$ and $\chi \in E(\Delta)$ from Definition 5, since $E(\Delta) \neq \emptyset$ (Cor. 4). From (a), we know that $\chi \in \Gamma$ for any $S(\Delta) \subseteq \Gamma$. Therefore, $\chi \in \Delta$ when $\Gamma = \Delta$ from (*) and $\Delta \in L(\chi)$ by definition of L. Therefore, $O^c(\Delta[\chi], \Delta[\chi], \Delta[\chi])$ follows by Definition 7. □

Now we need to prove the existence lemma again, since we have deleted lots of lmcss.

Lemma 3 (Existence lemma) *For any $\Delta[\chi_1] \in W^c$, if $\langle S\rangle\varphi \in \Delta$, then there exists $\Gamma[\neg\chi] \in W^c$ such that $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma[\neg\chi])$ and [either $\varphi \in \Gamma$, or for any lmcs $\Gamma'[\chi_2] \in W^c$, if $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma'[\chi_2])$, then $\neg\varphi \in \Gamma'$].*

Proof Let $\Delta[\chi_1]$ be an lmcs and $\langle S\rangle\varphi \in \Delta$. From $E(\Delta) \neq \emptyset$, we have $Rab(\Delta) = \{L(\neg\chi) \mid \neg\chi \in E(\Delta)\}$ by Def. 5. Based on $\varphi \in N(\Delta)$, we consider two cases:

- If $\neg\varphi \in E(\Delta)$, then $L(\neg\varphi) \in Rab(\Delta)$, and $\Delta[\neg\varphi] \in W^c$. By reflexivity (Prop. 44), $O^c(\Delta[\neg\varphi], \Delta[\neg\varphi], \Delta[\neg\varphi])$. It follows that $O^c(\Delta[\chi_1], \Delta[\neg\varphi], \Delta[\neg\varphi])$ by Definition 7 and $\Delta[\chi_1] \in W^c$. For any lmcs $\Gamma'[\chi_2]$, $O^c(\Delta[\chi_1], \Delta[\neg\varphi], \Gamma'[\chi_2])$ implies $\neg\varphi \in \Gamma'$ by Definition 7.
- If $\neg\varphi \notin E(\Delta)$ from $\varphi \in N(\Delta)$, we have $S(\Delta) \nvdash \neg\varphi$, and therefore $S(\Delta) \cup \{\varphi\}$ is consistent. Extend it into an mcs Γ. Then $\Gamma \in Ra(\Delta)$, and for the particular $L(\neg\bot) \in Rab(\Delta)$, we have that $O^c(\Delta[\chi_1], \Gamma[\neg\bot], \Gamma[\neg\bot])$ as $\Gamma \in L(\neg\bot)$, and $\varphi \in \Gamma$ by construction.

□

Like in the case of **ICS** we need the following result for the truth lemma.

Proposition 45 *If for some $\Gamma[\neg\chi] \in W^c$, $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma[\neg\chi])$ and [either $\psi \in \Gamma$, or for all $\Gamma'[\chi_2] \in W^c$, $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma'[\chi_2])$ implies $\neg\psi \in \Gamma'$], then $\langle S\rangle\psi \in \Delta$.*

Proof Proof by contraposition. Let $\langle S\rangle\psi \notin \Delta$, then $S\neg\psi \in \Delta$ by the mcs property and duality. We need to show that "for any $\Gamma[\neg\chi] \in W^c$, if $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma[\neg\chi])$ then [$\neg\psi \in \Gamma$ and for some $\Gamma' \in W^c$, $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma')$ and $\psi \in \Gamma'$]".

Assume that $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma[\neg\chi])$, then $\Gamma \in Ra(\Delta)$, $L(\neg\chi) \in Rab(\Delta)$ by definition. We immediately have $\neg\psi \in \Gamma$ from $\Gamma \in Ra(\Delta)$ and $S\neg\psi \in \Delta$. Since $E(\Delta) \neq \emptyset$, we have $Rab(\Delta) = \{L(\neg\chi) \mid \neg\chi \in E(\Delta)\}$ by definition. From Proposition 39, we have that $\{\neg\neg\psi, \neg\chi\}$ is consistent as $\neg\psi \in S(\Delta)$. Then for all mcs Γ' containing $\{\neg\neg\psi, \neg\chi\}$ such that $\Gamma'[\chi_2] \in W^c$ for any $\chi_2 \in \mathcal{L}_S$, we have $\Gamma'[\chi_2] \in L(\neg\chi)$ and $\psi \in \Gamma'$. □

Lemma 4 (Truth lemma) *Let $M^c = (W^c, O^c, V^c)$ be the canonical standard model for* **ICST**, *and $\Delta[\chi_1]$ be any lmcs in W^c. We have that for any formula $\varphi \in \mathcal{L}_S$, $M^c, \Delta[\chi_1] \models \varphi$ iff $\varphi \in \Delta[\chi_1]$.*

Proof The propositional cases are immediate. Consider the case $\varphi = \langle S\rangle\psi$.
(Left-to-right). If $M^c, \Delta[\chi_1] \models \langle S\rangle\psi$, then for some $\Gamma[\neg\chi] \in W^c$, $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma[\neg\chi])$ and [either $M^c, \Gamma[\neg\chi] \models \psi$, or for all $\Gamma'[\chi_2] \in W^c$: $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma'[\chi_2])$ implies $M, \Gamma'[\chi_2] \models \neg\psi$]. By IH, for some $\Gamma[\neg\chi] \in W^c$, $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma[\neg\chi])$ and [either $\psi \in \Gamma$, or for all $\Gamma'[\chi_2] \in W^c$: $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma'[\chi_2])$ implies $\neg\psi \in \Gamma'$]. We have that $\langle S\rangle\psi \in \Delta[\chi_1]$ from Proposition 45.
(Right-to-left). If $\langle S\rangle\psi \in \Delta[\chi_1]$, by Lemma 3, there exists $\Gamma[\neg\chi]$ such that $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma[\neg\chi])$ and [either $\psi \in \Gamma[\neg\chi]$, or for all $\Gamma'[\chi_2] \in W^c$, $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma'[\chi_2])$ implies $\neg\psi \in \Gamma'$]. It follows that there exists $\Gamma[\neg\chi]$ such that $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma[\neg\chi])$ and [either $M^c, \Gamma[\neg\chi] \models \psi$ (by IH), or for all $\Gamma'[\chi_2]$, $O^c(\Delta[\chi_1], \Gamma[\neg\chi], \Gamma'[\chi_2])$ implies $M^c, \Gamma'[\chi_2] \models \neg\psi$ (by IH)]. By semantics, $M^c, \Delta[\chi_1] \models \langle S\rangle\psi$. □

The following corollary follows immediately.

Corollary 5 *For any $\Delta[\neg\chi_1], \Delta[\neg\chi_2] \in W^c$ and for any $\varphi \in \mathcal{L}_S$,*

$$M^c, \Delta[\neg\chi_1] \models \varphi \textit{ iff } M^c, \Delta[\neg\chi_2] \models \varphi.$$

Proof Assume that $\Delta[\neg\chi_1], \Delta[\neg\chi_2] \in W^c$ are arbitrary, and $\varphi \in \mathcal{L}_S$ is an arbitrary formula. $M^c, \Delta[\neg\chi_1] \models \varphi$, iff $\varphi \in \Delta[\neg\chi_1]$ (by Lemma 4), iff $\varphi \in \Delta[\neg\chi_2]$, iff $M^c, \Delta[\neg\chi_2] \models \varphi$ (by Lemma 4). □

Theorem 8 (Completeness) ICST *is sound and strongly complete with respect to the class of reflexive standard models.*

Proof That (T) is valid on the class of reflxive standard models follows directly from the definition. Completeness: proof by contraposition. Let $\Gamma \nvdash_{\mathbf{ICST}} \varphi$. Then $\Gamma \cup \{\neg\varphi\}$ is consistent, and it can be extended into a maximal and **ICST**-consistent set Γ' such that $M^c, \Gamma'[\neg\bot] \models \neg\varphi$ by Lemma 4 and Corollary 5 ($\Gamma'[\neg\bot]$ is admissible from Corollary 4) and it then follows that $\Gamma \not\models \varphi$. □

Theorem 9 **ICST** *is sound and strongly complete with respect to the class of reflexive models.*

Proof From the proof of Theorem 8, we have $M^c, \Gamma'[\neg\bot] \models \neg\varphi$. By Theorem 2, $Tr(M^c), \Gamma'[\neg\bot] \models \neg\varphi$, and $Tr(M^c)$ is a (refined) model. It is easy to see that the translation of a reflexive standard model has R_a-reflexivity. By Proposition 37, we have that **ICST** is strongly complete with respect to the class of models satisfying both R_a and R_b reflexivity.

6 Discussion

In this paper we formalised the concept of "secretly knowing" as a modality on the abstraction level of standard Kripke models of epistemic logic. We studied the properties of the secretly knowing modalities in the form of valid formulas and validity preserving rules. These included interaction properties between secret knowledge and individual knowledge modalities, as well as between secret knowledge modalities for different agents.

The perhaps most interesting properties are those involving secretly knowing modalities for a single agent; these are the most fundamental properties of secretly knowing. We saw that the secret knowledge modalities are not normal. This should come as little surprise, given the non-standard interpretation in Kripke models using an "exist-forall" combination also found in other non-normal modal logics such as Coalition Logic (Pauly, 2002). Unlike the Coalition Logic modalities the secretly knowing modalities are not monotonic.

We gave a complete characterisation of the secret knowledge modality in the case of the weakest possible assumptions about individual knowledge (general Kripke models) considered in epistemic logic (Fagin et al., 1995) as well as under the (common) assumption that knowledge is veridial (reflexive Kripke models), by restricting the language to a single modality. These two restrictions, a single modality and weak individual knowledge, allowed us to untangle the most basic and fundamental properties of the secret knowledge modalities (that all also hold under stronger assumptions about the properties of knowledge). The main result is that secret knowledge modalities are characterised by what we called the Interpolation rule, which is weaker than Monotonicity but stronger than Equivalence, in addition to the fact that they are adjunctive (have the (C) axiom) and satisfy the (S) axiom – system **ICS** (or **ICST** if reflexivity is assumed).

Along the way we studied the Interpolation rule, which to our knowledge has not been studied in the literature, as a general sub-normal rule. In particular we showed that it is strictly between Equivalence and Monotonicity, and that in an adjunctive logic it can be used to derive the (K) axiom. While we are not aware of existing applications of the Interpolation rule, both negative and diamond necessitation, which in our case are admissible in the **S5** case but not in the **K** case, exist, in logics of agency and in deontic logics, respectively.

As **ICS** is an extension of the system **E**, the weakest system allowing neighbourhood semantics, a problem of obvious interest is to find a possibly corresponding class of neighbourhood models (using the standard interpretation of the language in neighbourhood models). We leave this for future work. However, as noted by Pacuit (2017), neighbourhood semantics is not always the best choice. Indeed, we have given a different alternative semantics, by what we called standard models. Standard models are in fact *n-ary relational models*, more specifically 3-ary in our case, proposed by Schotch and Jennings (1980). n-ary relational models were in fact originally developed for *non-adjunctive* logics (Pacuit, 2017). However, our interpretation of the language in standard models is different – a non-standard interpretation in (a particular class of) 3-ary models if you will. Finding a more natural and/or standard semantics, perhaps among the many existing semantics for weak modal logics remains a general open problem.

As pointed out by Van De Putte and McNamara (2021), modal logics with what is sometimes called "the normality schema" (K) but which nevertheless are sub-normal have been "widely neglected" when it comes to logical investigations, one possible reason being a lack of applications – maybe even a perception that it makes little sense to have (K) in a non-normal logic. In this paper we have shown that it actually does, it follows from a natural definition of secretly knowing in Kripke models. A notable exception to this neglect is (Van De Putte and McNamara, 2021) itself, which gives a constructive canonical[6] completeness proof for **EK**, **ECK** and some non-iterative relatives. We observe that one can in fact immediately get a strong completeness result for the logics **EKS** and **ECKS** with respect to neighbourhood models with properties $(k), (s)$ and $(k), (c), (s)$ (see Section 4.1), respectively, from the general results in (Van De Putte and McNamara, 2021).

If we add the assumption that individual knowledge has the S5 properties, secret knowledge inherits most of the properties of knowledge except negative introspection. This is similar to the situation for group knowledge and belief (Ågotnes and Wáng, 2021a): *general knowledge* ("everybody-knows") loses both negative and positive introspection on S5. Common knowledge, while retaining negative introspection on S5, loses it on every model class without the (B) axiom (symmetry) for individual knowledge. Negative introspection seems to be the most "fragile" property of knowledge. While the mentioned group knowledge modalities are normal, *somebody-knows* (Ågotnes and Wáng, 2021b) isn't, albeit for very different reasons than secretly-knowing: it is monotonic and has the necessitation rule, but lacks both (C) and (K).

Completeness of the $\mathcal{L}_S$ language interpreted in S5 models remains an open problem. We conjecture that the logic **ICT4S** is complete. The fact that there are relatively few existing completeness results for non-normal logics with non-iterative axioms such as (4) indicate that this might be a challenge. For example, while completeness of **ECKT** follows easily from the constructions in (Van De Putte and McNamara, 2021), we are not aware of existing completeness results for the sub-logic **ECK4**.

[6] Compared with existing non-canonical (Lewis, 1974) or non-constructive (Surendonk, 1997) proofs.

A related work is (van der Hoek and Lomuscio, 2004) where the concept of *ignorance* is studied in a very similar way to the way we studied secrets in this paper. A minimal modal language is used, with a single modality I with $I\varphi$ meaning that the agent is ignorant of whether φ (knows neither φ nor $\neg\varphi$). Like our modality, the ignorance modality can be expressed using the standard knowledge modality, but the goal is to tease out the basic principles of ignorance. To this end, a sound and complete axiomatisation is presented, for the case of no assumptions about the properties of knowledge (no restrictions on the accessibility relations, like our basic case). The ignorance modality is however, normal, the semantics being defined in terms of existential quantification over the accessibility relation. The secretly knowing modalities can be seen as a combination of knowledge and ignorance modalities, and it is this combination of universal and existential quantification that makes it sub-normal.

There are various ways our abstract definition of secretly knowing could be extended by bringing other aspects than individual knowledge and ignorance into the picture. We already mentioned intentionality in the introduction. Another is *common knowledge*. First, a stronger version of our definition of secretly knowing could be investigated, where not only a knows that b doesn't know φ, but that is common knowledge between a and b. This is similar to the pre- and post-conditions of *lying* in van Ditmarsch (2014). Second, as also mentioned in the introduction, sometimes secrets are known by a small number of people rather than a single person, and in this case common knowledge seems to play an important role. We leave further investigation to future work.

Of other conceptually related work, secrets play a key role in work on *gossip protocols* (Attamah et al., 2014, 2017; Apt et al., 2016; Apt and Wojtczak, 2018) which use logic to formalise reasoning about information flow. However, secrets are taken as a primary notion rather than derived from the underlying notion of knowledge[7], and the focus is not on the properties of secretly knowing. Also related are modal logics of access control (Abadi et al., 1993; Abadi, 2003; Garg and Abadi, 2008; Fong, 2011; Aceto et al., 2010). Some works in this area are concerned with properties of secrets of the type we consider in this paper, but they are (again) mostly taken as primary rather than derived from an underlying abstract epistemic framework.

Acknowledgements We are indebted to two anonymous AMAI reviewers, who helped us improve the paper significantly. We are grateful to Frederik Van De Putte for discussions about non-normal logics with (K) and for sharing the manuscript (Van De Putte and McNamara, 2021) with us. We also thank anonymous reviewers of the LAMAS 2020 workshop for valuable suggestions, and participants of the workshop for their comments.

[7] (Attamah et al., 2014, 2017) in fact contrasts two formalisations of "a knows the secret of b"; $K_aB \vee K_a\neg B$ vs. Kw_aB where Kw_a is a primitive "knowing-whether" modality and B is "the secret" of b. This is an interesting distinction, but it still takes the fact that B is "the secret" as a primary notion.

Conflict of Interest

Partial financial support was received by the first author by MOE Liberal arts and Social Sciences Foundation (project no. 20YJC7204002).

Data availability statement

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

References

Abadi M (2003) Logic in access control. In: 18th Annual IEEE Symposium of Logic in Computer Science, 2003. Proceedings., IEEE, pp 228–233

Abadi M, Burrows M, Lampson B, Plotkin G (1993) A calculus for access control in distributed systems. ACM Transactions on Programming Languages and Systems (TOPLAS) 15(4):706–734

Aceto L, Ingolfsdottir A, Sack J (2010) Resource bisimilarity and graded bisimilarity coincide. Information Processing Letters 111(2):68–76

Ågotnes T, Wáng YN (2021a) Group belief. Journal of Logic and Computation 31(8):1959–1978

Ågotnes T, Wáng YN (2021b) Somebody knows. In: Bienvenu M, Lakemeyer G, Erdem E (eds) Proceedings of the 18th International Conference on Principles of Knowledge Representation and Reasoning, KR 2021, pp 2–11

Apt KR, Wojtczak D (2018) Verification of distributed epistemic gossip protocols. Journal of Artificial Intelligence Research 62:101–132

Apt KR, Grossi D, van der Hoek W (2016) Epistemic protocols for distributed gossiping. arXiv preprint arXiv:160607516

Attamah M, Van Ditmarsch H, Grossi D, van der Hoek W (2014) Knowledge and gossip. In: ECAI, pp 21–26

Attamah M, van Ditmarsch H, Grossi D, van der Hoek W (2017) The pleasure of gossip. In: Başkent C, Moss LS, Ramanujam R (eds) Rohit Parikh on Logic, Language and Society, Springer, pp 145–163

Blackburn P, de Rijke M, Venema Y (2001) Modal Logic, Cambridge Tracts in Theoretical Computer Science, vol 53. Cambridge University Press

Craig W (1957) Three uses of the herbrand-gentzen theorem in relating model theory and proof theory. The Journal of Symbolic Logic 22(3):269–285

van Ditmarsch H (2014) Dynamics of lying. Synthese 191(5):745–777

Fagin R, Halpern JY, Moses Y, Vardi MY (1995) Reasoning about Knowledge. Cambridge, MA: The MIT Press

Fan J, Wang Y, Van Ditmarsch H (2015) Contingency and knowing whether. The Review of Symbolic Logic 8(1):75–107

Fong PW (2011) Relationship-based access control: protection model and policy language. In: Proceedings of the first ACM conference on Data and application security and privacy, pp 191–202

Garg D, Abadi M (2008) A modal deconstruction of access control logics. In: International Conference on Foundations of Software Science and Computational Structures, Springer, pp 216–230

Hart S, Heifetz A, Samet D (1996) "knowing whether","knowing that," and the cardinality of state spaces. journal of economic theory 70(1):249–256

Lewis D (1974) Intensional logics without iterative axioms. Journal of Philosophical Logic 3(4):457–466

Pacuit E (2017) Neighborhood semantics for modal logic. Springer

Pauly M (2002) A modal logic for coalitional power in games. Journal of Logic and Computation 12(1):149–166

Schotch PK, Jennings RE (1980) Modal logic and the theory of modal aggregation. Philosophia 9(2):265–278

Surendonk TJ (1997) Canonicity for intensional logics without iterative axioms. Journal of Philosophical Logic 26(4):391–409

Van De Putte F, McNamara P (2021) Neighbourhood canonicity for ek, eck, and relatives: A constructive proof. The Review of Symbolic Logic pp 1–17, DOI 10.1017/S1755020321000319

van der Hoek W, Lomuscio A (2004) A logic for ignorance. Electronic Notes in Theoretical Computer Science 85(2):117–133, lCMAS 2003, Logic and Communication in Multi-Agent Systems