Synthese
https://doi.org/10.1007/s11229-019-02104-0

A logic of knowing why

Chao Xu[1] · **Yanjing Wang**[1] · **Thomas Studer**[2]

Received: 6 February 2018 / Accepted: 21 January 2019

Abstract
When we say "I know why he was late", we know not only the fact that he was late, but also an explanation of this fact. We propose a logical framework of "knowing why" inspired by the existing formal studies on why-questions, scientific explanation, and justification logic. We introduce the $\mathcal{K}y_i$ operator into the language of epistemic logic to express "agent i knows why φ" and propose a Kripke-style semantics of such expressions in terms of knowing an explanation of φ. We obtain two sound and complete axiomatizations w.r.t. two different model classes depending on different assumptions about introspection. Finally we connect our logic with justification logic technically by providing an alternative semantics and an in-depth comparison on various design choices.

Keywords Knowing why · Why-questions · Scientific explanation · Epistemic logic · Justification logic · Axiomatization

1 Introduction

Ever since the seminal work by Hintikka (1962), epistemic logic has grown into a major subfield of philosophical logic, which has unexpected applications in other fields such as computer science, AI, and game theory [cf. the handbook by van Ditmarsch et al. (2015)]. Standard epistemic logic focuses on propositional knowledge expressed by "knowing that φ". However, there are various knowledge expressions in terms of "knowing whether", "knowing what", "knowing how", and so on, which have attracted

✉ Yanjing Wang
y.wang@pku.edu.cn

Chao Xu
c.xu@pku.edu.cn

Thomas Studer
tstuder@inf.unibe.ch

1 Department of Philosophy, Peking University, Beijing, China

2 Institut für Informatik, University of Bern, Bern, Switzerland

Published online: 29 January 2019

a growing interest in recent years [cf. e.g., Wang and Fan (2013), Fan et al. (2015), Wang (2018b), and the survey by Wang (2018a)].

Among those "knowing-wh",[1] "knowing why" is perhaps the most important driving force behind our advances in understanding the world and each other. For example, we may want to know why (Bird 1998):

- the window is broken.
- the lump of potassium dissolved.
- he stayed in the café all day.
- cheetahs can run at high speeds.
- blood circulates in the body.

Intuitively, each "knowing why" expression corresponds to an embedded why-question. To some extent, the process of knowing the world is to answer why-questions about the world (Hintikka 1981). In fact, there is a very general connection between knowledge and wh-questions discovered by Hinttika in the framework of quantified epistemic logic (Hintikka 1983). For example, consider the question Q : "Who murdered Mary?":

- The *presupposition* of Q is that the questioner knows that Mary was murdered by someone, formalized by $\mathcal{K}\exists x M(x, \text{Mary})$.
- The *desideratum* of Q is that the questioner knows who murdered Mary, which is formalized by $\exists x \mathcal{K} M(x, \text{Mary})$. The distinction between the desideratum and the presupposition highlights the difference between *de re* and *de dicto* readings of knowing who.
- One possible answer to Q is "John murdered Mary" formalized as $M(\text{John}, \text{Mary})$. However, telling the questioner this fact may not be enough to let the questioner know who murdered Mary since he or she may not have any idea on who John is. Therefore Hintikka also requires the following extra condition.
- *Conclusiveness* of the above answer requires that the questioner also knows who John is $(\exists x \mathcal{K}(\text{John} = x))$. Conclusive answers realize the desideratum.

However, Hintikka viewed why-questions, such as Q: "Why φ is the case?", as a special degenerated case where the presupposition and desideratum are the same:

- The *presupposition* of Q is $\mathcal{K}\varphi$;
- The *desideratum* of Q is $\mathcal{K}\varphi$.

A different logical theory of why-questions is developed by Hintikka and Halonen (1995) using the inquiry model and the interpolation theorem of first-order logic. However, we do not think why-questions are special if we can quantify over the possible answers to them. Intuitively, an answer to a question "Why φ?" is an explanation of the fact φ. In this paper, we take the view shared by Koura (1988) and Schurz (2005):

- The *presupposition* of Q is that the questioner knows that there is an explanation for the fact φ: $\mathcal{K}\exists x E(x, \varphi)$.
- The *desideratum* of Q is that the questioner knows why φ: $\exists x \mathcal{K} E(x, \varphi)$.

[1] "Wh" stands for the wh-question words.

Note that if explanations are *factive* $\exists x E(x, \varphi) \rightarrow \varphi$, then the presupposition $\mathcal{K}\exists x E(x, \varphi)$ also implies $\mathcal{K}\varphi$ in standard quantified (normal) modal logic.

Now we have a preliminary logical form of knowing why in terms of the desideratum $\exists x \mathcal{K} E(x, \varphi)$ of the corresponding why-question. The next questions are:

1. What are (good) *explanations*?
2. How can we capture the relation (E above) between an explanation and a proposition in logic?

The two questions are clearly related. To answer the first one, let us look back at the examples we mentioned at the beginning of this introduction. In fact there are different kinds of explanations (Bird 1998):

- Causal: The window broke because the stone was thrown at it.
- Nomic:[2] The lump of potassium dissolved since as a law of nature potassium reacts with water to form a soluble hydroxide.
- Psychological: He stayed in the café all day hoping to see her again.
- Darwinian: Cheetahs can run at high speeds because of the selective advantage this gives them in catching their prey.
- Functional: Blood circulates in order to supply the various parts of the body with oxygen and nutrients.

In philosophy of science, the emphasis is on *scientific explanations* to why questions, which mainly involve Nomic and Causal explanations in the above categorization (Bromberger 1966; Koura 1988; van Fraassen 1980). According to Schurz (1999) there are three major paradigms in understanding (scientific) explanations:[3]

- The *nomic expectability approach* initiated by Hempel (1965), where a good explanation to φ should make the explanandum φ predictable or increases φ's expectability.
- The *causality approach* [cf. e.g, Salmon (1984)], where an explanation to φ should give a complete list of causes or relevant factors to φ.
- The unification approach [cf. e.g., Kitcher (1981)] where the focus is on the global feature of explanations in a coherent picture.

Our initial inspiration comes from the deductive-nomological model proposed by Hempel and Oppenheim (1948) in the first approach mentioned above, which is the mostly discussed (and criticized) model of explanation. The basic idea is that an explanation is a *derivation* of the explanandum from some universally quantified laws and some singular sentences. Although such a logical empiricist's approach arouse debates for decades,[4] it draws our attention to the inner structure of explanations and its similarity to derivations in logic. In this paper, as the first step towards a logic of knowing why, we would like to stay neutral on different types of explanations and their models, and focus on the most abstract logical structure of (scientific) explanations.

[2] Nomic explanations are explanation in terms of laws of nature.

[3] There are also various dimensions of each paradigm, e.g, probabilistic versus non-probabilistic, singular events or laws to be explained.

[4] See Schurz (1995) and Weber et al. (2013) for critical surveys.

From a structuralist point of view, we only need to know how explanations compose and interact with each other without saying what they are exactly.

Now, as for the second question, how can we capture the explanatory relation between explanations and propositions in logic? Our next crucial inspiration came from *Justification Logic* proposed by Artemov (2008). Aiming at making up the gap between epistemic logic and the mainstream epistemology where justified true belief is the necessary basis of knowledge, justification logics are introduced based on the ideas of *Logic of Proof* (LP) by Artemov (1995, 2001).[5] Justification logic introduces formulas in the shape of $t:\varphi$ into the logical language, read as "t is a justification of φ".[6] Therefore, in justification logic we can talk about knowledge with an *explicit* justification. Moreover, justifications can be composed using various operations. For example, $t:(\varphi \to \psi) \wedge s:\varphi \to (t \cdot s):\psi$ is an axiom in the standard justification logic where $\cdot$ is the application operation of two justifications. Explanations may have similar compositional structures. If t is an explanation of the fact φ, and s is an explanation (e.g., a logical proof) for the material implication of $\varphi \to \psi$, then combining s and t should in principle give an explanation of the fact ψ.

On the other hand, conceptually, justifications are quite different from explanations. For example, the fact that the shadow of a flagpole is x meters long may justify that the length of the pole is y meters given the specific time and location on earth. However, the length of the shadow of a flagpole clearly does not explain why the pole is y meters long, if we are looking for causal explanations. In general, a justification of φ gives a reason to *believing* φ (though not necessarily true), but an explanation gives a reason to *being* φ, presupposing the truth of φ. In this paper, we only make use of some technical apparatus of justification logic, and there are quite some differences in our framework compared to justification logic, which will be discussed in Sect. 4.

Putting all the above ideas together, we are almost ready to lay out the basis of our logic of knowing why. Following Wang (2018a), we enrich the standard (multi-agent) epistemic language with a new "knowing why" operator $\mathcal{K}y_i$, instead of using a quantified modal language. Roughly speaking, $\mathcal{K}y_i\varphi$ is essentially $\exists t\mathcal{K}_i(t:\varphi)$, although we do not allow quantifiers and terms in the logical language. As in Wang (2015, 2018b) and Wang and Fan (2014), packing the quantifiers and modalities together will help us to control the expressive power of the logic in hopes of a computationally well-behaved logic. For a more general technical discussion of the advantages of such "packed" modalities, see Wang (2017). The semantics is based on the idea of Fitting model for justification logic.

Since the language has both the standard epistemic operator $\mathcal{K}_i$ and also the new "knowing why" modality $\mathcal{K}y_i$, there are lots of interesting things that can be expressed. For example,

- $\mathcal{K}_i p \wedge \neg\mathcal{K}y_i p$, e.g, I know that Fermat's last theorem is true but I do not know why.
- $\neg\mathcal{K}y_i p \wedge \mathcal{K}_i\mathcal{K}y_j p$, e.g., I do not know why Fermat's last theorem holds but I know that Andrew Wiles knows why.

[5] LP was invented to give an arithmetic semantics to intuitionistic logic under the Brouwer-Heyting-Kolmogorov provability interpretation.

[6] Similar ideas also appeared in van Benthem (1991).

- $\mathcal{K}_i\mathcal{K}_j p \wedge \neg \mathcal{K}y_i\mathcal{K}_j p$, e.g., I know that you know that the paper has been accepted, but I do not know why you know.
- $\mathcal{K}y_i\mathcal{K}_j p \wedge \mathcal{K}_i\neg \mathcal{K}y_j p$, e.g., I know why you know that the paper has been rejected, but I am sure you do not know why.

As we will see later, these situations are all satisfiable in our models.[7]

Before going into the technical details, it is helpful to summarize the aforementioned ideas:

- The language is inspired by the treatments of the logics of "knowing what", and "knowing how", where new modalities of such constructions are introduced, without using the full language of quantified epistemic logic.
- The formal treatment of explanations is inspired by the formal account of justifications in justification logics.
- The semantics of $\mathcal{K}y_i$ is inspired by Hintikka's logical formulation of the desideratum of Wh-questions: $\exists t\mathcal{K}_i(t{:}\varphi)$.

In the rest of the paper, Sect. 2 lays out the language, semantics and two proof systems of our knowing why logic; Sect. 3 proves the completeness of the two systems, and gives an alternative semantics closer to the standard justification logic; Sect. 4 gives a detailed comparison with various versions of justification logic; Sect. 5 concludes the paper with discussions and future directions.

2 A logic of knowing why

Definition 1 (*Language* **ELKy**) Given a countable set **I** of agent names and a countably infinite set **P** of basic propositional letters, the language of **ELKy** is defined as

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \mathcal{K}_i\varphi \mid \mathcal{K}y_i\varphi$$

where $p \in \mathbf{P}$ and $i \in \mathbf{I}$.

We use standard abbreviations for $\top$, $\bot$, $\varphi \to \psi$, $\varphi \vee \psi$, and $\widehat{\mathcal{K}_i}\varphi$ (the dual of $\mathcal{K}_i\varphi$). $\mathcal{K}y_i\varphi$ says that agent i knows why φ (is the case).

Intuitively, necessitation rule for $\mathcal{K}y_i$ should not hold, e.g., although something is a tautology, you may not know why it is a tautology. Borrowing the idea from justification logic, we introduce a special set of "self-evident" tautologies which the agents are assumed to know why. Please see Sect. 4 for the comparison with *constant specifications* in justification logic, which may contain all axioms of the logic.

Definition 2 (*Tautology Ground* Λ) Tautology Ground Λ is a set of propositional tautologies.

For example, Λ can be the set of all the instances of $\varphi \wedge \psi \to \varphi$ and $\varphi \wedge \psi \to \psi$. As we will see later, under such a Λ, $\mathcal{K}y_i(\varphi \wedge \psi \to \varphi)$ will be valid, which helps the agents to reason more.

[7] According to our semantics to be introduced later, it is also allowed to know why for different reasons (for different people), which might help to model mutual misunderstanding.

The model of our language **ELKy** is similar to the Fitting model of justification logic (Fitting 2005). Note that we do not have the justification terms in the logical language, but we do have a set E of explanations as semantic objects in the models. In this work, we require the accessibility relation to be equivalence relations to accommodate the S5 epistemic logic.

Definition 3 (**ELKy**-*Model*) An **ELKy**-model $\mathcal{M}$ is a tuple

$$(W, E, \{R_i \mid i \in \mathbf{I}\}, \mathcal{E}, V)$$

where:

- W is a non-empty set of possible worlds.
- E is a non-empty set of explanations satisfying the following conditions
  - (a) If $s, t \in E$, then a new explanation $(s \cdot t) \in E$;
  - (b) A special symbol e is in E.
- $R_i \subseteq W \times W$ is an equivalence relation over W.
- $\mathcal{E} : E \times \mathbf{ELKy} \to 2^W$ is an *admissible explanation function* satisfying the following conditions:
  - (I) $\mathcal{E}(s, \varphi \to \psi) \cap \mathcal{E}(t, \varphi) \subseteq \mathcal{E}(s \cdot t, \psi)$.
  - (II) If $\varphi \in \Lambda$, then $\mathcal{E}(e, \varphi) = W$.
- $V : \mathbf{P} \to 2^W$ is a valuation function.

Note that E does not depend on possible worlds, thus it can be viewed as a *constant domain* of explanations closed under an *application* operator $\cdot$ which combines two explanations into one. Note that we do not have other operators such as sum $(+)$ as in justification logic. We will come back to this with an in-depth comparison with justification logic in Sect. 4. The special element e in E is the *self-evident explanation*, which is uniform for all the self-evident formulas in Λ. The admissible explanation function $\mathcal{E}$ specifies the set of worlds where t is an explanation of φ. It is possible that some formula has *no* explanation at all on some world,[8] and some formula has more than one explanation on some world, e.g., one theorem may have different proofs.[9] The first condition of $\mathcal{E}$ captures the composition of explanations resembling the reasoning of knowing why by *modus ponens*, which amounts to the later axiom $\mathcal{K}y_i(\varphi \to \psi) \to (\mathcal{K}y_i\varphi \to \mathcal{K}y_i\psi)$.[10]

Definition 4 (*Semantics*)
The satisfaction relation of **ELKy** formulas on pointed models is as below:

[8] E.g., there is no proof for the consistency of PA within PA, given PA is consistent.

[9] Though as Johan van Benthem pointed out via personal communication, in many cases there is often a best or shortest explanation for the fact.

[10] Note that here $\to$ is a material implication which should not be treated as an arbitrary conditional.

$$
\begin{array}{ll}
\mathcal{M}, w \vDash p & \Longleftrightarrow w \in V(p) \\
\mathcal{M}, w \vDash \neg\varphi & \Longleftrightarrow \mathcal{M}, w \nvDash \varphi \\
\mathcal{M}, w \vDash \varphi \wedge \psi & \Longleftrightarrow \mathcal{M}, w \vDash \varphi \text{ and } \mathcal{M}, w \vDash \psi \\
\mathcal{M}, w \vDash \mathcal{K}_i\varphi & \Longleftrightarrow \mathcal{M}, v \vDash \varphi \text{ for each } v \text{ such that } w R_i v. \\
\mathcal{M}, w \vDash \mathcal{K}y_i\varphi & \Longleftrightarrow (1)\ \exists t \in E, \text{for each } v \text{ such that } w R_i v,\ v \in \mathcal{E}(t, \varphi); \\
 & \quad\ (2)\ \forall v \in W,\ w R_i v,\ v \vDash \varphi.
\end{array}
$$

Now it is clear that our $\mathcal{K}y_i\varphi$ is roughly $\exists t\mathcal{K}_i(t{:}\varphi) \wedge \mathcal{K}_i\varphi$ though there are subtle details to be discussed in Sect. 4 when compared to justification logic. Also note that $\mathcal{K}y_i\varphi \rightarrow \mathcal{K}_i\varphi$ is clearly valid, but $\mathcal{K}y_i$-necessitation is not since not all valid formulas are explained except those in Λ. Moreover, things we usually take for granted are not valid either, e.g., $\mathcal{K}y_i\varphi \wedge \mathcal{K}y_i\psi \rightarrow \mathcal{K}y_i(\varphi \wedge \psi)$ is not valid in general: The fact that I have explanations for φ and ψ, respectively, does not mean that I have an explanation for the co-occurrence of the two, e.g., quantum mechanics and general relativity have their own explanatory power on microcosm and macrocosm, respectively, but a "theory of everything" is not obtained by simply putting these two theories together.

As an example, in the following model (reflexive arrows are omitted), the formula $\mathcal{K}_i p \wedge \neg\mathcal{K}y_i p \wedge \mathcal{K}y_j p \wedge \mathcal{K}_i\mathcal{K}y_j p$ holds on the middle world, where we use $t :: \varphi$ on a world w to mean $w \in \mathcal{E}(t, \varphi)$.

$$
\begin{array}{ccccc}
p & & p & & p \\
t'{:}\, p & \text{—}i\text{—} & t{:}\, p & \text{—}j\text{—} & s{:}\, p \\
 & & s{:}\, p & &
\end{array}
$$

In this paper, we also consider models with special properties. First of all, we are interested in the models where explanations are always correct, i.e., if a proposition has an explanation on a world, then it must be true.

Definition 5 (*Explanation Factivity*) An **ELKy**-model $\mathcal{M}$ has the explanation factivity provided that, whenever $w \in \mathcal{E}(t, \varphi)$, then $\mathcal{M}, w \vDash \varphi$.

Note that this property is different from the factivity for KWi, i.e., $\mathcal{K}y_i\varphi \rightarrow \varphi$ is valid. Philosophically explanation factivity is debatable,[11] but as we will see later in Theorem 10, our logics stay neutral on it: assuming it or not will not affect the logics.

Besides explanation factivity, it is also debatable whether knowing why is introspective. Based on the current semantics, it is not hard to see that the following introspection axioms on knowing why are valid:

$$\mathcal{K}y_i\varphi \rightarrow \mathcal{K}_i\mathcal{K}y_i\varphi, \quad \neg\mathcal{K}y_i\varphi \rightarrow \mathcal{K}_i\neg\mathcal{K}y_i\varphi$$

However, what about the following ones? Note that they are not valid without further conditions on our models.

$$\mathcal{K}_i\varphi \rightarrow \mathcal{K}y_i\mathcal{K}_i\varphi, \quad \neg\mathcal{K}_i\varphi \rightarrow \mathcal{K}y_i\neg\mathcal{K}_i\varphi,$$
$$\mathcal{K}y_i\varphi \rightarrow \mathcal{K}y_i\mathcal{K}y_i\varphi, \quad \neg\mathcal{K}y_i\varphi \rightarrow \mathcal{K}y_i\neg\mathcal{K}y_i\varphi$$

[11] Martin Stokhof suggested an example where one explains to the other why he is the best candidate for a job, but in fact he is not, and his explanation may base on false premises.

One may argue that there is always a self-evident explanation to your own knowledge or ignorance, but another may say it happens a lot that you just forgot why you know some facts. Things can be even more complicated regarding nested $\mathcal{K}y_i$. Your explanation for why φ holds may be quite different from the explanation for why you know why φ, e.g., the window is broken (φ) because you know a stone was thrown at it, and you know why you know why φ because someone told you so. On the other hand, if you know why a theorem holds because of a proof, it seems reasonable to assume that you know why you know why the theorem holds: you can just verify the proof. The cases of negative introspection may invoke more debates.

As a first attempt to a logic of knowing why, we want to remain neutral in the philosophical debate, but would like to make it technically possible to handle the cases when introspection is considered reasonable. The following property guarantees that the above introspection axioms are valid.

Definition 6 (*Introspection Property*) An **ELKy**-model $\mathcal{M}$ has the introspection property provided that, whenever $\mathcal{M}, w \vDash \varphi$ and φ has the form of $\mathcal{K}_i\psi$ or $\neg\mathcal{K}_i\psi$ or $\mathcal{K}y_i\psi$ or $\neg\mathcal{K}y_i\psi$, then $\exists t \in E$, for each v such that wR_iv, $v \in \mathcal{E}(t, \varphi)$.

We use $\mathbb{C}$, $\mathbb{C}_F$, $\mathbb{C}_I$, $\mathbb{C}_{FI}$ to denote respectively the model classes of all **ELKy**-models, factive models, introspective models, and models with both properties. Obviously, we have $\mathbb{C}_F \subseteq \mathbb{C}$, $\mathbb{C}_I \subseteq \mathbb{C}$, $\mathbb{C}_{FI} \subseteq \mathbb{C}_F$, and $\mathbb{C}_{FI} \subseteq \mathbb{C}_I$. In the following, we write $\Gamma \vDash_{\mathbb{C}} \varphi$ if $\mathcal{M}, w \vDash \Gamma$ implies $\mathcal{M}, w \vDash \varphi$, for any $\mathcal{M} \in \mathbb{C}$ and any w in $\mathcal{M}$. Similar for $\mathbb{C}_F$, $\mathbb{C}_I$, $\mathbb{C}_{FI}$.

However, as we will see below, factivity does not affect the valid formulas. For an arbitrary $\mathcal{M} \in \mathbb{C}$, we can construct a new **ELKy**-model $\mathcal{M}^F \in \mathbb{C}_F$ which has factivity. Given $\mathcal{M} = (W, E, \{R_i \mid i \in \mathbf{I}\}, \mathcal{E}, V)$, let $\mathcal{M}^F = (W, E, \{R_i \mid i \in \mathbf{I}\}, \mathcal{E}^F, V)$ where:

$$\mathcal{E}^F(t, \varphi) = \mathcal{E}(t, \varphi) - \{u \mid \mathcal{M}, u \nvDash \varphi\}$$

We will show that $\mathcal{M}, w$ and $\mathcal{M}^F, w$ satisfy the same **ELKy** formulas, thus by the above definition of $\mathcal{E}^F$, it is clear that $\mathcal{M}^F$ has explanation factivity.

Lemma 7 *For any* **ELKy***-formula* φ *and any* $w \in W$, $\mathcal{M}, w \vDash \varphi$ *if and only if* $\mathcal{M}^F, w \vDash \varphi$.

Proof We can prove it by induction on the structure of formulas. It is trivial for the atomic, boolean, and $\mathcal{K}\psi$ cases since $\mathcal{M}^F$ only differs from $\mathcal{M}$ in $\mathcal{E}^F$. We just need to prove that $\mathcal{M}, w \vDash \mathcal{K}y_i\psi$ iff $\mathcal{M}^F, w \vDash \mathcal{K}y_i\psi$.

$\Rightarrow$ Suppose $\mathcal{M}, w \vDash \mathcal{K}y_i\psi$. Then $\exists t \in E$, for each v such that wR_iv, $v \in \mathcal{E}(t, \psi)$ and $v \vDash \psi$. Thus $v \notin \{u \mid \mathcal{M}, u \nvDash \psi\}$. Therefore we have $v \in \mathcal{E}^F(t, \psi)$. Hence by IH we have $\mathcal{M}^F, w \vDash \mathcal{K}y_i\psi$.

$\Leftarrow$ Suppose $\mathcal{M}^F, w \vDash \mathcal{K}y_i\psi$. Then $\exists t \in E$, for each v such that wR_iv, $v \in \mathcal{E}^F(t, \psi)$ and $v \vDash \psi$. By the definition of $\mathcal{E}^F$, we have $v \in \mathcal{E}(t, \psi)$. Hence by IH we get $\mathcal{M}, w \vDash \mathcal{K}y_i\psi$. □

Theorem 8 *For any set $\Gamma \cup \{\varphi\}$ of formulas, $\Gamma \models_{\mathbb{C}} \varphi$ if and only if $\Gamma \models_{\mathbb{C}_F} \varphi$.*

Proof $\Rightarrow$ Suppose $\Gamma \models_{\mathbb{C}} \varphi$ and $\Gamma \not\models_{\mathbb{C}_F} \varphi$. By $\Gamma \not\models_{\mathbb{C}_F} \varphi$, there exists a factive model $\mathcal{N} \in \mathbb{C}_F$ such that $\mathcal{N}, w \models \Gamma$ and $\mathcal{N}, w \not\models \varphi$ for some w in $\mathcal{N}$. Since $\mathbb{C}_F \subseteq \mathbb{C}$, we have $\mathcal{N} \in \mathbb{C}$. Thus $\Gamma \not\models_{\mathbb{C}} \varphi$. Contradiction.

$\Leftarrow$ Suppose $\Gamma \models_{\mathbb{C}_F} \varphi$ and $\Gamma \not\models_{\mathbb{C}} \varphi$. Then there exists a model $\mathcal{M} \in \mathbb{C}$ such that $\mathcal{M} \models \Gamma$ and $\mathcal{M} \not\models \varphi$. By Lemma 7, we can construct an $\mathcal{M}^F \in \mathbb{C}_F$ such that $\mathcal{M}^F \models \Gamma$ and $\mathcal{M}^F \not\models \varphi$. Thus $\Gamma \not\models_{\mathbb{C}_F} \varphi$. Contradiction. □

The above theorem is due to the lack of expressivity of our language: we cannot express that $t{:}\varphi$ as in justification logic. Then $\mathcal{M}^F$ can simply ignore the non-factive explanations for each φ without affecting the truth value of **ELKy** formulas.[12] Now we consider the introspective models.

Lemma 9 *If $\mathcal{M}$ is introspective, then so is $\mathcal{M}^F$.*

Proof Suppose $\mathcal{M}^F, w \models \varphi$ and φ has the form of $\mathcal{K}_i\psi$ or $\neg\mathcal{K}_i\psi$ or $\mathcal{K}y_i\psi$ or $\neg\mathcal{K}y_i\psi$. By Lemma 7, we have $\mathcal{M}, w \models \varphi$. Since $\mathcal{M}$ has introspection property, we have that $\exists t \in E$, for each v such that wR_iv, $v \in \mathcal{E}(t, \varphi)$. Since $\mathcal{M}^F, w \models \varphi$ and R_i is an equivalence relation, we have $\mathcal{M}^F, v \models \varphi$ for each v such that wR_iv. Thus $v \notin \{u \mid \mathcal{M}, u \not\models \varphi\}$. Thus $v \in \mathcal{E}^F(t, \varphi)$. Hence $\exists t \in E$, for each v such that wR_iv, $v \in \mathcal{E}^F(t, \varphi)$. Therefore $\mathcal{M}^F$ has introspection property. □

It is then easy to show:

Theorem 10 *For any set $\Gamma \cup \{\varphi\}$, $\Gamma \models_{\mathbb{C}_I} \varphi$ if and only if $\Gamma \models_{\mathbb{C}_{FI}} \varphi$.*

Theorems 8 and 10 showed that explanation factivity is neglectable w.r.t. the logic.

In the following, we present two proof systems which differ only on the introspection axioms of $\mathcal{K}y_i$ essentially. In the next section, we will show their completeness w.r.t. $\mathbb{C}$ and $\mathbb{C}_I$, respectively.

System $\mathbb{SKY}$

TAUT	Classical Propositional Axioms
DISTK	$\mathcal{K}_i(\varphi \to \psi) \to (\mathcal{K}_i\varphi \to \mathcal{K}_i\psi)$
DISTY	$\mathcal{K}y_i(\varphi \to \psi) \to (\mathcal{K}y_i\varphi \to \mathcal{K}y_i\psi)$
T	$\mathcal{K}_i\varphi \to \varphi$
4	$\mathcal{K}_i\varphi \to \mathcal{K}_i\mathcal{K}_i\varphi$
5	$\neg\mathcal{K}_i\varphi \to \mathcal{K}_i\neg\mathcal{K}_i\varphi$
IMP	$\mathcal{K}y_i\varphi \to \mathcal{K}_i\varphi$
4YK	$\mathcal{K}y_i\varphi \to \mathcal{K}_i\mathcal{K}y_i\varphi$

MP	Modus Ponens
NECK	$\vdash \varphi \Rightarrow \vdash \mathcal{K}_i\varphi$
NECKY	If $\varphi \in \Lambda$, then $\vdash \mathcal{K}y_i\varphi$

IMP says "knowing that" is necessary for "knowing why". 4YK is the positive introspection of "knowing why" by "knowing that".[13] The reader may wonder about the corresponding negative introspection of 4YK and it is provable in $\mathbb{SKY}$.

[12] The situation is like in Padmanabha et al. (2018) where the bundled modality $\exists\Box$ cannot distinguish constant domain and increasing domain models.

[13] Note that this is not one of the four introspection axioms of $\mathcal{K}y_i$ mentioned earlier.

Proposition 11 *5YK:* $\neg \mathcal{K}y_i\varphi \to \mathcal{K}_i\neg\mathcal{K}y_i\varphi$ *is provable in* $\mathbb{SKY}$.

Proof

(1)	$\mathcal{K}_i\mathcal{K}y_i\varphi \to \mathcal{K}y_i\varphi$	T
(2)	$\neg\mathcal{K}y_i\varphi \to \neg\mathcal{K}_i\mathcal{K}y_i\varphi$	Contraposition (1)
(3)	$\neg\mathcal{K}_i\mathcal{K}y_i\varphi \to \mathcal{K}_i\neg\mathcal{K}_i\mathcal{K}y_i\varphi$	5
(4)	$\mathcal{K}y_i\varphi \to \mathcal{K}_i\mathcal{K}y_i\varphi$	4YK
(5)	$\neg\mathcal{K}_i\mathcal{K}y_i\varphi \to \neg\mathcal{K}y_i\varphi$	Contraposition (4)
(6)	$\mathcal{K}_i(\neg\mathcal{K}_i\mathcal{K}y_i\varphi \to \neg\mathcal{K}y_i\varphi)$	NECK(5)
(7)	$\mathcal{K}_i\neg\mathcal{K}_i\mathcal{K}y_i\varphi \to \mathcal{K}_i\neg\mathcal{K}y_i\varphi$	MP(6), DISTK
(8)	$\neg\mathcal{K}y_i\varphi \to \mathcal{K}_i\neg\mathcal{K}_i\mathcal{K}y_i\varphi$	MP(2)(3)
(9)	$\neg\mathcal{K}y_i\varphi \to \mathcal{K}_i\neg\mathcal{K}y_i\varphi$	MP(7)(8)

□

Note that the choice of Λ and NECKY in $\mathbb{SKY}$ also give us some flexibility in the logic.

System $\mathbb{SKYI}$ is obtained by replacing 4, 5 and 4YK in $\mathbb{SKY}$ by the those four stronger introspection axioms of $\mathcal{K}y_i$:

4KY $\mathcal{K}_i\varphi \to \mathcal{K}y_i\mathcal{K}_i\varphi$	4Y $\mathcal{K}y_i\varphi \to \mathcal{K}y_i\mathcal{K}y_i\varphi$
5KY $\neg\mathcal{K}_i\varphi \to \mathcal{K}y_i\neg\mathcal{K}_i\varphi$	5Y $\neg\mathcal{K}y_i\varphi \to \mathcal{K}y_i\neg\mathcal{K}y_i\varphi$

It is straightforward to show that $\mathbb{SKYI}$ is deductively stronger than $\mathbb{SKY}$.

Proposition 12 *The following are provable in* $\mathbb{SKYI}$

4 $\mathcal{K}_i\varphi \to \mathcal{K}_i\mathcal{K}_i\varphi$	*4YK* $\mathcal{K}y_i\varphi \to \mathcal{K}_i\mathcal{K}y_i\varphi$
5 $\neg\mathcal{K}_i\varphi \to \mathcal{K}_i\neg\mathcal{K}_i\varphi$	*5YK* $\neg\mathcal{K}y_i\varphi \to \mathcal{K}_i\neg\mathcal{K}y_i\varphi$

3 Soundness and completeness

Due to Theorems 8 and 10, we only need to prove soundness and completeness w.r.t. $\mathbb{C}$ and $\mathbb{C}_I$ instead of $\mathbb{C}_F$ and $\mathbb{C}_{FI}$ respectively.

Theorem 13 (*Soundness*) $\mathbb{SKY}$ *and* $\mathbb{SKYI}$ *are sound for* $\mathbb{C}$ *and* $\mathbb{C}_I$ *respectively.*

Proof Since **ELKy**-models are based on S5 Kripke models, the standard axioms of system $\mathbb{S}5$ are all valid. So we just need to check the rest. First we check the non-trivial axioms and rules of $\mathbb{SKY}$ on $\mathbb{C}$.

DISTY: $\mathcal{K}y_i(\varphi \to \psi) \to (\mathcal{K}y_i\varphi \to \mathcal{K}y_i\psi)$

Suppose $w \vDash \mathcal{K}y_i(\varphi \to \psi)$ and $w \vDash \mathcal{K}y_i\psi$. Then by the definition of $\vDash$, $\exists s, t \in E$, for any v such that wR_iv, $v \in \mathcal{E}(s, \varphi \to \psi)$, $v \in \mathcal{E}(t, \varphi)$, $v \vDash \varphi \to \psi$, and $v \vDash \varphi$. We find $v \vDash \psi$ and $v \in \mathcal{E}(s, \varphi \to \psi) \cap \mathcal{E}(t, \varphi)$. By the condition (I) of $\mathcal{E}$, we have $v \in \mathcal{E}(s \cdot t, \psi)$. Hence $w \vDash \mathcal{K}y_i\psi$.

IMP: $\mathcal{K}y_i\varphi \to \mathcal{K}_i\varphi$

Suppose $w \vDash \mathcal{K}y_i\varphi$. Then for any v such that wR_iv, we have $v \vDash \varphi$. Thus $w \vDash \mathcal{K}_i\varphi$.

4YK: $\mathcal{K}y_i\varphi \to \mathcal{K}_i\mathcal{K}y_i\varphi$

Suppose $w \vDash \mathcal{K}y_i\varphi$. Then $\exists t \in E$, for any v such that wR_iv, $v \in \mathcal{E}(t, \varphi)$ and $v \vDash \varphi$. Let $u \in W$ be arbitrary with wR_iu. Since R_i is transitive, we find that uR_iv implies wR_iv. Thus $u \vDash \mathcal{K}y_i\varphi$. We conclude that $w \vDash \mathcal{K}_i\mathcal{K}y_i\varphi$.

NECKY Suppose $\varphi \in \Lambda$. Since Λ is a set of tautologies, we have $\forall w \in W, w \vDash \varphi$. By the condition (II) of $\mathcal{E}$, $\forall w \in W, \exists e \in E$, for any v such that wR_iv, $v \in \mathcal{E}(e, \varphi)$. Therefore it follows that $\vDash \mathcal{K}y_i\varphi$. Hence NECKY is valid.

Validity of the introspection axioms of $\mathbb{SKYI}$ on $\mathbb{C}_I$ are trivial based on the introspective property and the fact that R_i is an equivalence relation. □

Remark 1 Note that the rule of replacement is not valid, e.g., the validity of $\varphi \leftrightarrow \psi$ does not entail the validity of $\mathcal{K}y_i\varphi \leftrightarrow \mathcal{K}y_i\psi$.

To establish completeness, we build a canonical model for each consistent set of **ELKy** formulas. We will first show the completeness of $\mathbb{SKY}$ over $\mathbb{C}$, and the completeness of $\mathbb{SKYI}$ over $\mathbb{C}_I$ is then straightforward.

Let Ω be the set of all maximal $\mathbb{SKY}$-consistent sets of formulas. For any maximal consistent set (abbr. MCS) Γ, let $\Gamma_i^{\#} = \{\mathcal{K}y_i\varphi \mid \mathcal{K}y_i\varphi \in \Gamma\} \cup \{\varphi \mid \mathcal{K}_i\varphi \in \Gamma\}$.

Definition 14 (*Canonical model for* $\mathbb{SKY}$) The canonical model $\mathcal{M}^c$ for $\mathbb{SKY}$ is a tuple $(W^c, E^c, \{R_i^c \mid i \in \mathbf{I}\}, \mathcal{E}^c, V^c)$ where:

- E^c is defined in BNF: $t ::= e \mid \varphi \mid (t \cdot t)$ where $\varphi \in \mathbf{ELKy}$.
- $W^c = \{\langle \Gamma, F, \{f_i \mid i \in \mathbf{I}\}\rangle \mid \langle \Gamma, F\rangle \in \Omega \times \mathcal{P}(E^c \times \mathbf{ELKy}), f_i : \{\varphi \mid \mathcal{K}y_i\varphi \in \Gamma\} \to E^c$ such that F and $\vec{f}$ satisfy the conditions below$\}$:
  - (i) If $\langle s, \varphi \to \psi\rangle, \langle t, \varphi\rangle \in F$, then $\langle s \cdot t, \psi\rangle \in F$;
  - (ii) If $\varphi \in \Lambda$, then $\langle e, \varphi\rangle \in F$;
  - (iii) For any $i \in \mathbf{I}$, $\mathcal{K}y_i\varphi \in \Gamma$ implies $\langle f_i(\varphi), \varphi\rangle \in F$.
- $\langle \Gamma, F, \vec{f}\rangle R_i^c \langle \Delta, G, \vec{g}\rangle$ iff (1) $\Gamma_i^{\#} \subseteq \Delta$, and (2) $f_i = g_i$.
- $\mathcal{E}^c : E^c \times \mathbf{ELKy} \to 2^{W^c}$ defined by $\mathcal{E}^c(t, \varphi) = \{\langle \Gamma, F, \vec{f}\rangle \mid \langle t, \varphi\rangle \in F\}$.
- $V^c(p) = \{\langle \Gamma, F, \vec{f}\rangle \mid p \in \Gamma\}$.

In the above we write $\vec{f}$ for $\{f_i \mid i \in \mathbf{I}\}$. Essentially, f_i is a witness function picking one t for each formula in $\{\varphi \mid \mathcal{K}y_i\varphi \in \Gamma\}$. It can be used to construct the possible worlds for the existence lemma for $\neg\mathcal{K}y_i\varphi$. We do need such witness functions for each i, since i, j can have different explanations for φ. In the definition of R_i^c, we need to make sure the selected witnesses are the same for i. We include $\varphi \in \mathbf{ELKy}$ as building blocks in E^c for technical convenience, as it will become more clear below when we construct the successors. The component F in each world is used to encode the information of $\mathcal{E}^c$ locally, also for the technical convenience to define the canonical relations. Note that merely maximal consistent sets are not enough in constructing the canonical model, as in the case of the logic of knowing what by Wang and Fan (2013, 2014).

Now we need to show that the canonical model is well-defined:

- $\mathcal{E}^c$ satisfies conditions (I) and (II) in the definition of **ELKy**-models.
- R_i^c is an equivalence relation.
- W^c is not empty. Actually, we will prove a stronger one: for any $\Gamma \in \Omega$, there exist F and $\vec{f}$ such that $\langle \Gamma, F, \vec{f}\rangle \in W^c$.

Proposition 15 *$\mathcal{E}^c$ satisfies the conditions (I) and (II) of* **ELKy***-models.*

Proof (1) Suppose $\langle \Gamma, F, \vec{f}\rangle \in \mathcal{E}^c(s, \varphi \to \psi) \cap \mathcal{E}^c(t, \varphi)$. By the definition of $\mathcal{E}^c$, we have $\langle s, \varphi \to \psi\rangle, \langle t, \varphi\rangle \in F$. By the condition (i) of F in the definition of W^c, we have $\langle s \cdot t, \psi\rangle \in F$. Hence it follows that $\langle \Gamma, F, \vec{f}\rangle \in \mathcal{E}^c(s \cdot t, \psi)$. Therefore $\mathcal{E}^c(s, \varphi \to \psi) \cap \mathcal{E}^c(t, \varphi) \subseteq \mathcal{E}^c(s \cdot t, \psi)$.

(2) Suppose $\varphi \in \Lambda$. For an arbitrary $\langle \Gamma, F, \vec{f}\rangle \in W^c$, by condition (ii) in the definition of W^c, we have $\langle e, \varphi\rangle \in F$. By the definition of $\mathcal{E}^c$, we have $\langle \Gamma, F, \vec{f}\rangle \in \mathcal{E}^c(e, \varphi)$. Hence $\mathcal{E}^c(e, \varphi) = W^c$. □

Before proceeding further, we prove the following handy proposition.

Proposition 16 *If $\langle \Gamma, F, \vec{f}\rangle R_i^c \langle \Delta, G, \vec{g}\rangle$, then (1) $\mathcal{K}y_i\varphi \in \Gamma$ iff $\mathcal{K}y_i\varphi \in \Delta$ and (2) $\mathcal{K}_i\varphi \in \Gamma$ iff $\mathcal{K}_i\varphi \in \Delta$.*

Proof (1) Suppose $\mathcal{K}y_i\varphi \in \Gamma$. By the definition of R_i^c, we have $\mathcal{K}y_i\varphi \in \Delta$.

Suppose $\mathcal{K}y_i\varphi \in \Delta$ and $\mathcal{K}y_i\varphi \notin \Gamma$. By the property of MCS, we have $\neg\mathcal{K}y_i\varphi \in \Gamma$. By the provable 5YK ($\neg\mathcal{K}y_i\varphi \to \mathcal{K}_i\neg\mathcal{K}y_i\varphi$) and the property of MCS, we have $\mathcal{K}_i\neg\mathcal{K}y_i\varphi \in \Gamma$. By the definition of R_i^c, we have $\neg\mathcal{K}y_i\varphi \in \Delta$. Contradiction.

(2) Suppose $\mathcal{K}_i\varphi \in \Gamma$. By axiom 4 and the property of MCS, we have $\mathcal{K}_i\mathcal{K}_i\varphi \in \Gamma$. By the definition of R_i^c, we have $\mathcal{K}_i\varphi \in \Delta$.

Suppose $\mathcal{K}_i\varphi \in \Delta$ and $\mathcal{K}_i\varphi \notin \Gamma$. By the property of MCS, we have $\neg\mathcal{K}_i\varphi \in \Gamma$. By axiom 5 we have $\mathcal{K}_i\neg\mathcal{K}_i\varphi \in \Gamma$. Then we have $\neg\mathcal{K}_i\varphi \in \Delta$ by the definition of R_i^c. Contradiction. □

Proposition 17 *R_i^c is an equivalence relation.*

Proof We just need to prove R_i^c is reflexive, transitive, and symmetric.

(1) R_i^c is reflexive: For all $\mathcal{K}_i\psi \in \Gamma$, by axiom T we have $\psi \in \Gamma$. Hence we have $\langle \Gamma, F, \vec{f}\rangle R_i^c \langle \Gamma, F, \vec{f}\rangle$ by the definition of R_i^c.

(2) R_i^c is transitive: Suppose $\langle \Gamma, F, \vec{f}\rangle R_i^c \langle \Delta, G, \vec{g}\rangle$ and $\langle \Delta, G, \vec{g}\rangle R_i^c \langle \Theta, H, \vec{h}\rangle$. Suppose $\mathcal{K}y_i\varphi, \mathcal{K}_i\psi \in \Gamma$. By the definition of R_i^c, we have $f_i = g_i = h_i$. By Proposition 16, we have $\mathcal{K}y_i\varphi, \mathcal{K}_i\psi \in \Delta$. By the definition of R_i^c we get $\mathcal{K}y_i\varphi, \psi \in \Theta$. Therefore $\langle \Gamma, F, \vec{f}\rangle R_i^c \langle \Theta, H, \vec{h}\rangle$.

(3) R_i^c is symmetric: Suppose $\langle \Gamma, F, \vec{f}\rangle R_i^c \langle \Delta, G, \vec{g}\rangle$. Then we have $f_i = g_i$. Suppose $\mathcal{K}y_i\varphi, \mathcal{K}_i\psi \in \Delta$. By Proposition 16, we have $\mathcal{K}y_i\varphi \in \Gamma$ and $\mathcal{K}_i\psi \in \Gamma$. By axiom T, $\psi \in \Gamma$, thus $\langle \Delta, G, \vec{g}\rangle R_i^c \langle \Gamma, F, \vec{f}\rangle$. □

In order to establish that for any $\Gamma \in \Omega$, there exist F and $\vec{f}$ such that $\langle \Gamma, F, \vec{f}\rangle \in W^c$, we define the following construction.

Definition 18 Given any $\Gamma \in \Omega$, construct F^Γ and $\vec{f}^\Gamma$ as follows:

- $F_0^\Gamma = \{\langle \varphi, \varphi\rangle \mid \exists i \in \mathbf{I}, \mathcal{K}y_i\varphi \in \Gamma\} \cup \{\langle e, \varphi\rangle \mid \varphi \in \Lambda\}$
- $F_{n+1}^\Gamma = F_n^\Gamma \cup \{\langle s \cdot t, \psi\rangle \mid \langle s, \varphi \to \psi\rangle \in F_n^\Gamma, \langle t, \varphi\rangle \in F_n^\Gamma \text{ for some } \varphi\}$
- $F^\Gamma = \bigcup_{n\in\mathbb{N}} F_n^\Gamma$.
- $\forall i \in \mathbf{I}, f_i^\Gamma : \{\varphi \mid \mathcal{K}y_i\varphi \in \Gamma\} \to E^c, f_i^\Gamma(\varphi) = \varphi$.

By the construction of $F_n^\Gamma (n \in \mathbb{N})$, $\{F_n^\Gamma \mid n \in \mathbb{N}\}$ is monotonic. i.e., $\forall m, n \in N$, if $m \leqslant n$, then $F_m^\Gamma \subseteq F_n^\Gamma$.

Proposition 19 *For any* $\Gamma \in \Omega$, $\langle \Gamma, F^\Gamma, \vec{f}^\Gamma \rangle \in W^c$.

Proof To prove $\langle \Gamma, F^\Gamma, \vec{f}^\Gamma \rangle \in W^c$, we just need to show that F^Γ satisfies conditions (i)–(iii) in the definition of W^c.

- Suppose $\langle s, \varphi \to \psi \rangle, \langle t, \varphi \rangle \in F^\Gamma$. By monotonicity of $\{F_n^\Gamma \mid n \in \mathbb{N}\}$, there exists $k \in \mathbb{N}$ such that $\langle s, \varphi \to \psi \rangle, \langle t, \varphi \rangle \in F_k^\Gamma$. Thus we have $\langle s \cdot t, \psi \rangle \in F_{k+1}^\Gamma$ by the construction of $F_k^\Gamma (k \in \mathbb{N})$. Hence $\langle s \cdot t, \psi \rangle \in F^\Gamma$, thus F^Γ satisfies condition (i).
- Suppose $\varphi \in \Lambda$. By the construction of F_0^Γ, we have $\langle e, \varphi \rangle \in F_0^\Gamma$. Hence we get $\langle e, \varphi \rangle \in F^\Gamma$. Thus F satisfies condition (ii).
- Suppose $\mathcal{K}y_i \varphi \in \Gamma$. Then we have $\langle \varphi, \varphi \rangle \in F^\Gamma$ by the construction of F_0^Γ and F^Γ. Since $\mathcal{K}y_i \varphi \in \Gamma$, by the construction of f_i^Γ, we have $\varphi \in dom(f_i^\Gamma)$ and $f_i^\Gamma(\varphi) = \varphi$. Thus we have $\langle f_i^\Gamma(\varphi), \varphi \rangle \in F^\Gamma$. Hence, we have that F^Γ and $\vec{f}^\Gamma$ satisfy condition (iii). □

This completes the proof that $\mathcal{M}^c$ is well-defined. Now we can establish the existence lemmas for $\mathcal{K}_i$ and $\mathcal{K}y_i$.

Lemma 20 ($\mathcal{K}_i$ Existence Lemma) *For any* $\langle \Gamma, F, \vec{f} \rangle \in W^c$, *if* $\widehat{\mathcal{K}_i}\varphi \in \Gamma$, *then there exists a* $\langle \Delta, G, \vec{g} \rangle \in W^c$ *such that* $\langle \Gamma, F, \vec{f} \rangle R_i^c \langle \Delta, G, \vec{g} \rangle$ *and* $\varphi \in \Delta$.

Proof Suppose $\widehat{\mathcal{K}_i}\varphi \in \Gamma$. We will construct a $\langle \Delta, G, \vec{g} \rangle$ such that

$$\langle \Gamma, F, \vec{f} \rangle R^c \langle \Delta, G, \vec{g} \rangle \text{ and } \varphi \in \Delta.$$

Let Δ^- be $\{\varphi\} \cup \{\mathcal{K}y_i \psi \mid \mathcal{K}y_i \psi \in \Gamma\} \cup \{\chi \mid \mathcal{K}_i \chi \in \Gamma\}$. Then Δ^- is consistent. Suppose not, then there are $\mathcal{K}y_i \psi_1, \cdots, \mathcal{K}y_i \psi_m, \chi_1, \cdots, \chi_n \in \Delta^-$ such that

$$\vdash_{\mathbb{SKY}} \mathcal{K}y_i \psi_1 \wedge \cdots \wedge \mathcal{K}y_i \psi_m \wedge \chi_1 \wedge \cdots \wedge \chi_n \to \neg\varphi.$$

Then

$$\vdash_{\mathbb{SKY}} \mathcal{K}_i(\mathcal{K}y_i \psi_1 \wedge \cdots \wedge \mathcal{K}y_i \psi_m \wedge \chi_1 \wedge \cdots \wedge \chi_n) \to \mathcal{K}_i \neg\varphi.$$

Since

$$\begin{aligned} \vdash_{\mathbb{SKY}} & (\mathcal{K}_i \mathcal{K}y_i \psi_1 \wedge \cdots \wedge \mathcal{K}_i \mathcal{K}y_i \psi_m \wedge \mathcal{K}_i \chi_1 \wedge \cdots \wedge \mathcal{K}_i \chi_n) \\ & \to \mathcal{K}_i(\mathcal{K}y_i \psi_1 \wedge \cdots \wedge \mathcal{K}y_i \psi_m \wedge \chi_1 \wedge \cdots \wedge \chi_n), \end{aligned}$$

by propositional resoning,

$$\vdash_{\mathbb{SKY}} (\mathcal{K}_i \mathcal{K}y_i \psi_1 \wedge \cdots \wedge \mathcal{K}_i \mathcal{K}y_i \psi_m \wedge \mathcal{K}_i \chi_1 \wedge \cdots \wedge \mathcal{K}_i \chi_n) \to \mathcal{K}_i \neg\varphi.$$

By $\mathcal{K}y_i\psi_j \in \Gamma$ and axiom 4YK, we have $\mathcal{K}_i\mathcal{K}y_i\psi_j \in \Gamma$. Since $\mathcal{K}_i\chi_j \in \Gamma$, it follows that $\mathcal{K}_i\neg\varphi \in \Gamma$, i.e., $\neg\widehat{\mathcal{K}_i}\varphi \in \Gamma$. But this is impossible: Γ is an MCS containing $\widehat{\mathcal{K}_i}\varphi$. We conclude that Δ^- is consistent.

Let Δ be any MCS containing Δ^-, such extensions exist by a Lindenbaum-like argument. It follows that for any $\mathcal{K}y_i\varphi$, $\mathcal{K}y_i\varphi \in \Gamma$ iff $\mathcal{K}y_i\varphi \in \Delta$:

- Suppose $\mathcal{K}y_i\varphi \in \Gamma$. By the construction of Δ, we have $\mathcal{K}y_i\varphi \in \Delta$.
- Suppose $\mathcal{K}y_i\varphi \in \Delta$ and $\mathcal{K}y_i\varphi \notin \Gamma$. By the property of MCS, we have $\neg\mathcal{K}y_i\varphi \in \Gamma$. By Proposition 11, we have $\mathcal{K}_i\neg\mathcal{K}y_i\varphi \in \Gamma$. By the construction of Δ, we have $\neg\mathcal{K}y_i\varphi \in \Delta$. Contradiction.

In the following, we construct G and $\vec{g}$ to form a world $\langle\Delta, G, \vec{g}\rangle$ in W^c. Based on the above result, we can simply let $g_i = f_i$. We just need to construct g_j for $j \neq i$. Formally, let:

- $G_0 = F \cup \{\langle\varphi, \varphi\rangle \mid \mathcal{K}y_j\varphi \in \Delta \text{ for some } j \neq i\}$
- $G_{n+1} = G_n \cup \{\langle s \cdot t, \psi\rangle \mid \langle s, \varphi \to \psi\rangle, \langle t, \varphi\rangle \in G_n\}$
- $G = \bigcup_{n\in\mathbb{N}} G_n$

$$g_j(\varphi) = \begin{cases} f_j(\varphi) & j = i, \\ \varphi & j \neq i. \end{cases}$$

Since $F \subseteq G$ and G is closed under implication, conditions (i) and (ii) are obvious. For condition (iii), if $\mathcal{K}y_i\varphi \in \Delta$, then $\mathcal{K}y_i\varphi \in \Gamma$. Thus

$$\langle g_i(\varphi), \varphi\rangle = \langle f_i(\varphi), \varphi\rangle \in F \subseteq G.$$

Condition (iii) also holds if $\mathcal{K}y_j\varphi \in \Delta$ for $j \neq i$ by definition of G_0. It follows that $\langle\Delta, G, \vec{g}\rangle \in W^c$. By the construction of $\langle\Delta, G, \vec{g}\rangle$, we have $\varphi \in \Delta$, $\Gamma_i^{\#} \subseteq \Delta$, and $f_i = g_i$. Therefore there exists a state $\langle\Delta, G, \vec{g}\rangle \in W^c$ such that $\langle\Gamma, F, \vec{f}\rangle R_i^c \langle\Delta, G, \vec{g}\rangle$ and $\varphi \in \Delta$. □

To refute $\mathcal{K}y_i\psi$ semantically, for each explanation t for ψ at the current world, we need to construct an accessible world where t is not an explanation for ψ. This leads to the following lemma.

Lemma 21 ($\mathcal{K}y_i$ Existence Lemma) *For any $\langle\Gamma, F, \vec{f}\rangle \in W^c$, if $\mathcal{K}y_i\psi \notin \Gamma$ then for any $\langle t, \psi\rangle \in F$, there exists $\langle\Delta, G, \vec{g}\rangle \in W^c$ such that $\langle t, \psi\rangle \notin G$ and $\langle\Gamma, F, \vec{f}\rangle R_i^c \langle\Delta, G, \vec{g}\rangle$.*

Proof Suppose $\mathcal{K}y_i\psi \notin \Gamma$, $\langle\Gamma, F, \vec{f}\rangle \in W^c$, and $\langle t, \psi\rangle \in F$. We construct $\langle\Delta, G, \vec{g}\rangle$ as follows:

- $\Delta = \Gamma$
- $\Psi = \{\langle s, \varphi\rangle \mid \langle s, \varphi\rangle \in F \text{ and } \mathcal{K}y_i\varphi \notin \Gamma\}$
- $\Psi' = \{\langle t \cdot s, \varphi\rangle \mid \langle s, \varphi\rangle \in \Psi\}$
- $G_0 = (F \backslash \Psi) \cup \Psi'$
- $G_{n+1} = G_n \cup \{\langle r \cdot s, \varphi_2\rangle \mid \langle r, \varphi_1 \to \varphi_2\rangle, \langle s, \varphi_1\rangle \in G_n\}$

– $G = \bigcup_{n\in\mathbb{N}} G_n$
– For each $j \in \mathbf{I}$, $g_j : \{\varphi \mid \mathcal{K}y_j\varphi \in \Delta\} \to E^c$ is defined as:

$$g_j(\varphi) = \begin{cases} f_j(\varphi), & \langle f_j(\varphi), \varphi\rangle \notin \Psi \\ t \cdot f_j(\varphi), & \langle f_j(\varphi), \varphi\rangle \in \Psi \end{cases}$$

Throughout this proof we write $|s| > |t|$ to express that t is proper subterm of s. From the construction of G, it is clear that for any $\langle s, \varphi\rangle \in \Psi'$, we have $|s| > |t|$. We can show that for any $\mathcal{K}y_i\varphi \notin \Gamma$, if $\langle s, \varphi\rangle \in G_0$ then $\langle s, \varphi\rangle \in \Psi'$. Towards contradiction, suppose that $\mathcal{K}y_i\varphi \notin \Gamma$ and $\langle s, \varphi\rangle \in F\backslash\Psi$, then $\langle s, \varphi\rangle \in F$, thus $\langle s, \varphi\rangle \in \Psi$ by the definition of Ψ, contradiction. It follows:

$$\text{For any } \mathcal{K}y_i\varphi \notin \Gamma, \text{ if } \langle s, \varphi\rangle \in G_0 \text{ for some } s, \text{ then } |s| > |t|. \tag{1}$$

Thus in particular

$$\langle t, \psi\rangle \text{ is not in } G_0. \tag{2}$$

The idea behind the construction of G is to first replace any current explanation for ψ with something longer, and then take the closure w.r.t. implication. Note that for technical convenience, we treat all φ such that $\mathcal{K}y_i\varphi \notin \Gamma$ in the basic step together.

Now we prove the following claims.

Claim 1 $\langle \Delta, G, \vec{g}\rangle \in W^c$. i.e., G satisfies the conditions in the definition of W^c.

(i) Suppose $\langle r, \varphi_1 \to \varphi_2\rangle, \langle s, \varphi_1\rangle \in G$. By the construction of G, there exists $n \in \mathbb{N}$ such that $\langle r, \varphi_1 \to \varphi_2\rangle, \langle s, \varphi_1\rangle \in G_n$. By the construction of G_{n+1}, we have $\langle r \cdot s, \varphi_2\rangle \in G_{n+1}$. Thus $\langle r \cdot s, \varphi_2\rangle \in G$.
(ii) Suppose $\varphi \in \Lambda$. Then $\langle e, \varphi\rangle \in F$. Since φ is a tautology, by NECKY and the property of MCS, we have $\mathcal{K}y_i\varphi \in \Gamma$. Thus $\langle e, \varphi\rangle \notin \Psi$. Thus $\langle e, \varphi\rangle \in G_0$. Hence $\langle e, \varphi\rangle \in G$.
(iii) Suppose $\mathcal{K}y_j\varphi \in \Delta$ ($j \in \mathbf{I}$). By $\Delta = \Gamma$, we get $\mathcal{K}y_j\varphi \in \Gamma$. Thus $\langle f_j(\varphi), \varphi\rangle \in F$. We have two cases:

– $\langle f_j(\varphi), \varphi\rangle \notin \Psi$: Thus $g_j(\varphi) = f_j(\varphi)$. Thus we have $\langle g_j(\varphi), \varphi\rangle \in F$ and $\langle g_j(\varphi), \varphi\rangle \notin \Psi$. Thus $\langle g_j(\varphi), \varphi\rangle \in G_0$. Hence $\langle g_j(\varphi), \varphi\rangle \in G$.
– $\langle f_j(\varphi), \varphi\rangle \in \Psi$: Thus $g_j(\varphi) = t \cdot f_j(\varphi)$ and $\langle g_j(\varphi), \varphi\rangle \in \Psi'$. Thus we have $\langle g_j(\varphi), \varphi\rangle \in G_0$. Hence $\langle g_j(\varphi), \varphi\rangle \in G$.

Claim 2 $\langle \Gamma, F, \vec{f}\rangle R_i^c \langle \Delta, G, \vec{g}\rangle$

To prove this claim, we just need to check two conditions:

(i) Since $\Delta = \Gamma$, obviously, we have $\Gamma_i^{\#} \subseteq \Delta$.
(ii) Since $\Delta = \Gamma$, we have $\{\varphi \mid \mathcal{K}y_i\varphi \in \Gamma\} = \{\varphi \mid \mathcal{K}y_i\varphi \in \Delta\}$, i.e., dom($g_i$)=dom($f_i$). For any $\varphi \in \{\varphi \mid \mathcal{K}y_i\varphi \in \Delta\}$, since $\langle f_i(\varphi), \varphi\rangle \notin \Psi$, by the definition of g_i, we have $g_i(\varphi) = f_i(\varphi)$. Hence $g_i = f_i$.

To prove $\langle t, \psi\rangle \notin G$, we first prove the following useful claim:

Claim 3 If $\mathcal{K}y_i\varphi \notin \Gamma$ and $\langle s, \varphi\rangle \in G_{n+1}\backslash G_n$, then $|s| > |t|$.

Suppose $\mathcal{K}y_i\varphi \notin \Gamma$. Do induction on n:

- $n = 0$. Suppose $\langle s, \varphi\rangle \in G_1\backslash G_0$. Then there exists s_1, s_2, and χ such that $s = s_1 \cdot s_2$, $\langle s_1, \chi \to \varphi\rangle, \langle s_2, \chi\rangle \in G_0$. We have two cases
  - $\langle s_1, \chi \to \varphi\rangle \in \Psi'$ or $\langle s_2, \chi\rangle \in \Psi'$: Thus $|s_1| > |t|$ or $|s_2| > |t|$. Thus $|s| > |t|$.
  - $\langle s_1, \chi \to \varphi\rangle \notin \Psi'$ and $\langle s_2, \chi\rangle \notin \Psi'$: Since $\langle s_1, \chi \to \varphi\rangle, \langle s_2, \chi\rangle \in G_0$, thus $\langle s_1, \chi \to \varphi\rangle, \langle s_2, \chi\rangle \in F\backslash\Psi$. Thus $\mathcal{K}y_i(\chi \to \varphi), \mathcal{K}y_i\chi \in \Gamma$. Thus $\mathcal{K}y_i\varphi \in \Gamma$ by axiom DISTY. Contradiction.
- $n > 0$. Suppose $\langle s, \varphi\rangle \in G_{n+1}\backslash G_n$. Then there exist s_1, s_2, χ such that $s = s_1 \cdot s_2$ and $\langle s_1, \chi \to \varphi\rangle, \langle s_2, \chi\rangle \in G_n$. Moreover, we find that

$$\mathcal{K}y_i(\chi \to \varphi) \notin \Gamma \text{ or } \mathcal{K}y_i\chi \notin \Gamma$$

since $\mathcal{K}y_i\varphi \notin \Gamma$ and Γ is an MCS. We also have

$$\langle s_1, \chi \to \varphi\rangle \notin G_{n-1} \text{ or } \langle s_2, \chi\rangle \notin G_{n-1}$$

since otherwise $\langle s, \varphi\rangle \in G_n$ by the definition of G_n. Then we have

$$\langle s_1, \chi \to \varphi\rangle \in G_n\backslash G_{n-1} \text{ or } \langle s_2, \chi\rangle \in G_n\backslash G_{n-1}.$$

We have the following cases:

- $\mathcal{K}y_i(\chi \to \varphi) \notin \Gamma$ and $\langle s_1, \chi \to \varphi\rangle \in G_n\backslash G_{n-1}$. By IH, we have $|s_1| > |t|$. Hence $|s| > |t|$.
- $\mathcal{K}y_i\chi \notin \Gamma$ and $\langle s_2, \chi\rangle \in G_n\backslash G_{n-1}$. By IH, we have $|s_2| > |t|$. Hence $|s| > |t|$.
- $\mathcal{K}y_i(\chi \to \varphi) \notin \Gamma$ and $\langle s_1, \chi \to \varphi\rangle \in G_{n-1}$. If $\langle s_1, \chi \to \varphi\rangle \in G_0$, then we have $|s_1| > |t|$ by (1); If $\langle s_1, \chi \to \varphi\rangle \notin G_0$, then there exists $0 < k < n$ such that $\langle s_1, \chi \to \varphi\rangle \in G_k\backslash G_{k-1}$. Thus by IH we have $|s_1| > |t|$. Thus $|s| > |t|$.
- $\mathcal{K}y_i\chi \notin \Gamma$ and $\langle s_2, \chi\rangle \in G_{n-1}$. Similar to the above.

Claim 4 $\langle t, \psi\rangle \notin G$.

According to the construction of G, we just need to show that for all $n \in \mathbb{N}$, $\langle t, \psi\rangle \notin G_n$. By (2), we already know $\langle t, \psi\rangle \notin G_0$. Based on Claim 3, $\langle t, \psi\rangle$ cannot be added in any G_n for $n \geq 1$. We conclude $\langle t, \psi\rangle \notin G$. □

Finally we are ready to prove the truth lemma.

Lemma 22 (Truth Lemma) *For all* φ, $\langle \Gamma, F, \vec{f}\rangle \vDash \varphi$ *if and only if* $\varphi \in \Gamma$.

Proof This is established by standard induction on the complexity of φ. The atomic cases and the boolean cases are standard. The case when $\varphi = \mathcal{K}_i\psi$ is also routine based on Lemma 20.

Consider the case that φ is $\mathcal{K}y_i\psi$ for some ψ.

$\Leftarrow$ If $\mathcal{K}y_i\psi \in \Gamma$, then for any $\langle \Delta, G, \vec{g}\rangle$ such that $\langle \Gamma, F, \vec{f}\rangle R_i^c \langle \Delta, G, \vec{g}\rangle$, we have then $\mathcal{K}y_i\psi \in \Delta$ by the definition of R_i^c. Since $\vdash_{\mathbb{SKY}} \mathcal{K}y_i\psi \to \psi$ (by T and IMP), we have $\psi \in \Delta$. By IH, we have $\langle \Delta, G, \vec{g}\rangle \vDash \psi$. Since $\mathcal{K}y_i\psi \in \Gamma$ and $\mathcal{K}y_i\psi \in \Delta$, we have $\langle f_i(\psi), \psi\rangle \in F$ and $\langle g_i(\psi), \psi\rangle \in G$. By the definition of R_i^c, we have $f_i = g_i$. Thus there exists $g_i(\psi) = f_i(\psi) \in E^c$ such that $\langle \Delta, G, \vec{g}\rangle \in \mathcal{E}^c(g_i(\psi), \psi)$. Therefore we conclude $\langle \Gamma, F, \vec{f}\rangle \vDash \mathcal{K}y_i\psi$.

$\Rightarrow$ Suppose $\mathcal{K}y_i\psi \notin \Gamma$. We have two cases as follows:

- $\mathcal{K}_i\psi \notin \Gamma$: then by Lemma 20 and the semantics, $\langle \Gamma, F, \vec{f}\rangle \nvDash \mathcal{K}y_i\psi$.
- $\mathcal{K}\psi \in \Gamma$: We also have two cases:
  - $\langle t, \psi\rangle \notin F$ for all $t \in E$. By the semantics, $\langle \Gamma, F, \vec{f}\rangle \nvDash \mathcal{K}y_i\psi$.
  - There exists $t \in E$ such that $\langle t, \psi\rangle \in F$. By Lemma 21, there exists $\langle \Delta, G, \vec{g}\rangle \in W^c$ with $\langle t, \psi\rangle \notin G$ and $\langle \Gamma, F, \vec{f}\rangle R_i^c \langle \Delta, G, \vec{g}\rangle$. Hence we have $\langle \Gamma, F, \vec{f}\rangle \nvDash \mathcal{K}y_i\psi$.

□

Theorem 23 (Completeness of $\mathbb{SKY}$ over $\mathbb{C}$) *$\Sigma \vDash_{\mathbb{C}} \varphi$ implies $\Sigma \vdash_{\mathbb{SKY}} \varphi$.*

Proof Suppose $\Sigma \vDash_{\mathbb{C}} \varphi$. Towards a contradiction, suppose $\Sigma \nvdash_{\mathbb{SKY}} \varphi$. Then $\Sigma \cup \{\neg\varphi\}$ is consistent. Extend $\Sigma \cup \{\neg\varphi\}$ to a maximal consistent set Γ. By Proposition 19, there exist F and $\vec{f}$ such that $\langle \Gamma, F, \vec{f}\rangle \in W^c$. By Lemma 22, we have $\langle \Gamma, F, \vec{f}\rangle \vDash \Sigma \cup \{\neg\varphi\}$, thus $\Sigma \cup \{\neg\varphi\}$ is satisfiable, thus $\Sigma \vDash_{\mathbb{C}} \varphi$ is false. Contradiction. □

By Theorems 8 and 23, we have the following corollary.

Corollary 24 (Completeness of $\mathbb{SKY}$ over $\mathbb{C}_F$) *$\Sigma \vDash_{\mathbb{C}_F} \varphi$ implies $\Sigma \vdash_{\mathbb{SKY}} \varphi$.*

Now let us look at the completeness of $\mathbb{SKYI}$. The crucial observation is that we can use the same canonical model definition except now we let Ω be the set of all maximal $\mathbb{SKYI}$-consistent set of **ELKy** formulas. The similar propositions follow due to Proposition 12. The only extra thing is to check whether the new canonical model has the introspection property.

Proposition 25 *$\mathcal{M}^c$ has introspection property.*

Proof Suppose $\langle \Gamma, F, \vec{f}\rangle \vDash \varphi$ and φ has the form of $\mathcal{K}_i\psi$ or $\neg\mathcal{K}_i\psi$ or $\mathcal{K}y_i\psi$ or $\neg\mathcal{K}y_i\psi$. By Lemma 22, we have $\varphi \in \Gamma$. By the axioms 4KY-5Y and the properties of MCS, we have $\mathcal{K}y_i\varphi \in \Gamma$. By Lemma 22, we have $\langle \Gamma, F, \vec{f}\rangle \vDash \mathcal{K}y_i\varphi$. Thus $\exists r \in E^c$, $\langle \Delta, G, \vec{g}\rangle \in \mathcal{E}^c(r, \varphi)$ for each $\langle \Delta, G, \vec{g}\rangle$ such that $\langle \Gamma, F, \vec{f}\rangle R_i^c \langle \Delta, G, \vec{g}\rangle$. □

Based on the above proposition and Theorem 10 we have:

Theorem 26 (Completeness of $\mathbb{SKYI}$ over $\mathbb{C}_I$ and $\mathbb{C}_{FI}$)
If $\Sigma \vDash_{\mathbb{C}_I} \varphi$, then $\Sigma \vdash_{\mathbb{SKYI}} \varphi$. If $\Sigma \vDash_{\mathbb{C}_{FI}} \varphi$, then $\Sigma \vdash_{\mathbb{SKYI}} \varphi$.

4 Comparison with justification logic

In this section, we compare our framework with justification logic. We first explain our deviations from the standard justification logic, and then give an alternative semantics of our logic $\mathbb{SKY}$, which is technically closer to the standard setting of justification logic.

4.1 Similarities and differences

The language of the most classic justification logic **LP** [i.e., **JT4** by Artemov (2008)] includes both formulas φ and justification terms t:

$$\begin{aligned} \varphi &::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid t{:}\varphi \\ t &::= x \mid c \mid (t \cdot t) \mid (t + t) \mid !t \end{aligned}$$

The possible-world semantics of justification logic is based on the Fitting model $\langle S, R, \mathcal{E}, V\rangle$ where $\langle S, R, V\rangle$ is a single-agent Kripke model and $\mathcal{E}$ is an *evidence function* assigning justification terms t to formulas on each world, just as in our setting. The formula $t{:}\varphi$ has the following semantics [cf. e.g., Fitting (2016)]:

$$\mathcal{M}, w \Vdash t{:}\varphi \iff \begin{array}{l} \text{(a) } w \in \mathcal{E}(t, \varphi); \\ \text{(b) } v \Vdash \varphi \text{ for all } v \text{ such that } wRv. \end{array}$$

Compared to our semantics for $\mathcal{K}y_i\varphi$, note that (a) only requires that t is a justification of φ on the current world w. The Fitting models for **LP** are assumed to have further conditions:[14]

(1) $\mathcal{E}(s, \varphi \to \psi) \cap \mathcal{E}(t, \varphi) \subseteq \mathcal{E}(s \cdot t, \psi)$
(2) $\mathcal{E}(t, \varphi) \cup \mathcal{E}(s, \varphi) \subseteq \mathcal{E}(s + t, \varphi)$
(3) $\mathcal{E}(t, \varphi) \subseteq \mathcal{E}(!t, t{:}\varphi)$
(4) Monotonicity: $w \in \mathcal{E}(t, \varphi)$ and wRv implies $v \in \mathcal{E}(t, \varphi)$.
(5) R is reflexive and transitive.

Note that we also require (1) and (5) above and include $\cdot$ as an operation on explanations in E semantically. On the other hand, we leave out (2) (3) (4) and the operations $+$ and $!$ for specific considerations in our setting. For the case of $+$, consider the following model where φ has two possible explanations and agent i cannot distinguish them (thus $\neg\mathcal{K}y_i\varphi$ holds).

$$t{:}\varphi \text{ —}i\text{— } s{:}\varphi$$

If we impose condition (2) then $s + t$ is a uniform explanation of φ on both worlds, which makes $\mathcal{K}y_i\varphi$ true. More generally, for any finite model where φ has some explanations on each world, $\mathcal{K}y_i\varphi$ will always be true under condition (2), which is counterintuitive in our setting. Conceptually, the explanation should be *precise*, you

[14] The "S5 version" of justification logic **JT45** also adds another condition about negative introspection: $\overline{\mathcal{E}(t, \varphi)} \subseteq \mathcal{E}(?t, \neg(t{:}\varphi))$, and requires strong evidence, where ? is a new operation for justification terms in the language, cf. Artemov (2008). To simplify the discussion, we focus on **LP** here.

cannot explain a theorem by saying one of all the possible proofs up to a certain length works. Knowing there is a proof does not mean you know why the theorem holds.

Operation ! and conditions (3) and (4) are relevant to the validity of the axiom $t{:}\varphi \rightarrow !t{:}(t{:}\varphi)$ in the justification logic **LP**, which is used to realize axiom 4 in modal logic. Intuitively, ! is the proof checker and $!t$ will always be a justification of $t{:}\varphi$.[15] Although we do not have $t{:}\varphi$ in the language, it may sound reasonable to include ! and require $\mathcal{E}(t,\varphi) \subseteq \mathcal{E}(!t, \mathcal{K}y_i\varphi)$. However, t being an explanation for φ does not entail that t can be transformed uniformly into an explanation for $\mathcal{K}y_i\varphi$. For example, the window is broken since someone threw a rock at it, but there can be different explanations for an agent to know why the window is broken: she saw it, or someone told her about it, and so on.

The technically motivated condition (4) in justification logic requires that any accessible possible world has more explanations than the actual world, which is not reasonable in our setting: an undesired consequence of condition (4) would be $w \in \mathcal{E}(t,\varphi)$ and $w \vDash \mathcal{K}_i\varphi$ imply $w \vDash \mathcal{K}y_i\varphi$.

There are justification logics available with both $\Box\varphi$ and $t{:}\varphi$, see, e.g., Artemov and Nogina (2005) and Kuznets and Studer (2012). Justification terms are used to represent explicit knowledge whereas the $\Box$-operator is used for implicit knowledge. Hence these logics feature the principle

$$t{:}\varphi \rightarrow \Box\varphi, \tag{3}$$

which is based on the idea that one may implicitly know more than what is explicitly justified. Note that in the presence of the !-operation, the principle (3) implies

$$t{:}\varphi \rightarrow \Box t{:}\varphi. \tag{4}$$

Indeed, by (3) we have $!t{:}(t{:}\varphi) \rightarrow \Box(t{:}\varphi)$, which together with the axiom $t{:}\varphi \rightarrow !t{:}(t{:}\varphi)$ yields (4). The formulas (3) and (4) correspond in our setting to the axioms $\mathcal{K}y_i\varphi \rightarrow \mathcal{K}_i\varphi$ and $\mathcal{K}y_i\varphi \rightarrow \mathcal{K}_i\mathcal{K}y_i\varphi$, respectively.

In some versions of multi-agent justification logic, e.g. Bucheli et al. (2011) and Renne (2012), the evidence function $\mathcal{E}$ is agent-dependent (or, equivalently, each agent has her own justifications), and correspondingly the formula $t{:}_i\,\varphi$ is introduced into the language to express that t is a justification of φ for i. However, in our models, we use a uniform function $\mathcal{E}$ for all agents since we think the explanatory relation between explanations and formulas is also part of the possible worlds, just like basic propositional facts, which are interpreted by a valuation function independent from the agents.

Justification logics are parameterized by a constant specification (CS), a collection of $c{:}\varphi$ formulas where c is a justification constant and φ is an axiom of the justification logic. It controls which axioms the logic provides justifications for, i.e. which axioms an agent may use in her justified reasoning process. A justification logic model meets the requirement of a given CS if $W = \mathcal{E}(c,\varphi)$ for all $c{:}\varphi \in$ CS. In contrast, we include

[15] In the multi-agent setting, $!_i$ was introduced to capture the proof check done by each agent (Yavorskaya 2006).

only tautologies (but not all axioms) in our tautology ground Λ. For example, if we had $(\mathcal{K}_i\varphi \to \varphi) \in \Lambda$, then we could derive $\mathcal{K}y_i(\mathcal{K}_i\varphi \to \varphi)$ by NECKY, which would imply $\mathcal{K}y_i\mathcal{K}_i\varphi \to \mathcal{K}y_i\varphi$ by DISTY. That, however, would be a strange consequence: e.g., I know why I know that the window is broken implies that I know why it is broken.

The table below highlights the similarities between our axioms (or derivable theorems in $\mathbb{SKY}$ and $\mathbb{SKYI}$) and axioms in (variants of) justification logic when viewing $t{:}\varphi$ as $\mathcal{K}y_i\varphi$:

Justification Logic	Our work
$t{:}(\varphi \to \psi) \to s{:}\varphi \to (t \cdot s){:}\psi$	$\mathcal{K}y_i(\varphi \to \psi) \to (\mathcal{K}y_i\varphi \to \mathcal{K}y_i\psi)$
$t{:}\varphi \to (s + t){:}\varphi$	$\mathcal{K}y_i\varphi \to \mathcal{K}y_i\varphi$
$t{:}\varphi \to \varphi$	$\mathcal{K}y_i\varphi \to \varphi$
$t{:}\varphi \to !t{:}(t{:}\varphi)$	$\mathcal{K}y_i\varphi \to \mathcal{K}y_i\mathcal{K}y_i\varphi$
$\neg t{:}\varphi \to ?t{:}(\neg t{:}\varphi)$	$\neg\mathcal{K}y_i\varphi \to \mathcal{K}y_i\neg\mathcal{K}y_i\varphi$
$t{:}\varphi \to \Box\varphi$ (Artemov and Nogina 2005)	$\mathcal{K}y_i\varphi \to \mathcal{K}_i\varphi$
$t{:}\varphi \to \Box t{:}\varphi$ (Artemov and Nogina 2005)	$\mathcal{K}y_i\varphi \to \mathcal{K}_i\mathcal{K}y_i\varphi$
$\neg t{:}\varphi \to \Box\neg t{:}\varphi$ (Artemov and Nogina 2005)	$\neg\mathcal{K}y_i\varphi \to \mathcal{K}_i\neg\mathcal{K}y_i\varphi$

To close this comparison, note that Fitting proposed a quantified justification logic by Fitting (2008), and discussed briefly in the end what can be expressed if the language also includes the normal knowledge operator. Since $\mathcal{K}y_i$ implicitly includes quantification over explanations, our language can then be viewed as a fragment of this quantified justification logic extended with $\mathcal{K}$.

4.2 An alternative semantics

The similarities between our work and justification logic make it technically possible to give a more standard justification logic semantics to **ELKy**-formulas. In the following we evaluate formulas over multi-agent Fitting models, see, e.g., Bucheli et al. (2011) and Renne (2012), where each agent has her own accessibility relation and evidence function.[16] We call these models **JL**-models. The interpretation of *agent i knows why* φ is given as *agent i knows that* φ *and has some justification of* φ, that is *knowing why* translates to *having a justification*.

We consider our results about **JL**-models to be technical observations that show why we developed **ELKy**-models to give semantics to the logic of knowing why instead of simply using standard justification logic models. **JL**-models are in fact rather weak for our purpose. This can be seen from the fact that a lot of information is lost in the translation from **ELKy**-models to **JL**-models, in particular **JL**-models only store known explanations but all other possible explanations are dropped. Hence **JL**-models cannot really talk about the difference between $\exists x\mathcal{K}E(x,\varphi)$ and $\mathcal{K}\exists xE(x,\varphi)$, which is essential for our analysis of *knowing why*. Moreover, this **JL**-like semantics cannot handle conditional knowledge-why, as will be introduced formally in the last section

[16] The alternative semantics does not work if we just have only one evidence function.

of this paper. For example, it is reasonable to have a situation where I don't know why φ right now, but I know why φ *given* the information ψ, since the extra information of ψ may rule out some possibilities such that there is a uniform explanation on the remaining possibilities. Due to Remark 2, this is not possible in a monotonic S5 (or S4) **JL**-model.

Definition 27 (**JL**-Model) A **JL**-model $\mathcal{M}$ is a tuple

$$(W, E, \{R_i \mid i \in \mathbf{I}\}, \{\mathcal{E}_i \mid i \in \mathbf{I}\}, V)$$

where:

- W is a non-empty set of possible worlds.
- E is a non-empty set of explanations satisfying the following conditions
  - (a) If $s, t \in E$, then a new explanation $(s \cdot t) \in E$;
  - (b) A special symbol e is in E.
- $R_i \subseteq W \times W$ is an equivalence relation over W.
- $\mathcal{E}_i : E \times \mathbf{ELKy} \to 2^W$ is an *admissible evidence function* satisfying the following conditions:
  - (I) $\mathcal{E}_i(s, \varphi \to \psi) \cap \mathcal{E}_i(t, \varphi) \subseteq \mathcal{E}_i(s \cdot t, \psi)$.
  - (II) If $\varphi \in \Lambda$, then $\mathcal{E}_i(e, \varphi) = W$.
  - (III) Monotonicity: $w \in \mathcal{E}_i(t, \varphi)$ and wR_iv implies $v \in \mathcal{E}_i(t, \varphi)$.
- $V : \mathbf{P} \to 2^W$ is a valuation function.

Remark 2 Note that by imposing monotonicity on S5 models, all i-indistinguishable worlds have the same justifications for the same formula, i.e., if wR_iv then

$$w \in \mathcal{E}_i(t, \varphi) \text{ iff } v \in \mathcal{E}_i(t, \varphi).$$

Definition 28 (*Semantics*)
The satisfaction relation of **ELKy**-formulas on pointed **JL**-models is as below:

$\mathcal{M}, w \models^{\mathsf{J}} p$	$\iff w \in V(p)$
$\mathcal{M}, w \models^{\mathsf{J}} \neg\varphi$	$\iff \mathcal{M}, w \not\models^{\mathsf{J}} \varphi$
$\mathcal{M}, w \models^{\mathsf{J}} \varphi \wedge \psi$	$\iff \mathcal{M}, w \models^{\mathsf{J}} \varphi$ and $\mathcal{M}, w \models^{\mathsf{J}} \psi$
$\mathcal{M}, w \models^{\mathsf{J}} \mathcal{K}_i\varphi$	$\iff \mathcal{M}, v \models^{\mathsf{J}} \varphi$ for each v such that wR_iv.
$\mathcal{M}, w \models^{\mathsf{J}} \mathcal{K}y_i\varphi$	$\iff$ (1) $\exists t \in E$ such that $w \in \mathcal{E}_i(t, \varphi)$;
	(2) $\forall v \in W$, wR_iv implies $\mathcal{M}, v \models^{\mathsf{J}} \varphi$.

Compared to our semantics, the crucial difference in the above semantics of $\mathcal{K}y_i\varphi$ is that it only requires that t is a justification on the *current* world w.

Theorem 29 (Soundness) $\mathbb{SKY}$ *is sound for* **JL**-*models.*

Proof Since **JL**-models are based on S5 Kripke models, the standard axioms of system $\mathbb{S}5$ are all valid. So we just need to check the rest.

DISTY: $\mathcal{K}y_i(\varphi \to \psi) \to (\mathcal{K}y_i\varphi \to \mathcal{K}y_i\psi)$

Suppose $w \Vdash^{\mathsf{J}} \mathcal{K}y_i(\varphi \to \psi)$ and $w \Vdash^{\mathsf{J}} \mathcal{K}y_i\varphi$. By soundness of DISTK, we obtain $\forall v \in W$, wR_iv implies $\mathcal{M}, v \Vdash^{\mathsf{J}} \psi$. Further, by the definition of $\Vdash^{\mathsf{J}}$, there exist $s, t \in E$ with $w \in \mathcal{E}_i(s, \varphi \to \psi)$ and $w \in \mathcal{E}_i(t, \varphi)$. By the closure conditions on admissible evidence functions we get $w \in \mathcal{E}_i(s \cdot t, \psi)$. Hence $w \Vdash^{\mathsf{J}} \mathcal{K}y_i\psi$.

IMP: $\mathcal{K}y_i\varphi \to \mathcal{K}_i\varphi$

Follows immediately from the definition of $\Vdash^{\mathsf{J}}$.

4YK: $\mathcal{K}y_i\varphi \to \mathcal{K}_i\mathcal{K}y_i\varphi$

Suppose $w \Vdash^{\mathsf{J}} \mathcal{K}y_i\varphi$ and let $v \in W$ be arbitrary with wR_iv. By transitivity of R_i we find that $\forall u \in W$, vR_iu implies $\mathcal{M}, u \Vdash^{\mathsf{J}} \varphi$. Further there exists t with $w \in \mathcal{E}_i(t, \varphi)$ and monotonicity of $\mathcal{E}_i$ yields $v \in \mathcal{E}_i(t, \varphi)$. We obtain $v \Vdash^{\mathsf{J}} \mathcal{K}y_i\varphi$ and conclude $w \Vdash^{\mathsf{J}} \mathcal{K}_i\mathcal{K}y_i\varphi$.

NECKY Suppose $\varphi \in \Lambda$. By condition (II) on $\mathcal{E}_i$, we get $w \in \mathcal{E}_i(e, \varphi)$ for any w. Since Λ is a set of tautologies, we also have that wR_iv implies $v \Vdash^{\mathsf{J}} \varphi$. Hence $\mathcal{K}y_i\varphi$ is valid. □

To establish completeness of $\mathbb{SKY}$ with respect to **JL**-models, we show how to transform a given **ELKy**-model into an equivalent **JL**-model. Then completeness for **JL**-models is a consequence of completeness for **ELKy**-models.

Definition 30 (*Corresponding* **JL**-model) Given an **ELKy**-model

$$\mathcal{M} = (W, E, \{R_i \mid i \in \mathbf{I}\}, \mathcal{E}, V)$$

we define the corresponding **JL**-model $\mathcal{M}^{\mathsf{J}}$ as

$$(W, E, \{R_i \mid i \in \mathbf{I}\}, \{\mathcal{E}_i^{\mathsf{J}} \mid i \in \mathbf{I}\}, V)$$

where

$$\mathcal{E}_i^{\mathsf{J}}(t, \varphi) := \{w \mid \forall v \in W,\ wR_iv \text{ implies } v \in \mathcal{E}(t, \varphi)\}.$$

The above definition indeed yields a **JL**-model. Moreover, any given **ELKy**-model $\mathcal{M}$ and its corresponding **JL**-model $\mathcal{M}^{\mathsf{J}}$ satisfy the same formulas. We have the following two lemmas.

Lemma 31 *Let $\mathcal{M}$ be an* **ELKy***-model, then $\mathcal{M}^{\mathsf{J}}$ is a* **JL***-model.*

Proof We have to verify the conditions on $\mathcal{E}^{\mathsf{J}}$:

(I) Suppose $w \in \mathcal{E}_i^{\mathsf{J}}(s, \varphi \to \psi) \cap \mathcal{E}_i^{\mathsf{J}}(t, \varphi)$. Thus for each $v \in W$, wR_iv implies

$$v \in \mathcal{E}(s, \varphi \to \psi) \cap \mathcal{E}(t, \varphi)$$

and hence also $v \in \mathcal{E}(s \cdot t, \psi)$. By the definition of $\mathcal{E}_i^{\mathsf{J}}$ we conclude $w \in \mathcal{E}_i^{\mathsf{J}}(s \cdot t, \psi)$.

(II) If $\varphi \in \Lambda$, then $\mathcal{E}(e, \varphi) = W$ and hence also $\mathcal{E}_i^{\mathsf{J}}(e, \varphi) = W$.
(III) Assume $w \in \mathcal{E}_i^{\mathsf{J}}(t, \varphi)$ and wR_iv. Let u be arbitrary with vR_iu. Since R_i is transitive, we get wR_iu. Thus by the definition of $\mathcal{E}_i^{\mathsf{J}}$ we find $u \in \mathcal{E}(t, \varphi)$. We conclude $v \in \mathcal{E}_i^{\mathsf{J}}(t, \varphi)$. □

Lemma 32 *Let $\mathcal{M} = (W, E, \{R_i \mid i \in \mathbf{I}\}, \mathcal{E}, V)$ be an* **ELKy***-model. For each $w \in W$ and each formula φ,*

$$\mathcal{M}, w \vDash \varphi \text{ if and only if } \mathcal{M}^{\mathsf{J}}, w \vDash^{\mathsf{J}} \varphi.$$

Proof By induction on φ.
Case φ is of the form $\mathcal{K}y_i\psi$. Observe that by the definition of $\mathcal{E}_i^{\mathsf{J}}$ we have that

$$\exists t \in E, \text{ for each } v \text{ such that } wR_iv,\ v \in \mathcal{E}(t, \psi)$$

if and only if

$$\exists t \in E \text{ with } w \in \mathcal{E}_i^{\mathsf{J}}(t, \psi).$$

Hence $\mathcal{M}, w \vDash \mathcal{K}y_i\psi$ if and only if $\mathcal{M}^{\mathsf{J}}, w \vDash^{\mathsf{J}} \mathcal{K}y_i\psi$.
All other cases are trivial. □

Remark 3 The above result also holds if we consider S4-based **ELKy**-models and **JL**-models, i.e., when the relations R_i are only reflexive and transitive.

Corollary 33 (Completeness) $\mathbb{SKY}$ *is strongly complete for* **JL***-models.*

Proof Suppose $\Gamma \nvdash \varphi$. By completeness with respect to **ELKy**-models, there is an **ELKy**-model $\mathcal{M}$ with a world w such that $\mathcal{M}, w \vDash \Gamma$ but $\mathcal{M}, w \nvDash \varphi$. By the previous two lemmas, we find a **JL**-model $\mathcal{M}^{\mathsf{J}}$ such that $\mathcal{M}^{\mathsf{J}}, w \vDash^{\mathsf{J}} \Gamma$ but $\mathcal{M}^{\mathsf{J}}, w \nvDash^{\mathsf{J}} \varphi$. □

The results in this section show that $\mathbb{SKY}$ does have a standard justification logic semantics although its language does not include explicit justifications. In this sense, the logic of knowing why sits in between modal logic and justification logic. Hence it is natural to ask whether there is a realization theorem for $\mathbb{SKY}$ and which system of justification logic $\mathbb{SKY}$ corresponds to (Artemov 2001; Fitting 2016; Kuznets and Studer 2019). For now, we have to leave this question open.

5 Conclusions and future work

In this paper, we present an attempt to formalize the logic of knowing why. In the language we have both the standard knowing that operator $\mathcal{K}_i$ and the new knowing why operator $\mathcal{K}y_i$. A semantics based on Fitting-like models for justification logic is given, which interprets knowing why φ as *there exists an explanation such that I know it is one explanation for* φ. We gave two proof systems, one weaker and one stronger depending on the choice of introspection axioms, and showed their completeness over various model classes.

Note that, in the logic of knowing value (Wang and Fan 2013, 2014), there is one and only one value for each constant. However, there can be different explanations for the same fact in our setting. This difference also leads to some technical complications in the completeness proof: to negate $\mathcal{K}y_i\varphi$, it is not enough to just construct another world. Instead, for *each* explanation t of φ at the current world, we need to construct one world to refute it.

As the title shows, it is by no means *the* logic of knowing why. Besides the introspection axioms, there are a lot to be discussed.[17] For example, although DISTY looks reasonable in a setting focusing on deductive explanations, it may cause troubles if causal explanations or other types of explanations are considered. Recall our example about the flagpole and its shadow. It is reasonable to assume that I know why the shadow is y meters long ($\mathcal{K}y_i\, p$), and I also know why that the shadow is y meters implies the pole is x meters long ($\mathcal{K}y_i(p \to q)$). However, it does not entail that I know why the pole is x meters long ($\mathcal{K}y_i q$) if we are looking for causal explanation (or functional explanation). One way to go around is to replace the material implication by some other relevant (causal) conditional, then $\mathcal{K}y_i(p \to q)$ may not hold in this setting anymore.

It seems that we often do not have clear semantic intuition about non-trivial expressions of knowing why. One reason is that there may be different readings of the same statement of knowing why φ regarding different aspects of φ and different types of desired explanations. For example, "I know why Frank went to Beijing on Monday" may have different meanings depending on the *contrast* the speaker wants to emphasize (van Fraassen 1980):

- I know why *Frank*, not Mary, went to Beijing on Monday.
- I know why Frank went to *Beijing*, not Shanghai, on Monday.
- I know why Frank went to Beijing on *Monday*, not on Tuesday.

Following Koura (1988), we may partially handle this by adding contrast formulas, e.g., turn $\mathcal{K}y_i\varphi$ into $\mathcal{K}y_i(\varphi \wedge \neg\psi \wedge \neg\chi \wedge \dots)$ depending on the emphasis. However, we cannot handle the changes of types of explanations depending on the contrast.

Another future direction is to study the inner structure of explanations further. The early work by Hintikka and Halonen (1995) may turn out to be helpful, where explanations can be of the form of universally quantified formulas, which connects better with the existing theories of scientific explanations in philosophy of science.[18] Moreover, we may be interested in saying whether an explanation is *true*, and knowing why an explanation holds. An extended language with explanations as formulas can also let us handle more reasoning patterns w.r.t. wh-compliments such as those discussed by Groenendijk and Stokhof (1982), e.g., given John knows why Mary was late, then from Mary was late because of the traffic jam, we can infer that John knows that there was a traffic jam.

A promising future study is about group notions of knowing why. For example, how do we define everyone knows why φ? Simply having a conjunction of $\mathcal{K}y_i\varphi$ for each i may not be enough, since people can have different explanations for φ. The case

[17] We may also discuss whether $\mathcal{K}y_i\varphi \to \mathcal{K}y_i\mathcal{K}_i\varphi$ is reasonable.

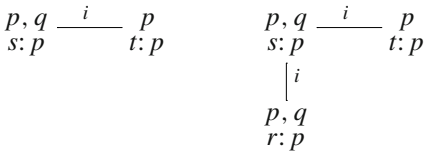
[18] There are also modal logic approaches to handle scientific explanations cf. e.g, Šešelja and Straßer (2013) and Sedlár and Halas (2015).

of *commonly* knowing why φ is more interesting. For example, we may have different definitions:

- It is (standard) common knowledge that everyone knows why φ w.r.t. the same explanation.
- Everyone knows why …everyone knows why φ.

In contrast to standard epistemic logic, such definitions can be quite different from each other. Since each iteration of $\mathcal{K}y_i$ may ask for a new explanation, we then have a much richer spectrum of such common knowledge notions, e.g., for the second definition, we may ask the agents to have exactly the same explanation for each level of "iteration of everyone knows why". It will be interesting to compare such notions with justified common knowledge (Artemov 2006; Bucheli et al. 2011).

Of course, we can also consider the dynamics of knowing why, similar to the dynamics in justification logic (Bucheli et al. 2014; Kuznets and Studer 2013; Renne 2008, 2012). Clearly, public announcements can change knowledge-why. However, in contrast to public announcement logic (Plaza 2007), adding public announcement will increase the expressive power of the logic, e.g., $[q]\mathcal{K}y_i\, p$ can distinguish the following two pointed models (the left-hand-side worlds as designated), which cannot be distinguished by formulas in **ELKy** (a simple inductive proof on the structure of the formula suffices):



In particular, $[q]\mathcal{K}y_i\, p$ is not equivalent to $q \rightarrow \mathcal{K}y_i(q \rightarrow p)$. To handle public announcements, we can follow the idea from Wang and Fan (2013) and generalize the knowing why operator to a conditional one to express that the agent i knows why φ given the condition ψ ($\mathcal{K}y_i(\psi, \varphi)$):

$$\mathcal{M}, w \vDash \mathcal{K}y_i(\psi, \varphi) \iff \exists t \in E, \text{for each } v \text{ such that } wR_iv \text{ and } \mathcal{M}, v \vDash \psi: (1)\ v \in \mathcal{E}(t, \varphi) \text{ and } (2)\ \mathcal{M}, v \vDash \varphi.$$

$\mathcal{K}y_i(\psi, \varphi)$ is similar to $[\psi]\mathcal{K}y_i\varphi$, and may be used to encode the announcements under further restrictions on models. An axiomatization of this conditionalized logic is presented by Pischke (2017).

On the other hand, there can be other natural dynamics, e.g., publicly announcing why, which is similar to public inspection introduced by Gattinger et al. (2017) in the setting of knowing values. A deeper connection between knowing why and dynamic epistemic logic is possible based on the observation that we do update according to events because we know why they happened (the preconditions). It is suggested by Olivier Roy that there is also a close connection with forward induction in games, where it is crucial to guess why someone did an apparently irrational move.

Finally, our work is also related to *explicit knowledge*, which aims to avoid logical omniscience. In fact, knowledge with justification or explanation can be viewed as

a type of explicit knowledge. One important approach to define explicit knowledge is by using *awareness*: φ is a piece of explicit knowledge of i ($X_i\varphi$) if i is aware of φ ($A_i\varphi$) and i implicitly knows that φ ($\mathcal{K}_i\varphi$), where awareness is often defined syntactically [cf. Fagin et al. (1995)]. Accordingly, the axioms are also changed, e.g., the K axiom now becomes $X_i(\varphi \to \psi) \wedge X_i\varphi \wedge A_i\psi \to X_i\psi$. Other approaches to explicit knowledge use the idea of *algorithmic knowledge* (Halpern and Pucella 2011). We may explore the concrete connection in the future.

Acknowledgements We thank Albert Anglberger, Johan van Benthem, Huimin Dong, Mel Fitting, Dominik Klein, Fenrong Liu, Olivier Roy, Martin Stokhof, Che-Ping Su, Frank Veltman, Lu Wang, Wei Wang, Junhua Yu, Liying Zhang, and Yuncheng Zhou for discussions and useful suggestions on earlier versions of this paper. The insightful comments from the anonymous reviewers of this journal also helped in improving the paper. Yanjing Wang acknowledges the support from the National Program for Special Support of Eminent Professionals and NSSF major Project 12&ZD119. Thomas Studer is supported by the Swiss National Science Foundation Grant 200021_165549.

References

Artemov, S. (1995). *Operational modal logic*. Technical report MSI 9529, Cornell University.
Artemov, S. (2001). Explicit provability and constructive semantics. *Bulletin of Symbolic Logic*, *7*(1), 1–36.
Artemov, S. (2006). Justified common knowledge. *Theoretical Computer Science*, *357*(1), 4–22.
Artemov, S. (2008). The logic of justification. *The Review of Symbolic Logic*, *1*(04), 477–513.
Artemov, S., & Nogina, E. (2005). Introducing justification into epistemic logic. *Journal of Logic and Computation*, *15*(6), 1059–1073.
Bird, A. (1998). *Philosophy of science*. Abingdon: Routledge.
Bromberger, S. (1966). Questions. *The Journal of Philosophy*, *63*(20), 597–606.
Bucheli, S., Kuznets, R., & Studer, T. (2011). Justifications for common knowledge. *Journal of Applied Non-classical Logics*, *21*(1), 35–60.
Bucheli, S., Kuznets, R., & Studer, T. (2014). Realizing public announcements by justifications. *Journal of Computer and System Sciences*, *80*(6), 1046–1066.
Fagin, R., Halpern, J., Moses, Y., & Vardi, M. (1995). *Reasoning about knowledge*. Cambridge: MIT Press.
Fan, J., Wang, Y., & van Ditmarsch, H. (2015). Contingency and knowing whether. *The Review of Symbolic Logic*, *8*, 75–107.
Fitting, M. (2005). The logic of proofs, semantically. *Annals of Pure and Applied Logic*, *132*(1), 1–25.
Fitting, M. (2008). A quantified logic of evidence. *Annals of Pure and Applied Logic*, *152*(1), 67–83.
Fitting, M. (2016). Modal logics, justification logics, and realization. *Annals of Pure and Applied Logic*, *167*(8), 615–648.
Gattinger, M., van Eijck, J., & Wang, Y. (2017). Knowing values and public inspection. In *Proceedings of ICLA '17* (forthcoming).
Groenendijk, J., & Stokhof, M. (1982). Semantic analysis of wh-complements. *Linguistics and Philosophy*, *5*(2), 175–233.
Halpern, J. Y., & Pucella, R. (2011). Dealing with logical omniscience: Expressiveness and pragmatics. *Artificial Intelligence*, *175*(1), 220–235.
Hempel, C. (1965). *Aspects of scientific explanation and other essays in the philosophy of science*. Mankato: The Free Press.
Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science*, *15*(2), 135–175.
Hintikka, J. (1962). *Knowledge and belief: An introduction to the logic of the two notions* (Vol. 181). Ithaca: Cornell University Press.
Hintikka, J. (1981). On the logic of an interrogative model of scientific inquiry. *Synthese*, *47*(1), 69–83.
Hintikka, J. (1983). *New foundations for a theory of questions and answers* (pp. 159–190). Dordrecht: Springer.
Hintikka, J., & Halonen, I. (1995). Semantics and pragmatics for why-questions. *The Journal of Philosophy*, *92*(12), 636–657.

Kitcher, P. (1981). Explanatory unification. *Philosophy of Science*, *48*(4), 507–531.
Koura, A. (1988). An approach to why-questions. *Synthese*, *74*(2), 191–206.
Kuznets, R., & Studer, T. (2012). Justifications, ontology, and conservativity. In T. Bolander, T. Braüner, S. Ghilardi, & L. Moss (Eds.), *Advances in modal logic* (Vol. 9, pp. 437–458). London: College Publications.
Kuznets, R., & Studer, T. (2013). Update as evidence: Belief expansion. In S. Artemov & A. Nerode (Eds.), *Logical foundations of computer science*, LFCS 2013, Springer, LNCS, Vol. 7734, pp. 266–279.
Kuznets, R., & Studer, T. (2019). *Justification logic*. London: College Publications.
Padmanabha, A., Ramanujam, R., & Wang, Y. (2018). Bundled fragments of first-order modal logic: (Un)decidability. In *Proceedings of FSTTCS 2018* (pp. 43:1–43:20).
Pischke, N. (2017). Dynamic extensions for the logic of knowing why with public announcements of formulas. arXiv e-prints 1707.05617.
Plaza, J. (2007). Logics of public communications. *Synthese*, *158*(2), 165–179.
Renne, B. (2008). *Dynamic epistemic logic with justification*. PhD thesis, New York, NY, USA, aAI3310607.
Renne, B. (2012). Multi-agent justification logic: Communication and evidence elimination. *Synthese*, *185*(1), 43–82.
Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.
Schurz, G. (1995). Scientific explanation: A critical survey. *Foundations of Science*, *1*(3), 429–465.
Schurz, G. (1999). Explanation as unification. *Synthese*, *120*(1), 95–114.
Schurz, G. (2005). Explanations in science and the logic of why-questions: Discussion of the Halonen–Hintikka-approachand alternative proposal. *Synthese*, *143*(1), 149–178.
Sedlár, I., & Halas, J. (2015). Modal logics of abstract explanation frameworks. In *Abstract in proceedings of CLMPS 15*.
Šešelja, D., & Straßer, C. (2013). Abstract argumentation and explanation applied to scientific debates. *Synthese*, *190*(12), 2195–2217.
van Benthem, J. (1991). Reflections on epistemic logic. *Logique and Analyse*, *34*(133–134), 5–14.
van Ditmarsch, H., Halpern, J. Y., van der Hoek, W., & Kooi, B. (Eds.). (2015). *Handbook of epistemic logic*. London: College Publications.
van Fraassen, B. C. (1980). *The scientific image*. Oxford: Oxford University Press.
Wang, Y. (2015). A logic of knowing how. In *Proceedings of LORI-V* (pp. 392–405).
Wang, Y. (2017). A new modal framework for epistemic logic. In *Proceedings of TARK '17* (pp. 515–534).
Wang, Y., & Fan, J. (2013). Knowing that, knowing what, and public communication: Public announcement logic with Kv operators. In *Proceedings of IJCAI'13* (pp. 1139–1146).
Wang, Y. (2018a). Beyond knowing that: A new generation of epistemic logics. In H. van Ditmarsch & G. Sandu (Eds.), *Jaakko Hintikka on knowledge and game theoretical semantics, outstanding contributions to logic* (Vol. 12, pp. 499–533). Berlin: Springer.
Wang, Y. (2018b). A logic of goal-directed knowing how. *Synthese*, *195*(10), 4419–4439.
Wang, Y., & Fan, J. (2014). Conditionally knowing what. *Advances in Modal Logic*, *10*, 569–587.
Weber, E., van Bouwel, J., & De Vreese, L. (2013). *Scientific explanation*. Berlin: Springer.
Yavorskaya, T. (2006). Multi-agent explicit knowledge. In D. Grigoriev, J. Harrison, & E. A. Hirsch (Eds.), *Proceedings of CSR 2006* (pp. 369–380). Berlin: Springer.