Logic and **L**ogical **P**hilosophy
Online First Articles (2025)
DOI: 10.12775/LLP.2025.027

Arkadiusz Wójcik

Fitch's Paradox in Fusions of Epistemic and Alethic Logics

Abstract. In this paper, we analyze Fitch's paradox of knowability in the framework of fusions of epistemic and alethic modal logics. The paradox arises from accepting the knowability principle, which states that all truths are knowable. However, this leads to the unacceptable conclusion that all truths are known. We introduce a logical system that incorporates all assumptions used by Fitch in his original reasoning, including the knowability principle. We present a natural semantics for this logic, proving the soundness and completeness theorem. Additionally, we present a new semantic proof of the knowability paradox, demonstrating that the problematic conclusion can be derived independently of Fitch's original proof and showing that the knowability principle itself is the source of the paradox. Using the formal tools introduced, we conduct a semantic analysis of the paradox, which allows us to identify the root cause of its occurrence. Finally, we propose a weakened version of the knowability principle that avoids paradoxical conclusions.

Keywords: Fitch's paradox; knowability paradox; knowability principle; epistemic logic; fusions of modal logics; modal correspondence theory

1. Introduction

One of the fundamental philosophical questions concerns the limits of knowledge. The debate over this question centers around the following thesis:

- *All truths are knowable.* (The knowability principle)

This can be symbolized as

$$\varphi \to \Diamond K\varphi, \tag{KP}$$

Received January 27, 2025. *Accepted* November 12, 2025. *Published online* December 5, 2025

where $\Diamond$ is the operator of possibility (i.e., $\Diamond\varphi$ should be read as "it is possible that φ"), and K is the knowledge operator (i.e., $K\varphi$ should be interpreted as "it is known that φ" or "someone knows that φ").

The knowability principle is primarily endorsed by verificationists and semantic anti-realists. As highlighted by Salerno [38], it is also accepted by proponents of various philosophical positions, including Putnam's internal realism, the logical positivisms of the Berlin and Vienna Circles, Peirce's pragmatism, Kant's transcendental idealism, and Berkeley's metaphysical idealism. Kinkaid [24] adds Husserl and representatives of phenomenology to this list, while Bueno [6] includes Field and proponents of fictionalism in the philosophy of mathematics. Other notable authors, for reasons unrelated to the aforementioned philosophical positions, emphasize the intuitive plausibility of the knowability principle (see, e.g., [3, 34, 35, 36, 44]).

Nevertheless, Fitch [17] has shown that (KP) leads to a paradoxical conclusion. To demonstrate this, he made as assumptions two axioms:

$$K(\varphi \wedge \psi) \to (K\varphi \wedge K\psi), \tag{$\mathtt{M}_K$}$$

$$K\varphi \to \varphi, \tag{$\mathtt{T}_K$}$$

and two rules:

$$\text{if} \vdash \varphi, \text{then} \vdash \Box\varphi, \tag{RN}$$

$$\text{if} \vdash \Box\neg\varphi, \text{then} \vdash \neg\Diamond\varphi. \tag{R$\Box\neg$}$$

These assumptions appear plausible. According to ($\mathtt{M}_K$), knowledge is distributive over conjunction (*monotonic*), whereas ($\mathtt{T}_K$) expresses the factivity of knowledge. (RN) is the rule of necessitation, i.e., a standard rule of modal logics, stating that if formula φ is provable, then formula $\Box\varphi$ is also provable, where $\Box\varphi$ is interpreted as "it is necessary that φ". Rule (R$\Box\neg$) is also uncontroversial, because $\Box$ and $\Diamond$ are interdefinable.

Fitch's proof starts by substituting for (KP) the sentence known as Moore sentence:

$$(p \wedge \neg Kp) \to \Diamond K(p \wedge \neg Kp). \tag{$\mathtt{KP}_{\mathtt{MS}}$}$$

According to the following argument, Moore sentence $p \wedge \neg Kp$ cannot be the object of knowledge:

1. $K(p \wedge \neg Kp)$	premise for a reductio
2. $K(p \wedge \neg Kp) \to (Kp \wedge K\neg Kp)$	substitution for ($\mathtt{M}_K$)
3. $Kp \wedge K\neg Kp$	from 1 and 2 by classical logic
4. Kp	from 3 by classical logic

5. $K\neg Kp$ from 3 by classical logic
6. $K\neg Kp \to \neg Kp$ substitution for ($\mathtt{T}_K$)
7. $\neg Kp$ from 5 and 6 by classical logic

Because we obtain a contradiction, we can infer the following thesis:

$$\neg K(p \wedge \neg Kp). \tag{$\star$}$$

The next steps in the reasoning are as follows:

1. $\Box\neg K(p \wedge \neg Kp)$ from ($\star$) by the rule (RN)
2. $\neg\Diamond K(p \wedge \neg Kp)$ from 1 by the rule ($\mathrm{R}\Box\neg$)
3. $\neg(p \wedge \neg Kp)$ from ($\mathtt{KP}_{\mathtt{MS}}$) and 2 by classical logic
4. $p \to Kp$ from 3 by classical logic

In item 4, the variable p could be substituted with any formula, leading to the conclusion

$$\varphi \to K\varphi. \tag{OP}$$

This result appears paradoxical because (OP) formalizes a clearly false principle of omniscience:

- *All truths are known.* (The omniscience principle)

The issue identified by Fitch is known in the literature as "the knowability paradox", "Fitch's paradox of knowability", or, due to the role of Church's suggestion [see, e.g., 39], "the Church–Fitch paradox".

The knowability paradox has been the subject of hundreds of papers. This extensive literature reveals several standard responses to the paradox: (i) revising one of the assumptions about the knowledge operator, (ii) revising classical logic, or (iii) restricting the knowability principle. None of the proposed solutions have gained widespread acceptance, leaving the problem relevant and open for further inquiry.

Strategy (i) [see, e.g., 8, 32] seems the least plausible. Both ($\mathtt{M}_K$) and ($\mathtt{T}_K$) are difficult to challenge, and furthermore, there are alternative versions of Fitch's proof that employ weaker versions of these assumptions [see 23, 31, 50].

Choosing solution (ii), which involves replacing classical logic with a non-classical system such as intuitionistic [11, 14, 48], relevance [47], or paraconsistent logic [3, 34], seems too radical and *ad hoc*. Proponents of semantic anti-realism can respond to this challenge because their use of intuitionistic logic is motivated by reasons independent of Fitch's paradox. Nevertheless, intuitionistic logic has led to other counterintuitive

claims ([33]). While some of these implausible consequences have been justified within the intuitionistic framework [48, 49], there is still ambiguity as to whether these reinterpretations are not *ad hoc* [29].

Strategy (iii) [see, e.g., 13, 15, 40, 44] appears most promising, though its applications are often criticized as *ad hoc* [see, e.g., 5, 12, 20]. To avoid this objection, it is necessary to provide justification for the proposed restriction independent of the knowability paradox or to develop a diagnosis of the paradox that supports the suggested solution.

It should be noted that the proof of the knowability paradox, as reconstructed above, is purely syntactic. However, Fitch's original argument was not conducted within any specific logical system. Rather, it was a formal representation of reasoning based on the stated assumptions and several rules of classical logic. This naturally leads to the question: How must the logic and its appropriate semantics be structured to replicate the paradox? To develop a logic that incorporates all of Fitch's assumptions, it seems appropriate to consider a method of combining two modal logical systems. Because assumptions ($\mathtt{M}_K$) and ($\mathtt{T}_K$) hold in all standard epistemic modal logics, and assumptions (RN) and (R$\Box\neg$) are accepted in all alethic modal logics, combining these two types of systems would yield the required logic. The method that appears most suitable for achieving such a combination is the so-called fusion of modal logics. This approach not only allows for a semantic analysis of the knowability paradox but also enables the study of the actual strength of various principles related to ($\mathtt{KP}$). Consequently, it becomes possible to formulate a weakened version of the knowability principle that avoids paradoxical conclusions.

Outline of the paper. In Section 2, we provide preliminary information that introduces the basics of modal systems and the methods by which they can be combined. Section 3 presents a fusion of two modal logics that satisfies the assumptions ($\mathtt{M}_K$), ($\mathtt{T}_K$), (RN), and (R$\Box\neg$). This base logic is then extended in Section 4 with the addition of ($\mathtt{KP}$), where we prove the soundness and completeness theorem. In Section 5, we use the introduced formalism to conduct a semantic analysis of Fitch's paradox. Finally, in Section 6, we propose a restriction on the knowability principle that resolves the paradox.

2. Preliminaries: Modal logics and their fusions

To develop a logic that incorporates all of Fitch's assumptions, we must formulate it in a language that includes both epistemic and alethic modal operators.

DEFINITION 2.1. Let Var denote the set of propositional variables. The *epistemic-alethic language* $\mathcal{L}_{K,\Box}$ is defined inductively as follows:

$$\varphi \ ::= \ p \mid \neg\varphi \mid \varphi \to \psi \mid K\varphi \mid \Box\varphi,$$

where $p \in Var$. The set of all $\mathcal{L}_{K,\Box}$ formulas is denoted by $\Gamma_{\mathcal{L}_{K,\Box}}$. The fragment of the language without the operator $\Box$ is denoted by $\mathcal{L}_K$, and the fragment without the operator K by $\mathcal{L}_\Box$. The sets of formulas of these languages are denoted by $\Gamma_{\mathcal{L}_K}$ and $\Gamma_{\mathcal{L}_\Box}$, respectively.

The language $\mathcal{L}_{K,\Box}$ extends the language of propositional logic with the knowledge operator K and the necessity operator $\Box$. We use the symbol $\top$ as an abbreviation for $p \vee \neg p$ and $\bot$ to denote $\neg\top$. Other classical logical constants are defined in the standard way. Additionally, the dual operators $\langle K\rangle$ and $\Diamond$ are defined as follows:

$$\langle K\rangle\varphi := \neg K\neg\varphi, \qquad (\texttt{Def}\,\langle K\rangle)$$
$$\Diamond\varphi := \neg\Box\neg\varphi. \qquad (\texttt{Def}\,\Diamond)$$

The semantics for the language $\mathcal{L}_{K,\Box}$ is constructed based on the semantics proposed by Kripke [26, 27]. However, in this language, there are two types of modalities, so its interpretation requires the consideration of two distinct accessibility relations: epistemic and alethic. This distinction is reflected in the following definitions of key semantic terms.

DEFINITION 2.2. An *epistemic-alethic frame* is a tuple $\mathcal{F} = (W, R_K, R_\Box)$, where

- $W \neq \emptyset$ is a set of states,
- $R_K \subseteq W \times W$ is an epistemic accessibility relation,
- $R_\Box \subseteq W \times W$ is an alethic accessibility relation.

The notions of a (purely) *epistemic frame* (W, R_K) and a (purely) *alethic frame* $(W, R_\Box)$ are obtained by dropping $R_\Box$ and R_K, respectively. A *model* is a pair $\mathcal{M} = (\mathcal{F}, v)$, where $\mathcal{F}$ is a frame and $v\colon Var \to \mathcal{P}(W)$ is a valuation function; such a model is said to be *based on* $\mathcal{F}$.

DEFINITION 2.3. Let $\mathcal{M} = (W, R_K, R_\Box, v)$ be an epistemic-alethic model, and let $s \in W$. The *satisfaction relation* $\models$ is defined inductively as follows:

$$\begin{array}{lll} \mathcal{M}, s \models p & \text{iff} & s \in v(p), \\ \mathcal{M}, s \models \neg\varphi & \text{iff} & \mathcal{M}, s \not\models \varphi, \\ \mathcal{M}, s \models \varphi \to \psi & \text{iff} & \text{if } \mathcal{M}, s \models \varphi, \text{ then } \mathcal{M}, s \models \psi, \\ \mathcal{M}, s \models K\varphi & \text{iff} & \text{for all } t \in W\text{: if } sR_Kt, \text{ then } \mathcal{M}, t \models \varphi, \\ \mathcal{M}, s \models \Box\varphi & \text{iff} & \text{for all } t \in W\text{: if } sR_\Box t, \text{ then } \mathcal{M}, t \models \varphi, \end{array}$$

where $p \in Var$ and $\varphi, \psi \in \Gamma_{\mathcal{L}_{K,\Box}}$.

The other terms are defined following standard conventions in modal logic. A formula φ is *true in a model* $\mathcal{M}$ (written $\mathcal{M} \models \varphi$) if for any state s, $\mathcal{M}, s \models \varphi$. A formula φ is *true in a frame* $\mathcal{F}$ (written $\mathcal{F} \models \varphi$) if for any model $\mathcal{M}$ based on the frame $\mathcal{F}$, $\mathcal{M} \models \varphi$. A formula φ is *valid* (written $\models \varphi$) if for any model $\mathcal{M}$, $\mathcal{M} \models \varphi$. A formula φ is *valid with respect to a class of frames* $\mathcal{C}$ (written $\mathcal{C} \models \varphi$) if for any $\mathcal{F} \in \mathcal{C}$, $\mathcal{F} \models \varphi$. A formula scheme S is *true* (or *valid*) *in a model/frame/class of frames* if every instance of S is true (or valid) in that model/frame/class of frames.

Having established the basic semantic framework for interpreting the language $\mathcal{L}_{K,\Box}$, we now turn to the presentation of specific modal logics.

DEFINITION 2.4. Let $\nabla \in \{K, \Box\}$. The *logic* $\mathbf{K}_\nabla$ is the smallest set of formulas that contains:

- all instantiations of propositional tautologies, (PC)
- the following axiom scheme:

$$\nabla(\varphi \to \psi) \to (\nabla\varphi \to \nabla\psi), \qquad (\mathtt{K}_\nabla)$$

and is closed under the following rules:

- from $\varphi \to \psi$ and φ, infer ψ, (modus ponens)
- from φ, infer $\nabla\varphi$. (Gödel's rule for ∇)

The logic $\mathbf{K}_\Box$ is the minimal normal alethic modal logic. A *normal alethic modal logic* is a set of formulas that contains PC, every instance of the axiom scheme ($\mathtt{K}_\Box$), i.e. ($\mathtt{K}_\nabla$) for $\nabla = \Box$, and is closed under the rules of modus ponens and Gödel's rule (RN) for $\Box$. Normal epistemic modal logics are obtained from normal alethic modal logics by replacing the operator $\Box$ with the operator K. In semantic terms, the logics $\mathbf{K}_\Box$ and $\mathbf{K}_K$ can be defined as the sets of formulas that are valid on all alethic and all epistemic frames, respectively.

Name	Axiom	Property of R_∇
$\mathtt{D}_\nabla$	$\nabla\varphi \to \neg\nabla\neg\varphi$	serial
$\mathtt{T}_\nabla$	$\nabla\varphi \to \varphi$	reflexive
$\mathtt{4}_\nabla$	$\nabla\varphi \to \nabla\nabla\varphi$	transitive
$\mathtt{5}_\nabla$	$\neg\nabla\varphi \to \nabla\neg\nabla\varphi$	Euclidean

Table 1. Some basic modal axiom schemes for $\nabla \in \{K, \Box\}$

Stronger normal alethic (or epistemic) modal logics are obtained by adding all instances of a given axiom scheme to $\mathbf{K}_\Box$ (or $\mathbf{K}_K$). Semantically, these logics are modeled by restricting the class of frames to those that satisfy certain conditions on the accessibility relation. These conditions are determined by the specific axiom scheme added to the logic. A formula scheme S *corresponds* to a first-order formula Φ if the following holds for all frames $\mathcal{F}$: $\mathcal{F} \models S$ iff $\mathcal{F}$ satisfies the condition expressed by Φ. In that case, we also say that the scheme S *defines* the class of frames with condition Φ. Table 1 provides the fundamental axiom schemes for both epistemic and alethic operators, which will be employed in the subsequent sections, along with their corresponding frame conditions.

Let $\nabla \in \{K, \Box\}$. We adopt a variant of Lemmon's naming convention for modal logics, where the notation $\mathbf{KA} \dots \mathbf{Z}_\nabla$ denotes the logic obtained by adding the axiom schemes A, ..., Z (each specialized to ∇) to the logic $\mathbf{K}_\nabla$. Some exceptions to this convention apply to certain well-known normal logics, among which the following will be referred to in this work: $\mathbf{T}_\nabla = \mathbf{KT}_\nabla$, $\mathbf{S5}_\nabla = \mathbf{KT5}_\nabla$.

For readability, in some cases we also use the notation $\mathbf{L} + S$ for the smallest normal modal logic extending $\mathbf{L}$ by all instances of the axiom scheme S.

Given that $\mathbf{L}$ is a normal modal logic, we use $\mathbf{L} \vdash \varphi$ to denote that $\varphi \in \mathbf{L}$; in such a case, we say that φ is *provable* in $\mathbf{L}$ or simply $\mathbf{L}$-*provable*. A scheme S is provable in $\mathbf{L}$ if every instance of S is provable in $\mathbf{L}$. Logic $\mathbf{L}$ is *sound* with respect to a class of frames $\mathcal{C}$ if for any φ, $\mathbf{L} \vdash \varphi$ implies $\mathcal{C} \models \varphi$. Logic $\mathbf{L}$ is *complete* with respect to a class of frames $\mathcal{C}$ if for any φ, $\mathcal{C} \models \varphi$ implies $\mathbf{L} \vdash \varphi$. The standard normal modal logics are sound and complete with respect to classes of frames whose accessibility relations satisfy properties expressible in first-order logic. Based on Table 1, we can establish many such results for various normal modal logics. For example, the logic $\mathbf{S5}_\Box$ is sound and complete with

respect to the class of all reflexive and Euclidean alethic frames. This follows from the fact that $\mathbf{S5}_{\Box} = \mathbf{KT5}_{\Box}$, and the axiom schemes ($\mathtt{T}_{\Box}$) and ($\mathtt{5}_{\Box}$) define the classes of reflexive and Euclidean alethic frames, respectively. As reflexivity and Euclideanness entail both symmetry and transitivity, $\mathbf{S5}_{\Box}$ is equivalently sound and complete with respect to the class of all equivalence alethic frames.

For the purposes of this article, it is crucial to identify effective methods for combining epistemic modal logics with alethic modal logics. We will focus on one method of combining modal systems: the fusion of modal logics.

DEFINITION 2.5. Let $\mathbf{L}_K$ and $\mathbf{L}_{\Box}$ be normal modal logics formulated in the languages $\mathcal{L}_K$ and $\mathcal{L}_{\Box}$, respectively. The *fusion* $\mathbf{L}_K \otimes \mathbf{L}_{\Box}$ of the logics $\mathbf{L}_K$ and $\mathbf{L}_{\Box}$ is the smallest normal modal logic formulated in the language $\mathcal{L}_{K,\Box}$ that contains $\mathbf{L}_K \cup \mathbf{L}_{\Box}$.

In particular, if $\mathbf{L}_K$ and $\mathbf{L}_{\Box}$ are axiomatized by the sets Ax_1 and Ax_2, respectively, then $\mathbf{L}_K \otimes \mathbf{L}_{\Box}$ is axiomatized by the union $Ax_1 \cup Ax_2$. By a general result of Thomason [41], the fusion $\mathbf{L}_K \otimes \mathbf{L}_{\Box}$ is a conservative extension of both $\mathbf{L}_K$ and $\mathbf{L}_{\Box}$, which means that $\mathbf{L}_K = (\mathbf{L}_K \otimes \mathbf{L}_{\Box}) \cap \Gamma_{\mathcal{L}_K}$ and $\mathbf{L}_{\Box} = (\mathbf{L}_K \otimes \mathbf{L}_{\Box}) \cap \Gamma_{\mathcal{L}_{\Box}}$.

Established results demonstrate the transference of specific metalogical properties from the components to the fusions of modal logics (see, e.g., [19] or [28] for a survey). Particularly relevant in this context is the result by Kracht and Wolter [25] and Fine and Schurz [18], which states the preservation of soundness and completeness. As a direct consequence of this result, the following theorem holds for epistemic-alethic fusions:

THEOREM 2.1. *Let* $\mathbf{L_K}$ *and* $\mathbf{L}_{\Box}$ *be normal modal logics formulated in the languages* $\mathcal{L}_K$ *and* $\mathcal{L}_{\Box}$ *that are sound and complete with respect to classes of frames* $\mathcal{C}_1$ *and* $\mathcal{C}_2$*, respectively. Then, the fusion* $\mathbf{L}_K \otimes \mathbf{L}_{\Box}$ *is sound and complete with respect to the class*

$$\mathcal{C}_1 \otimes \mathcal{C}_2 = \{(W, R_K, R_{\Box}) : (W, R_K) \in \mathcal{C}_1,\ (W, R_{\Box}) \in \mathcal{C}_2\}.$$

3. Towards a base logic for Fitch's paradox

In order to obtain a logic that satisfies assumptions ($\mathtt{M}_K$), ($\mathtt{T}_K$), (RN), and (R$\Box\neg$), which were adopted in the proof of Fitch's paradox, we

propose considering the fusion $\mathbf{T}_K \otimes \mathbf{K}_\Box$.[1] In the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box$, all necessary assumptions hold. Any formula with the scheme ($\mathtt{M}_K$) is provable in the logic $\mathbf{K}_K$, and therefore, it is also provable in the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box$.

Proposition 3.1. $\mathbf{T}_K \otimes \mathbf{K}_\Box \vdash K(\varphi \wedge \psi) \to (K\varphi \wedge K\psi)$.

Scheme ($\mathtt{T}_K$) is adopted as an axiom scheme by using ($\mathtt{T}_\nabla$) from Table 1 for $\nabla = K$, while rule (RN) is simply Gödel's rule for $\Box$. Assumption (R$\Box\neg$) holds by virtue of ($\mathtt{Def}\,\Diamond$) and the law of double negation: for any formula $\varphi \in \Gamma_{\mathcal{L}_{K,\Box}}$, we have $(\neg\Diamond\varphi) \leftrightarrow (\neg\neg\Box\neg\varphi) \leftrightarrow (\Box\neg\varphi)$. The axiom scheme ($\mathtt{K}_\Box$) and Gödel's rule for K were not used in Fitch's proof, but they must be included to ensure that the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box$ is a fusion of two normal modal logics.

As a direct consequence of Theorem 2.1 and the soundness and completeness of the logics $\mathbf{T}_K$ and $\mathbf{K}_\Box$ with respect to their corresponding classes of frames, the following theorem holds:

Theorem 3.1. *Logic $\mathbf{T}_K \otimes \mathbf{K}_\Box$ is sound and complete with respect to the class of all epistemic-alethic frames where the relation R_K is reflexive.*

Although the axiom schemes of $\mathbf{T}_K \otimes \mathbf{K}_\Box$ do not include schemes that combine epistemic and alethic operators, formulas of this type can still be provable in this logic.

Proposition 3.2. 1. $\mathbf{T}_K \otimes \mathbf{K}_\Box \vdash \Diamond K\varphi \to \Diamond\varphi$,
2. $\mathbf{T}_K \otimes \mathbf{K}_\Box \vdash \Box K\varphi \to \Box\varphi$.

However, it is crucial to note that the scheme corresponding to the knowability principle is not provable in the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box$.

Proposition 3.3. $\mathbf{T}_K \otimes \mathbf{K}_\Box \nvdash \varphi \to \Diamond K\varphi$.

Proof. Consider the model $\mathcal{M} = (W, R_K, R_\Box, v)$ such that $W = \{s\}$, $R_K = \{(s,s)\}$, $R_\Box = \emptyset$, $v(p) = \{s\}$. Note that $\mathcal{M}, s \not\models p \to \Diamond Kp$. Hence $\mathcal{M} \not\models p \to \Diamond Kp$. Since $\mathcal{M}$ is a $\mathbf{T}_K \otimes \mathbf{K}_\Box$-model, we obtain $\mathbf{T}_K \otimes \mathbf{K}_\Box \not\models \varphi \to \Diamond K\varphi$. By Theorem 3.1, it follows that $\mathbf{T}_K \otimes \mathbf{K}_\Box \nvdash \varphi \to \Diamond K\varphi$. ⊣

This raises the question: Is there a plausible fusion of epistemic and alethic logics in which ($\mathtt{KP}$) is provable? Given that the standard epistemic logic is considered to be $\mathbf{S5}_K$, and furthermore, the axioms of

[1] The suggestion to investigate Fitch's paradox using a fusion of modal logics was also proposed by Costa-Leite [9, 10]. Nevertheless, the lack of a formulated semantics for the proposed logic limited the scope of his analysis.

the logic $\mathbf{S5}_\Box$ characterizing the necessity operator also seem intuitively acceptable, it would be preferable if (KP) were provable in the fusion $\mathbf{S5}_K \otimes \mathbf{S5}_\Box$. Yet, by constructing an appropriate countermodel, it can be easily demonstrated that this scheme is not $\mathbf{S5}_K \otimes \mathbf{S5}_\Box$-provable.

PROPOSITION 3.4. $\mathbf{S5}_K \otimes \mathbf{S5}_\Box \nvdash \varphi \to \Diamond K\varphi$.

PROOF. Consider the model $\mathcal{M} = (W, R_K, R_\Box, v)$ such that $W = \{s, t\}$, $R_K = \{(s,s),(t,t),(s,t),(t,s)\}$, $R_\Box = \{(s,s),(t,t)\}$, and $v(p) = \{s\}$. Note that $\mathcal{M}, s \not\models p \to \Diamond Kp$. Hence $\mathcal{M} \not\models p \to \Diamond Kp$. Since $\mathcal{M}$ is a $\mathbf{S5}_K \otimes \mathbf{S5}_\Box$-model, we obtain $\mathbf{S5}_K \otimes \mathbf{S5}_\Box \not\models \varphi \to \Diamond K\varphi$. By soundness, it follows that $\mathbf{S5}_K \otimes \mathbf{S5}_\Box \nvdash \varphi \to \Diamond K\varphi$. ⊣

It is clear that the model presented in the proof of the above proposition also serves as a countermodel for (OP).

Nevertheless, it is possible to construct a fusion of epistemic and alethic logics in which any formula of the scheme $\varphi \to \Diamond K\varphi$ is provable. Let the fusion consist of the alethic logic $\mathbf{T}_\Box$ and the epistemic logic that contains the axiom scheme ($\mathtt{T}_K$) along with the following axiom scheme:

$$\langle K \rangle \varphi \to K\varphi, \tag{$\mathtt{D}_{cK}$}$$

which is the converse of the standard axiom scheme ($\mathtt{D}_K$), i.e. ($\mathtt{D}_\nabla$) from Table 1 for $\nabla = K$. Moreover, axiom ($\mathtt{T}_\Box$) defines the class of reflexive alethic frames, while ($\mathtt{D}_{cK}$) defines the class of epistemic frames satisfying the following condition:

$$\forall x, y, z((xR_Ky \wedge xR_Kz) \to y = z). \tag{Φ}$$

Based on Theorem 2.1, the logic $\mathbf{TD_c}_K \otimes \mathbf{T}_\Box$ is sound and complete with respect to the class of all epistemic-alethic frames where both relations R_K and $R_\Box$ are reflexive, along with the additional satisfaction of Φ. It can be easily checked that in this logic, (KP) is provable and that Fitch's paradox arises within it.

PROPOSITION 3.5. 1. $\mathbf{TD_c}_K \otimes \mathbf{T}_\Box \vdash \varphi \to \Diamond K\varphi$,
2. $\mathbf{TD_c}_K \otimes \mathbf{T}_\Box \vdash \varphi \to K\varphi$.

To consider this result significant and philosophically interesting, it would be necessary to justify ($\mathtt{D}_{cK}$). However, formulating such a justification does not seem feasible. In the logic $\mathbf{TD_c}_K \otimes \mathbf{T}_\Box$, the axiom scheme ($\mathtt{D}_K$) is provable. Ultimately, we obtain $\mathbf{TD_c}_K \otimes \mathbf{T}_\Box \vdash K\varphi \leftrightarrow \langle K \rangle \varphi$. This indicates that the logic $\mathbf{TD_c}_K \otimes \mathbf{T}_\Box$ does not distinguish between a scenario where an agent knows that φ and a scenario where φ is consis-

tent with the agent's entire knowledge. This limitation undermines its status as a plausible epistemic logic (i.e., a logic that captures the basic intuitions associated with the functioning of expressions such as "it is known that" and "agent knows that" in natural language).

4. The logic $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{KP})$

Since (KP) is not provable in $\mathbf{T}_K \otimes \mathbf{K}_\Box$ or in any natural epistemic-alethic fusion, it seems reasonable to add it as an additional axiom to the base logic.

DEFINITION 4.1. The logic $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{KP})$ is an extension of the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box$ by the addition of the axiom scheme (KP).

The primary aim now is to construct the semantics for the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{KP})$ and prove its soundness and completeness. A critical question is whether (KP) corresponds to a first-order formula that expresses the properties of the relations R_K and $R_\Box$[2]. To answer these queries, we must present several important results from modal correspondence theory.

Modal correspondence theory applies not only to epistemic and alethic modal logics but also to other systems, such as temporal or deontic logics. In many standard publications, definitions and theorems are presented in a general form applicable to all modal languages. Specifically, these languages can have any modal similarity type. A *modal similarity type* is defined as a pair $\tau = (O, \rho)$, where O is a non-empty set of modal operators and $\rho : O \mapsto \mathbb{N}$ is a function that represents their arities. For the purpose of presenting the key results of correspondence theory, we will focus exclusively on the language $\mathcal{L}_{K,\Box}$. We will begin by introducing two auxiliary definitions and defining an important class of formulas known as Sahlqvist formulas, named after Henrik Sahlqvist, who first described them in his work [37].

DEFINITION 4.2. Let $\varphi \in \Gamma_{\mathcal{L}_{K,\Box}}$ be any formula containing only the constants $\neg$, $\wedge$, and $\vee$. An occurrence of a propositional variable p in φ is *positive* (or *negative*) if it is under the scope of an even (or odd) number

[2] This question is valid because not all modal axioms correspond to properties of accessibility relations expressible in first-order logic. A frequently cited example is the McKinsey axiom scheme: $\Box\Diamond\varphi \to \Diamond\Box\varphi$ (see [42]).

of negations. A formula φ is *positive* (or *negative*) if all occurrences of propositional variables in φ are positive (or negative).

DEFINITION 4.3. Let $p \in Var$. A *boxed atom* is a formula of the form

$$\nabla_1 \dots \nabla_n p,$$

where $n \geqslant 0$, and $\nabla_i \in \{K, \Box\}$ for each $i = 1, \dots, n$. In the case where $n = 0$, the boxed atom $\nabla_1 \dots \nabla_n p$ is simply the propositional variable p.

DEFINITION 4.4. A *Sahlqvist antecedent* is a formula constructed from $\top$, $\bot$, negative formulas, and boxed atoms, using only $\wedge$, $\vee$, $\Diamond$, and $\langle K \rangle$. A *Sahlqvist implication* is an implication $\varphi \to \psi$ where φ is a Sahlqvist antecedent and ψ is a positive formula. A *Sahlqvist formula* is a formula constructed from Sahlqvist implications by freely applying K, $\Box$, $\wedge$, and by applying $\vee$ to formulas that share no common propositional variables. A *Sahlqvist scheme* is a scheme constructed from a Sahlqvist formula.

Remark 4.1. The scheme $\varphi \to \Diamond K \varphi$ is a Sahlqvist scheme.

Remark 4.1 is significant because, based on it and the following theorem, we can establish the existence of a corresponding first-order formula for $\varphi \to \Diamond K \varphi$.

THEOREM 4.1 ([37]). *Let S be a Sahlqvist scheme. Then, S corresponds to a first-order formula that is effectively computable from φ.*

The detailed proof of this theorem can be found in the work of Blackburn, de Rijke, and Venema [4, p. 165], where it is proven for any modal similarity type. Alternatively, the proof is available in Sahlqvist's original article [37, p. 121–123], although Sahlqvist's proof was formulated for a language with only one modal operator.

The proof of Theorem 4.1 relies on the Sahlqvist-van Benthem algorithm, independently provided by Sahlqvist [37] and van Benthem [43]. This algorithm allows for the computation of a corresponding first-order formula from a Sahlqvist formula. In Appendix A, we present the version of this algorithm for the language $\mathcal{L}_{K,\Box}$, along with a discussion of its application to the knowability principle. Subsequently, we prove the following proposition (with the corresponding number in parentheses):

PROPOSITION 4.1 (A.1). *Let $\mathcal{F} = (W, R_K, R_\Box)$ be an epistemic-alethic frame. Then, the following are equivalent:*

(a) $\mathcal{F} \models \varphi \to \Diamond K \varphi$,
(b) $\forall x \in W \; \exists y \in W (x R_\Box y \wedge \forall z \in W (y R_K z \to x = z))$.

As a result, we can accurately define the concept of a frame for the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{KP})$ as follows:

DEFINITION 4.5. An *epistemic-alethic frame for the logic* $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{KP})$ (in short, $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{KP})$*-frame*) is a frame $\mathcal{F} = (W, R_K, R_\Box)$, where

- $W \neq \emptyset$ is a set of epistemic states,
- $R_K \subseteq W \times W$ and $R_\Box \subseteq W \times W$ are relations that satisfy the following conditions:
  (i) $\forall x \in W(xR_Kx)$,
  (ii) $\forall x \in W\ \exists y \in W(xR_\Box y \wedge \forall z \in W(yR_Kz \to x = z))$.

The significance of Sahlqvist's research lies not only in proving the correspondence between Sahlqvist formulas and first-order logic formulas but also in illuminating the close connection between the structure of modal formulas and the soundness and completeness of normal modal logics.

THEOREM 4.2 ([37]). *Let* $\mathbf{L}$ *be a normal modal logic that is sound and complete with respect to the class of frames* $\mathcal{C}$*, and let* S *be a Sahlqvist scheme that defines the class of frames* $\mathcal{C}'$*. Then, the logic* $\mathbf{L} + S$ *is sound and complete with respect to the class of frames* $\mathcal{C} \cap \mathcal{C}'$*.*

Sahlqvist's completeness theorem enables the proof of the completeness of the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{KP})$.

THEOREM 4.3. *The logic* $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{KP})$ *is sound and complete with respect to the class of all* $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{KP})$*-frames.*

PROOF. According to Definition 4.1, the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{KP})$ is obtained by adding the axiom scheme $\varphi \to \Diamond K\varphi$ to the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box$. By virtue of Proposition 4.1, we know that this scheme defines the class of frames $\mathcal{F} = (W, R_K, R_\Box)$, such that $\forall x \in W \exists y \in W(xR_\Box y \wedge \forall z \in W(yR_Kz \to x = z))$. Furthermore, from Theorem 3.1, we know that the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box$ is sound and complete with respect to the class of frames $\mathcal{F} = (W, R_K, R_\Box)$, such that $\forall x \in W(xR_Kx)$. Therefore, by Theorem 4.2 and Remark 4.1, we conclude that the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{KP})$ is sound and complete with respect to the class of all $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{KP})$-frames. ⊣

5. Fitch's paradox in the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{KP})$

The established semantics allows us to prove that any formula of the scheme $\varphi \to K\varphi$ is valid in the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{KP})$.

THEOREM 5.1. $\mathbf{T}_K \otimes \mathbf{K}_\Box{+}(\mathtt{KP}) \models \varphi \to K\varphi$.

PROOF. Let $\mathcal{M} = (W, R_K, R_\Box, v)$ be a $\mathbf{T}_K \otimes \mathbf{K}_\Box{+}(\mathtt{KP})$-model. Then, the following conditions hold:

$$\forall x \in W\, \exists y \in W\, (xR_\Box y \wedge \forall z \in W\, (yR_K z \to x = z)), \tag{$*$}$$

$$\forall x \in W (xR_K x). \tag{$**$}$$

Let $s \in W$ be such that $\mathcal{M}, s \models \varphi$. We aim to show that $\mathcal{M}, s \models K\varphi$. Let $t \in W$ such that $sR_K t$. Since $s \in W$, by ($*$), choose $y \in W$ such that $sR_\Box y$ and the following condition holds:

$$\forall z \in W (yR_K z \to s = z). \tag{$***$}$$

By ($**$), $yR_K y$. Since $yR_K y$, from ($***$) we obtain $s = y$. Thus, from $s = y$ and $sR_K t$, it follows that $yR_K t$. Given that $yR_K t$, we conclude from ($***$) that $s = t$. Finally, from $s = t$ and $\mathcal{M}, s \models \varphi$, it follows that $\mathcal{M}, t \models \varphi$. Therefore, $\mathcal{M}, s \models K\varphi$. This concludes the proof that $\mathbf{T}_K \otimes \mathbf{K}_\Box{+}(\mathtt{KP}) \models \varphi \to K\varphi$. ⊣

This proof can be seen as a novel semantic proof of Fitch's paradox. As a direct consequence of Theorem 5.1 and Theorem 4.3, we obtain the following theorem:

THEOREM 5.2. $\mathbf{T}_K \otimes \mathbf{K}_\Box{+}(\mathtt{KP}) \vdash \varphi \to K\varphi$.

Independently of Fitch's original proof — without recourse to Moore sentences of the form $p \wedge \neg Kp$ — we have replicated the original result. This outcome provides a compelling argument against endorsing the knowability principle. Furthermore, we have demonstrated that, contrary to the assertions made by some authors [see, e.g., 29, 30], the possibility of deriving (OP) from (KP) does not arise from any fallacy in Fitch's original reasoning.

The adopted method not only enables a novel proof of Fitch's paradox but also offers several other significant advantages. Firstly, the presented semantics provides a basis for formulating a semantic explanation for the provability of $\varphi \to K\varphi$: accepting the knowability principle alongside the remaining assumptions acknowledged by Fitch implies that every world is epistemically isolated, meaning it has no epistemic alternatives other than itself.

Proposition 5.1. *Let* $\mathcal{F} = (W, R_K, R_\Box)$ *be a* $\mathbf{T}_K \otimes \mathbf{K}_\Box$+(KP)*-frame. Then* $\forall x, y \in W(xR_K y \to x = y)$.[3]

Proof. Let $\mathcal{F} = (W, R_K, R_\Box)$ be a $\mathbf{T}_K \otimes \mathbf{K}_\Box$+(KP)-frame. Then, the following conditions hold:

$$\forall x \in W \exists y \in W \left(xR_\Box y \wedge \forall z \in W \left(yR_K z \to x = z\right)\right), \tag{$*$}$$

$$\forall x \in W(xR_K x). \tag{$**$}$$

Let $x, y \in W$ be such that $xR_K y$. We aim to show that $x = y$. Since $x \in W$, by ($*$), choose $a \in W$ such that $xR_\Box a$ and the following condition holds:

$$\forall z \in W(aR_K z \to x = z). \tag{$***$}$$

By ($**$), $aR_K a$. Since $aR_K a$, from ($***$) we obtain $x = a$. From $xR_K y$ and $x = a$, it follows that $aR_K y$. Since $aR_K y$, from ($***$) we obtain $x = y$. ⊣

Interpreting the knowledge operator within relational semantics creates a strict dependency between the number of worlds in the epistemic accessibility relation for a given agent and what that agent knows. Explaining his use of possible worlds semantics for epistemic logic in *Knowledge and Belief* [21], Hintikka described the relationship between agents' knowledge as follows:

> *a* knows more than *b* if and only if the class of possible worlds compatible with what he knows is smaller than the class of possible worlds compatible with what *b* knows. [22, p. 157]

An agent who is completely ignorant — meaning they possess no knowledge — considers all worlds as possible; an agent who is omniscient considers only the actual world as possible. The standard interpretation of the scheme $\varphi \to K\varphi$, which expresses the omniscience of the agent, aligns with how omniscience is expressed in model-theoretic terms. This serves as an argument for the adequacy of using Kripke semantics to analyze the knowability paradox.

A second advantage is the ability to link the obtained results with Moore sentences, namely sentences of the form $p \wedge \neg Kp$. If for any formula φ, it holds that $\varphi \to \Diamond K\varphi$, then it must also hold for the formula

[3] The property of the relation R_K mentioned in Proposition 5.1 implies that $\mathbf{T}_K \otimes \mathbf{K}_\Box$+(KP) $\models \varphi \to K\varphi$. Referring to Proposition 5.1 and the fact that the scheme $\varphi \to K\varphi$ corresponds to a first-order formula $\forall x, y(xR_K y \to x = y)$, suffices to prove Theorem 5.1.

$p \wedge \neg Kp$; therefore, $(p \wedge \neg Kp) \rightarrow \Diamond K(p \wedge \neg Kp)$ must be valid. How can this be guaranteed? This can be achieved by ensuring the truth of the consequent or the falsity of the antecedent. However, in any epistemic-alethic system where the relation R_K is reflexive, such as in the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box$ and any of its extensions, the formula $\neg\Diamond K(p \wedge \neg Kp)$ is valid.

PROPOSITION 5.2. $\mathbf{T}_K \otimes \mathbf{K}_\Box \models \neg\Diamond K(p \wedge \neg Kp)$.

PROOF. Let $\mathcal{M} = (W, R_K, R_\Box, v)$ be a model of the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box$, and let $s \in W$. Suppose, for the sake of contradiction, that $\mathcal{M}, s \models \Diamond K(p \wedge \neg Kp)$. Thus, there exists $t \in W$ such that $sR_\Box t$ and $\mathcal{M}, t \models K(p \wedge \neg Kp)$. Given the reflexivity of the relation R_K, we have $tR_K t$. From $tR_K t$ and $\mathcal{M}, t \models K(p \wedge \neg Kp)$, it follows that $\mathcal{M}, t \models p \wedge \neg Kp$. Since $\mathcal{M}, t \models \neg Kp$, there exists $u \in W$ such that $tR_K u$ and $\mathcal{M}, u \not\models p$. However, since $tR_K u$, and given that $\mathcal{M}, t \models K(p \wedge \neg Kp)$, we obtain $\mathcal{M}, u \models p \wedge \neg Kp$. Therefore, we have $\mathcal{M}, u \models p$ and $\mathcal{M}, u \not\models p$. Thus, we have shown that $\mathcal{M}, s \models \neg\Diamond K(p \wedge \neg Kp)$. This concludes the proof that $\mathbf{T}_K \otimes \mathbf{K}_\Box \models \neg\Diamond K(p \wedge \neg Kp)$. ⊣

To ensure the falsity of the antecedent of the formula $(p \wedge \neg Kp) \rightarrow \Diamond K(p \wedge \neg Kp)$, it is necessary to reject the reflexivity of the relation R_K, which in turn implies rejecting the axiom ($\mathtt{T}_K$). However, in Fitch's assumptions, the axiom ($\mathtt{T}_K$) was explicitly adopted, so this route of defending ($\mathtt{KP}$) against counterexamples in the form of $p \wedge \neg Kp$ is closed.

What about ensuring the truth of the consequent? The semantic implications of adopting ($\mathtt{KP}$) ensure this, because, as stated in Proposition 5.1, adopting ($\mathtt{KP}$) implies that no world has any epistemic alternative other than itself. Therefore, the formula $p \wedge \neg Kp$ is false in every $\mathbf{T}_K \otimes \mathbf{K}_\Box$+($\mathtt{KP}$)-model. Let $\mathcal{M} = (W, R_K, v)$ be any epistemic model where the relation R_K satisfies $\forall x, y \in W(xR_K y \rightarrow x = y)$, and let $s \in W$ be given. Suppose, for the sake of contradiction, that $\mathcal{M}, s \models p \wedge \neg Kp$. By assumption, there exists $t \in W$ such that $sR_K t$ and $\mathcal{M}, t \not\models p$. Because R_K satisfies the condition $\forall x, y \in W(xR_K y \rightarrow x = y)$ and $sR_K t$, it follows that $s = t$. Therefore, we have $\mathcal{M}, s \models p$ and $\mathcal{M}, s \not\models p$. Thus, in the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box$+($\mathtt{KP}$) — in contrast to systems that only contain standard epistemic axioms — the formula $(p \wedge \neg Kp) \rightarrow \Diamond K(p \wedge \neg Kp)$ is valid, since the antecedent $p \wedge \neg Kp$ is unsatisfiable in every model of this logic.

The third advantage of the presented method of analysis is the ability to examine the role of the axiom ($\mathtt{T}_K$) in deriving the paradoxical consequence. Although Fitch used this axiom in his proof, the possibility

of deriving (OP) from (KP) without invoking the factivity of knowledge has not yet been ruled out. Now, it can be demonstrated that if the base logic is chosen to be the fusion $\mathbf{K}_K \otimes \mathbf{K}_\Box$, where no constraints are imposed on either the relation $R_\Box$ or the relation R_K, then, despite adding (KP), the scheme $\varphi \to K\varphi$ remains unprovable.

PROPOSITION 5.3. $\mathbf{K}_K \otimes \mathbf{K}_\Box{+}(\mathtt{KP}) \nvdash \varphi \to K\varphi$.

PROOF. Consider the model $\mathcal{M} = (W, R_K, R_\Box, v)$ such that $W = \{s, t\}$, $R_K = R_\Box = \{(s,t),(t,s)\}$, and $v(p) = \{s\}$. Note that $\mathcal{M}, s \not\models p \to Kp$, which implies $\mathcal{M} \not\models p \to Kp$. Given that $\mathcal{M}$ is a $\mathbf{K}_K \otimes \mathbf{K}_\Box{+}(\mathtt{KP})$-model, we obtain $\mathbf{K}_K \otimes \mathbf{K}_\Box{+}(\mathtt{KP}) \not\models \varphi \to K\varphi$. By soundness, this is equivalent to $\mathbf{K}_K \otimes \mathbf{K}_\Box{+}(\mathtt{KP}) \nvdash \varphi \to K\varphi$. ⊣

Without significant difficulty, it can also be shown, by constructing appropriate countermodels, that replacing the axiom scheme ($\mathtt{T}_K$) with any of the standard epistemic axioms schemes listed in Table 1 will not result in the provability of the scheme $\varphi \to K\varphi$.

However, (OP) will be provable in an epistemic-alethic logic that differs from the system $\mathbf{T}_K \otimes \mathbf{K}_\Box{+}(\mathtt{KP})$ by adopting two axioms, ($\mathtt{D}_K$) and ($\mathtt{4}_K$), instead of the axiom ($\mathtt{T}_K$).

PROPOSITION 5.4. $\mathbf{KD4}_K \otimes \mathbf{K}_\Box{+}(\mathtt{KP}) \vdash \varphi \to K\varphi$.

PROOF. Let $\mathcal{M} = (W, R_K, R_\Box, v)$ be a model of the logic $\mathbf{KD4}_K \otimes \mathbf{K}_\Box{+}(\mathtt{KP})$. Then, the following conditions hold:

$$\forall x \in W\ \exists y \in W(xR_\Box y \wedge \forall z \in W(yR_K z \to x = z)), \tag{$*$}$$

$$\forall x \in W\ \exists y \in W(xR_K y), \tag{$**$}$$

$$\forall x, y, z \in W((xR_K y \wedge yR_K z) \to xR_K z). \tag{$***$}$$

Let $s \in W$ be such that $\mathcal{M}, s \models \varphi$. The goal is to show that $\mathcal{M}, s \models K\varphi$. Let $t \in W$ be such that $sR_K t$. Given $s \in W$, by ($*$), choose $y \in W$ such that $sR_\Box y$ and the following condition holds:

$$\forall z \in W(yR_K z \to s = z). \tag{$\#$}$$

Since $y \in W$, by ($**$), choose $z \in W$ such that $yR_K z$. By ($\#$), $s = z$. From $s = z$ and $sR_K t$, it follows that $zR_K t$. From $yR_K z$ and $zR_K t$, by ($***$), we get $yR_K t$. However, since $yR_K t$, by ($\#$), we obtain $s = t$. Finally, $s = t$ and $\mathcal{M}, s \models \varphi$ imply $\mathcal{M}, t \models \varphi$. Thus, we have shown that $\mathcal{M}, s \models K\varphi$. This concludes the proof that $\mathbf{KD4}_K \otimes \mathbf{K}_\Box{+}(\mathtt{KP}) \models \varphi \to K\varphi$. Since the completeness of the logic $\mathbf{KD4}_K \otimes \mathbf{K}_\Box{+}(\mathtt{KP})$ can be

proven analogously to the completeness of the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{KP})$, we obtain $\mathbf{KD4}_K \otimes \mathbf{K}_\Box + (\mathtt{KP}) \vdash \varphi \to K\varphi$. ⊣

This result is significant because, since Hintikka's foundational works in epistemic logic, the logic **KD4** — or its extension **KD45** — has been regarded as the standard doxastic logic. Consequently, if the operator K is interpreted not as a knowledge operator but as a belief operator, we obtain a doxastic version of Fitch's paradox: the *believability paradox*.

Another significant benefit of the adopted method for analyzing the knowability paradox is that the obtained semantics allow us to identify other schemes that are provable under the given assumptions. Some of these schemes, such as the formulas of the form $\varphi \to K\varphi$, should not be provable in a system we consider an adequate epistemic-alethic logic. Below, we present several schemes that are provable in the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{KP})$.

PROPOSITION 5.5. 1. $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{KP}) \vdash K\varphi \to KK\varphi$,
2. $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{KP}) \vdash \neg K\varphi \to K\neg K\varphi$,
3. $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{KP}) \vdash \Box\varphi \to \varphi$,
4. $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{KP}) \vdash \Box\varphi \to K\varphi$,
5. $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{KP}) \vdash \langle K \rangle \varphi \to K\varphi$,
6. $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{KP}) \vdash \Box\varphi \to \Box K\varphi$.

The fact that the schemes listed in points 1, 2, and 3 are provable is not inherently problematic, even though these formulas are not derivable within the fusion $\mathbf{T}_K \otimes \mathbf{K}_\Box$ itself. However, the presence of 4, 5, and 6 is highly undesirable. Nevertheless, it should be observed that none of the formulas from 4–6 are provable in the fusion $\mathbf{T}_K \otimes \mathbf{K}_\Box$ itself, nor in any stronger fusion obtained from it by adding the standard modal axioms for both modal operators, such as $\mathbf{S5}_K \otimes \mathbf{S5}_\Box$. This can be shown by means of a simple countermodel $\mathcal{M} = (W, R_K, R_\Box, v)$ such that $W = \{s, t\}$, $R_K = \{(s,s),(t,t),(s,t),(t,s)\}$, $R_\Box = \{(s,s),(t,t)\}$, and $v(p) = \{s\}$. The model $\mathcal{M}$ is a model of the logic $\mathbf{S5}_K \otimes \mathbf{S5}_\Box$, and hence also of the weaker fusion $\mathbf{T}_K \otimes \mathbf{K}_\Box$.

The provability of formulas associated with the scheme in point 4 — a consequence of the inclusion $R_K \subseteq R_\Box$, which itself results from adding the axiom scheme (KP) to $\mathbf{T}_K \otimes \mathbf{K}_\Box$ — is not surprising. If it is provable that all truths are known, it should also be provable that all necessarily true truths are known. Particular attention should be given to the scheme $\Box\varphi \to \Box K\varphi$. Since $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{KP}) \vdash \varphi \to K\varphi$, then, by (RN), we get

$\mathbf{T}_K \otimes \mathbf{K}_\Box{+}(\mathtt{KP}) \vdash \Box(\varphi \to K\varphi)$. From this, applying the axiom ($\mathtt{K}_\Box$), we obtain $\mathbf{T}_K \otimes \mathbf{K}_\Box{+}(\mathtt{KP}) \vdash \Box\varphi \to \Box K\varphi$. Therefore, if the assumptions in this paper are justified, it follows that the principle of knowability implies the existence of truths that are necessarily known by someone.

6. Restricting the knowability principle

In the context of dynamic epistemic logics (see [46] for a survey), it is possible to model the phenomenon of an unsuccessful update. This phenomenon occurs when a true sentence becomes false due to its announcement. To illustrate, let's assume that agent A is waiting for a letter informing them of their admission to university. Agent B picks up the letter from the post office, reads its contents, and announces to A: "You don't know it, but you have been admitted to university". B conveys two pieces of information in his announcement: (i) A has been admitted to university; (ii) A does not know that A has been admitted to university. Assuming agent A considers B to be a reliable source of information, conveying piece of information (i) renders piece of information (ii) false. Therefore, the true sentence becomes false due to its announcement.

Using the formalism of dynamic epistemic logics, it can be demonstrated that the phenomenon of an unsuccessful update plays a key role in the emergence of the knowability paradox [see 44, 51]. This formalism also allows for defining an important class of formulas: successful formulas [see 45]. A formula is considered successful if, after its announcement, it retains its truth, meaning it cannot change its logical value during the verification process, which could lead to an unsuccessful update. In [2] is shown that restricting the formalization of the knowability principle in dynamic epistemic logic to formulas in the class of successful formulas effectively blocks the paradox. The results obtained support the hypothesis that the cause of Fitch's paradox is the phenomenon of unsuccessful update, and an adequate solution to the paradox should involve restricting the knowability principle to sentences that cannot lead to this phenomenon.

In standard epistemic-alethic logics, it is not possible to express that a given formula is successful. However, in epistemic-alethic systems, one can express that a certain formula is stable — where stability is understood as the preservation of the formula's truth in all possible states of the verification process. To express this property, it is necessary to

modify the interpretation of the alethic operators from the language $\mathcal{L}_{K,\Box}$. This reinterpretation is based on the proposal by Artemov and Protopopescu [1].

According to the standard interpretation, the operator $\Diamond$ represents logical possibility. We have adopted this interpretation to accurately reflect how Fitch and other authors understand the knowability principle. However, Artemov and Protopopescu argue that to adequately formalize the knowability of φ — especially from a verificationist standpoint — the interpretation of the alethic operators $\Diamond$ and $\Box$ should be modified. They propose viewing these operators as quantifiers over states of discovery. According to this view, the formula $\Diamond\varphi$ should be interpreted as "there exists a state of the discovery process in which φ holds", while the formula $\Box\varphi$ should be interpreted as "in all states of the discovery process, φ holds". A consequence of this reinterpretation is a different understanding of the accessibility relation $R_\Box$. The presence of the relation $R_\Box$ between elements of the model's universe — referred to as information states — indicates a potential direction in the verification process. If $sR_\Box t$ holds, it means that the discovery or verification process can lead from state s to state t.

As a result of this reinterpretation, we obtain the original system of dynamic epistemic logic, which not only provides a formal representation of knowledge but also models the phenomenon of knowledge change at various stages of the verification process. Furthermore, due to the modification in the interpretation of the alethic operators described above, there is a new way to understand the knowability principle:

- *If φ is true, then there exists a state of the discovery process in which φ is known.*

Of course, as demonstrated by Fitch's proof, if no restrictions are imposed on the formulas φ, the above thesis is false because the formula $p \wedge \neg Kp$ cannot be known in any state of the discovery process. This is because a proposition may change its logical value during the discovery process.

Nevertheless, if the alethic operators are understood as described above, it becomes possible to express within the language $\mathcal{L}_{K,\Box}$ that a formula φ is stable, meaning it retains its truth throughout the discovery process. More precisely, a formula φ is *stable* in a given model $\mathcal{M}$ if for any state s, it holds that $\mathcal{M}, s \models \varphi \to \Box\varphi$. Consequently, the following modification of (KP) can be made, which can be termed the principle of

stable knowability:

$$(\varphi \wedge (\varphi \to \Box\varphi)) \to \Diamond K\varphi.$$

Note that in any model where the accessibility relation used to interpret the operator $\Box$ is reflexive, the formulas $\varphi \to \Box\varphi$ and $\Box\varphi$ are equivalent. Because the authors assume that the relation $R_\Box$ is reflexive, the stable knowability principle can be equivalently expressed as follows:

$$\Box\varphi \to \Diamond K\varphi. \tag{SK}$$

The above formula scheme should be interpreted as follows:

- *If φ holds at all states of discovery, then there exists a state of the discovery process in which φ becomes known.*

Artemov and Protopopescu argue that this thesis represents the correct interpretation of the verificationist approach because it explicitly establishes the relationship between verification, knowledge, and truth[4]. They present an argument that, in their view, demonstrates that accepting the stable knowability principle does not imply the omniscience principle. Specifically, they provide a model of the logic $\mathbf{S5}_K \otimes \mathbf{T}_\Box$ and show that the formula $\Box\varphi \to \Diamond K\varphi$ is true in this model, while $\varphi \to K\varphi$ fails to hold in any state. However, the argument presented by Artemov and Protopopescu does not establish that, in the logic $\mathbf{S5}_K \otimes \mathbf{T}_\Box$+(SK), the scheme $\varphi \to K\varphi$ is not valid. The authors did not determine which first-order formula corresponds to the scheme $\Box\varphi \to \Diamond K\varphi$, and, therefore, did not construct a semantics for the logic $\mathbf{S5}_K \otimes \mathbf{T}_\Box$+(SK).[5]

We will demonstrate that it is not necessary to restrict (KP) to stable formulas to block the provability of formulas of the form $\varphi \to K\varphi$. Instead, it is sufficient to restrict (KP) to formulas that can be termed

[4] It is important to note that this thesis was considered in the literature before the publication of Artemov and Protopopescu's article. However, the authors who addressed it interpreted the operators $\Box$ and $\Diamond$ in the standard way. (SK) was considered by Burgess [7], who ultimately advocated for its temporal counterpart, and by Egré [16], who argued that adopting (SK) is the appropriate response to Fitch's paradox.

[5] It can be demonstrated, however, that the scheme $\Box\varphi \to \Diamond K\varphi$ corresponds to a first-order formula Ψ: $\forall x \exists y(xR_\Box y \wedge \forall z(yR_K z \to xR_\Box z))$. It can also be proven that the logic $\mathbf{S5}_K \otimes \mathbf{T}_\Box$+(SK) is sound and complete with respect to the class of all frames where the relation R_K is an equivalence relation, the relation $R_\Box$ is reflexive, and Ψ holds. Consequently, a countermodel for $\varphi \to K\varphi$ could be, for example, the model $\mathcal{M} = (W, R_K, R_\Box, v)$, such that $W = \{s,t\}$, $R_K = R_\Box = \{(s,s),(t,t),(s,t),(t,s)\}$, $v(p) = \{s\}$.

locally stable formulas. A formula φ is considered *locally stable* in a given model $\mathcal{M}$ if for any state s, it holds that $\mathcal{M}, s \models \Diamond\Box\varphi$. Consequently, we obtain the following thesis, referred to as the principle of locally stable knowability:

$$(\varphi \wedge \Diamond\Box\varphi) \to \Diamond K\varphi. \tag{LSK}$$

According to this principle, for a true formula φ to be considered knowable in state s, it suffices that there exists a state t such that the discovery process can lead from s to t, with φ remaining true in all states reachable from t via the verification process. Restricting (KP) to locally stable formulas is, of course, less rigorous than restricting it to stable formulas. Under (LSK), φ can be knowable in state s even if it is false in some states reachable from s via $R_\Box$, whereas (SK) excludes this possibility. However, the restriction imposed by (LSK) is sufficient to eliminate all those truths that become false as a result of the very act of discovery.

Next, we will investigate the logic obtained by extending the base logic with (LSK) as an additional axiom scheme.

DEFINITION 6.1. The *logic* $\mathbf{T}_K \otimes \mathbf{K}_\Box$+(LSK) is an extension of the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box$ by the addition of the axiom scheme (LSK).

Remark 6.1. The scheme $(\varphi \wedge \Diamond\Box\varphi) \to \Diamond K\varphi$ is a Sahlqvist scheme.

By Remark 6.1, and according to Sahlqvist's correspondence theorem (Theorem 4.1) for the language $\mathcal{L}_{K,\Box}$, there is a first-order formula corresponding to $(\varphi \wedge \Diamond\Box\varphi) \to \Diamond K\varphi$. In Appendix B, we present a version of the Sahlqvist-van Benthem algorithm applicable to (LSK), followed by a proof of the following proposition:

PROPOSITION 6.1 (B.1). *Let $\mathcal{F} = (W, R_K, R_\Box)$ be an epistemic-alethic frame. Then, the following are equivalent:*

(a) $\mathcal{F} \models (\varphi \wedge \Diamond\Box\varphi) \to \Diamond K\varphi$,
(b) $\forall x, y \in W(xR_\Box y \to \exists a \in W(xR_\Box a \wedge \forall b \in W(aR_K b \to (yR_\Box b \vee b = x))))$.

Based on these results, we can define a frame for the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box$+(LSK) as follows:

DEFINITION 6.2. An *epistemic-alethic frame for* $\mathbf{T}_K \otimes \mathbf{K}_\Box$+(LSK) (in short, $\mathbf{T}_K \otimes \mathbf{K}_\Box$+(LSK)*-frame*) is a frame $\mathcal{F} = (W, R_K, R_\Box)$, where

- $W \neq \emptyset$,
- $R_K \subseteq W \times W$ and $R_\Box \subseteq W \times W$ are relations satisfying the following conditions:

(i) $\forall x \in W(xR_Kx)$,
(ii) $\forall x, y \in W(xR_\Box y \to \exists a \in W(xR_\Box a \wedge \forall b \in W(aR_Kb \to (yR_\Box b \vee b = x))))$.

The application of Theorem 4.2 and the established propositions and theorems is sufficient to prove the soundness and completeness of the $\mathbf{T}_K \otimes \mathbf{K}_\Box$+(LSK) logic.

THEOREM 6.1. *Logic* $\mathbf{T}_K \otimes \mathbf{K}_\Box$+(LSK) *is sound and complete with respect to the class of all* $\mathbf{T}_K \otimes \mathbf{K}_\Box$+(LSK)*-frames.*

PROOF. According to Definition 6.1, the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box$+(LSK) is obtained by adding the axiom scheme $(\varphi \wedge \Diamond\Box\varphi) \to \Diamond K\varphi$ to the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box$. By Proposition 6.1, we know that this scheme defines the class of frames $\mathcal{F} = (W, R_K, R_\Box)$, such that $\forall x, y \in W(xR_\Box y \to \exists a \in W(xR_\Box a \wedge \forall b \in W(aR_Kb \to (yR_\Box b \vee b = x))))$. Furthermore, from Theorem 3.1, we know that the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box$ is sound and complete with respect to the class of frames $\mathcal{F} = (W, R_K, R_\Box)$, such that $\forall x \in W(xR_Kx)$. Therefore, by Theorem 4.2 and Remark 6.1, we conclude that the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box$+(LSK) is sound and complete with respect to the class of all $\mathbf{T}_K \otimes \mathbf{K}_\Box$+(LSK)-frames. ⊣

Referring to the above theorem, we can demonstrate the difference between the logics $\mathbf{T}_K \otimes \mathbf{K}_\Box$+(KP) and $\mathbf{T}_K \otimes \mathbf{K}_\Box$+(LSK) in terms of the provability of paradoxical formulas. A key observation in the context of Fitch's paradox is that the standard formalization of the omniscience principle is not provable in the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box$+(LSK).

PROPOSITION 6.2. $\mathbf{T}_K \otimes \mathbf{K}_\Box$+(LSK) $\nvdash \varphi \to K\varphi$.

PROOF. Consider the model $\mathcal{M} = (W, R_K, R_\Box, v)$ such that $W = \{s, t\}$, $R_K = \{(s, s), (t, t), (s, t)\}$, $R_\Box = \emptyset$, and $v(p) = \{s\}$. Since sR_Kt and $\mathcal{M}, t \not\Vdash p$, we have $\mathcal{M}, s \not\Vdash Kp$. Thus, $\mathcal{M}, s \not\Vdash p \to Kp$. Therefore, $\mathcal{M} \not\Vdash p \to Kp$. Note that $\mathcal{M}$ is a $\mathbf{T}_K \otimes \mathbf{K}_\Box$+(LSK)-model: the relation R_K is reflexive, and since $R_\Box = \emptyset$, the condition $\forall x, y \in W(xR_\Box y \to \exists a \in W(xR_\Box a \wedge \forall b \in W(aR_Kb \to (yR_\Box b \vee b = x))))$ holds. Consequently, we obtain $\mathbf{T}_K \otimes \mathbf{K}_\Box$+(LSK) $\not\Vdash \varphi \to K\varphi$, and thus, by Theorem 6.1, we conclude $\mathbf{T}_K \otimes \mathbf{K}_\Box$+(LSK) $\nvdash \varphi \to K\varphi$. ⊣

By restricting the knowability principle to locally stable formulas, we successfully eliminate the main negative consequence of adopting (KP). Notably, in the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box$+(LSK), formulas of the form $(\varphi \wedge \Diamond\Box\varphi) \to$

$K\varphi$ are not provable, implying that not all locally stable truths are known.

PROPOSITION 6.3. $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{LSK}) \nvdash (\varphi \wedge \Diamond\Box\varphi) \to K\varphi$.

PROOF. Consider the model $\mathcal{M} = (W, R_K, R_\Box, v)$ such that $W = \{s, t, u\}$, $R_K = \{(s,s), (t,t), (u,u), (s,t)\}$, $R_\Box = \{(s,u), (u,u)\}$, and $v(p) = \{s, u\}$. Since sR_Kt and $\mathcal{M}, t \not\models p$, we have $\mathcal{M}, s \not\models Kp$. Due to the choice of $R_\Box$ and the fact that $\mathcal{M}, u \models p$, we obtain $\mathcal{M}, u \models \Box p$. Given that $sR_\Box u$ and $\mathcal{M}, u \models \Box p$, it follows that $\mathcal{M}, s \models \Diamond\Box p$. Thus, $\mathcal{M}, s \models p \wedge \Diamond\Box p$ and $\mathcal{M}, s \not\models Kp$, which means that $\mathcal{M}, s \not\models (p \wedge \Diamond\Box p) \to Kp$. Therefore, $\mathcal{M} \not\models (p \wedge \Diamond\Box p) \to Kp$. Since $\mathcal{M}$ is a $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{LSK})$-model, we obtain $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{LSK}) \not\models (\varphi \wedge \Diamond\Box\varphi) \to K\varphi$, which implies $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{LSK}) \nvdash (\varphi \wedge \Diamond\Box\varphi) \to K\varphi$ by Theorem 6.1. ⊣

In Proposition 5.5, we highlighted other schemes beyond $\varphi \to K\varphi$ that are provable in $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{KP})$. Formulas of these schemes should not be regarded as theorems of a system that we would consider to be an adequate epistemic-alethic logic. Now, we can demonstrate that these schemes are not provable in the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{LSK})$.

PROPOSITION 6.4. 1. $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{LSK}) \nvdash \Box\varphi \to K\varphi$,
2. $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{LSK}) \nvdash \langle K \rangle \varphi \to K\varphi$,
3. $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{LSK}) \nvdash \Box\varphi \to \Box K\varphi$.

PROOF. It can be easily verified that a countermodel for items 1 and 2 is the model provided in the proof of Proposition 6.2. We now focus on proving point 3. Consider the model $\mathcal{M} = (W, R_K, R_\Box, v)$ such that $W = \{s, t, u\}$, $R_K = \{(s,s), (t,t), (u,u), (t,u)\}$, $R_\Box = \{(s,s), (s,t)\}$, and $v(p) = \{s, t\}$. Because $R_\Box = \{(s,s), (s,t)\}$, and given that $\mathcal{M}, s \models p$ and $\mathcal{M}, t \models p$, it follows that $\mathcal{M}, s \models \Box p$. Notably, since tR_Ku and $\mathcal{M}, u \not\models p$, we have $\mathcal{M}, t \not\models Kp$. Consequently, given $sR_\Box t$, we obtain $\mathcal{M}, s \not\models \Box Kp$. Thus, we have $\mathcal{M}, s \models \Box p$ and $\mathcal{M}, s \not\models \Box Kp$, which means $\mathcal{M}, s \not\models \Box p \to \Box Kp$. Therefore, $\mathcal{M} \not\models \Box p \to \Box Kp$. Since $\mathcal{M}$ is a $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{LSK})$-model, we obtain $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{LSK}) \not\models \Box\varphi \to \Box K\varphi$, which, by Theorem 6.1, is equivalent to $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{LSK}) \nvdash \Box\varphi \to \Box K\varphi$. ⊣

Given the argument presented, the following objection may arise: although replacing (KP) with the weaker axiom scheme (LSK) has blocked the provability of the omniscience principle and other paradoxical formulas, the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{LSK})$ does not include all possible axioms for the operators K and $\Box$. Thus, it remains possible that accepting additional

epistemic and alethic axioms may cause these paradoxical formulas to be provable in an extended version of the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{LSK})$. To counter this objection, we will demonstrate that these formulas are also not provable in the system $\mathbf{S5}_K \otimes \mathbf{S5}_\Box + (\mathtt{LSK})$.

The logic $\mathbf{S5}_K \otimes \mathbf{S5}_\Box + (\mathtt{LSK})$ is sound and complete with respect to the class of all epistemic-alethic frames where the relations R_K and $R_\Box$ are equivalence relations, and the condition $\forall x, y(xR_\Box y \to \exists a(xR_\Box a \wedge \forall b(aR_K b \to (yR_\Box b \vee b = x))))$ holds. Based on this, we can establish the following proposition:

PROPOSITION 6.5. 1. $\mathbf{S5}_K \otimes \mathbf{S5}_\Box + (\mathtt{LSK}) \nvdash \varphi \to K\varphi$,
2. $\mathbf{S5}_K \otimes \mathbf{S5}_\Box + (\mathtt{LSK}) \nvdash (\varphi \wedge \Diamond\Box\varphi) \to K\varphi$,
3. $\mathbf{S5}_K \otimes \mathbf{S5}_\Box + (\mathtt{LSK}) \nvdash \Box\varphi \to K\varphi$,
4. $\mathbf{S5}_K \otimes \mathbf{S5}_\Box + (\mathtt{LSK}) \nvdash \langle K\rangle \varphi \to K\varphi$,
5. $\mathbf{S5}_K \otimes \mathbf{S5}_\Box + (\mathtt{LSK}) \nvdash \Box\varphi \to \Box K\varphi$,

PROOF. We will prove point 2. The countermodel provided for the scheme $(\varphi \wedge \Diamond\Box\varphi) \to K\varphi$ will also be a countermodel for all other schemes. Consider the model $\mathcal{M} = (W, R_K, R_\Box, v)$ such that $W = \{s, t, u, w\}$, $R_K = \{(s,s),(t,t),(u,u),(w,w),(s,t),(t,s)\}$, $R_\Box = \{(s,s),(t,t),(u,u),(w,w),(s,u),(u,s),(t,w),(w,t)\}$, and $v(p) = \{s, u\}$. This model is a $\mathbf{S5}_K \otimes \mathbf{S5}_\Box + (\mathtt{LSK})$-model. Since $sR_K t$ and $\mathcal{M}, t \not\models p$, we have $\mathcal{M}, s \not\models Kp$. From the choice of $R_\Box$, given that $\mathcal{M}, s \models p$ and $\mathcal{M}, u \models p$, it follows that $\mathcal{M}, u \models \Box p$. Furthermore, since $\mathcal{M}, u \models \Box p$ and $sR_\Box u$, we obtain $\mathcal{M}, s \models \Diamond\Box p$. Thus, we have $\mathcal{M}, s \models p \wedge \Diamond\Box p$ and $\mathcal{M}, s \not\models Kp$, i.e., $\mathcal{M}, s \not\models (p \wedge \Diamond\Box p) \to Kp$. Consequently, $\mathcal{M} \not\models (p \wedge \Diamond\Box p) \to Kp$. Hence, we obtain $\mathbf{S5}_K \otimes \mathbf{S5}_\Box + (\mathtt{LSK}) \not\models (\varphi \wedge \Diamond\Box\varphi) \to K\varphi$, which, by the soundness of the logic $\mathbf{S5}_K \otimes \mathbf{S5}_\Box + (\mathtt{LSK})$, implies $\mathbf{S5}_K \otimes \mathbf{S5}_\Box + (\mathtt{LSK}) \nvdash (\varphi \wedge \Diamond\Box\varphi) \to K\varphi$. ⊣

Consequently, even strengthening the base logic to the fusion of the systems $\mathbf{T}_K$ and $\mathbf{K}_\Box$ does not result in the provability of paradoxical formulas when the knowability principle is restricted to locally stable formulas.

7. Conclusion

The key contributions of this study are as follows:

- Demonstration that (OP) is not provable in any natural fusion of epistemic and alethic logics unless (KP) is assumed.

- Formulation of a new semantic proof of the knowability paradox, showing that the problematic conclusion can be reached independently of Fitch's original proof, emphasizing that the knowability principle itself is responsible for the paradox.
- Development of a semantic explanation for the provability of (OP), demonstrating that it arises from the fact that models of the logic $\mathbf{T}_K \otimes \mathbf{K}_\Box + (\mathtt{KP})$ are, in a sense, abnormal, meaning that no world possesses an epistemic alternative other than itself.
- Detection of paradoxical schemes other than (OP), provable due to the adoption of (KP) in conjunction with the other assumptions present in Fitch's proof.
- Identification of the root cause of the paradox, namely the possibility of a formula changing its truth value during the process of verification.
- Demonstration that the paradox can be blocked in some dynamic epistemic logics by restricting the knowability principle to locally stable formulas (i.e., formulas that retain their truth value from a specific stage of the verification process).
- Demonstration that the proposed restriction prevents the provability of other paradoxical formulas that were theorems of the logic where (KP) was assumed.

The boundary between the principles related to the knowability principle that lead to the provability of paradoxical formulas and those that avoid such undesirable consequences can be termed "the Fitch boundary". The method of fusing modal logics proves to be a valuable tool in drawing this boundary. Although the restriction on the knowability principle proposed in this paper is less stringent than other suggestions found in the literature, it is possible that even less restrictive limitations may be identified in the future, thereby bringing us closer to the Fitch boundary while remaining on its non-paradoxical side. However, as a result of the conceptual framework for the knowability problem presented here, this issue takes on a more technical character, providing us with new formal tools for its analysis.

Acknowledgments. The author would like to thank the anonymous reviewers for their several useful comments, which helped improve the paper's presentation. This research was supported by the National Science Centre, Poland, grant no. 2022/45/N/HS1/01080.

References

[1] Artemov, S., and T. Protopopescu, "Discovering knowability: A semantic analysis", *Synthese*, 190 (2013): 3349–3376. DOI: 10.1007/s11229-012-0168-x

[2] Balbiani, P., A. Baltag, H. van Ditmarsch, A. Herzig, T. Hoshi, and T. De Lima, "Knowable as known after an announcement", *Review of Symbolic Logic*, 1, 3 (2008): 305–334. DOI: 10.1017/S1755020308080210

[3] Beall, J. C., "Knowability and possible epistemic oddities", pages 105–125 in J. Salerno (ed.), *New Essays on the Knowability Paradox*, New York: Oxford University Press, 2009. DOI: 10.1093/acprof:oso/9780199285495.003.0009

[4] Blackburn, P., M. de Rijke, and Y. Venema, *Modal Logic*, Cambridge University Press, Cambridge, 2001. DOI: 10.1017/CBO9781107050884

[5] Brogaard, B., and J. Salerno, "Clues to the paradoxes of knowability: reply to Dummett and Tennant", *Analysis*, 62, 2 (2002): 143–150. DOI: 10.1093/analys/62.2.143

[6] Bueno, O., "Fitch's paradox and the philosophy of mathematics", pages 252–280 in J. Salerno (ed.), *New Essays on the Knowability Paradox*, New York: Oxford University Press, 2009. DOI: 10.1093/acprof:oso/9780199285495.003.0017

[7] Burgess, J., "Can truth out?", pages 147–162 in J. Salerno (ed.), *New Essays on the Knowability Paradox*, New York: Oxford University Press, 2009. DOI: 10.1017/CBO9780511487347.012

[8] Church, A., "Referee reports on Fitch's *A definition of value*", pages 13–20 in J. Salerno (ed.), *New Essays on the Knowability Paradox*, New York: Oxford University Press, 2009. DOI: 10.1093/acprof:oso/9780199285495.003.0002

[9] Costa-Leite, A., "Fusions of modal logics and Fitch's paradox", *Croatian Journal of Philosophy*, 6, 2 (2006): 281–290. DOI: 10.5840/croatjphil20066220

[10] Costa-Leite, A., *Interactions of Metaphysical and Epistemic Concepts*, PhD Thesis, University of Neuchatel, Switzerland, 2007.

[11] DeVidi, D., and G. Solomon, "Knowability and intuitionistic logic", *Philosophia*, 28 (2001): 319–334. DOI: 10.1007/BF02379783

[12] Douven, I., "A principled solution to Fitch's paradox", *Erkenntnis*, 62, 1 (2005): 47–69. DOI: 10.1007/s10670-004-9563-0

[13] Dummett, M., "Victor's error", *Analysis*, 61, 1 (2001): 1–2. DOI: 10.1093/analys/61.1.1

[14] Dummett, M., "Fitch's paradox of knowability", pages 51–52 in J. Salerno (ed.), *New Essays on the Knowability Paradox*, New York: Oxford University Press, 2009. DOI: 10.1093/acprof:oso/9780199285495.003.0005

[15] Edgington, D., "The paradox of knowability", *Mind*, 94, 376 (1985): 557–568. DOI: 10.1093/mind/xciv.376.557

[16] Égré, P., "Le paradoxe de Fitch dans l'œil du positiviste: y a-t-il des vérités inconnaissables", *Les Études Philosophiques*, 1 (2008): 71–95. DOI: 10.3917/leph.081.0071

[17] Fitch, F. B., "A logical analysis of some value concepts", *Journal of Symbolic Logic*, 28, 2 (1963): 135–142. DOI: 10.2307/2271594

[18] Fine, K., and G. Schurz, "Transfer theorems for multimodal logics", pages 169–213 in B. J. Copeland (ed.), *Logic and Reality: Essays on the Legacy of Arthur Prior*, Oxford: Cambridge University Press, 1996. DOI: 10.1093/oso/9780198240600.003.0009

[19] Gabbay, D. M, A. Kurucz, F. Wolter, and M. Zakharyaschev, *Many-Dimensional Modal Logics: Theory and Applications*, Studies in Logic and the Foundations of Mathematics, vol. 148, Elsevier, Amsterdam, 2003.

[20] Hand, M., and J. L. Kvanvig, "Tennant on knowability", *Australasian Journal of Philosophy*, 77, 4 (1999): 422–428. DOI: 10.1080/00048409912349191

[21] Hintikka, J., *Knowledge and Belief*, Cornell University Press, Ithaca, 1962.

[22] Hintikka, J., "On the logic of perception", pages 151–183 in J. Hintikka, *Models for Modalities: Selected Essays*, Dordrecht: Springer, 1969. DOI: 10.1007/978-94-010-1711-4_8

[23] Jago, M., "Closure on knowability", *Analysis*, 70, 4 (2010): 648–659. DOI: 10.1093/analys/anq067

[24] Kinkaid, J., "Phenomenology, anti-realism, and the knowability paradox", *European Journal of Philosophy*, 30, 3 (2022): 1010–1027. DOI: 10.1111/ejop.12762

[25] Kracht, M., and Wolter, F., "Properties of independently axiomatizable bimodal logics", *Journal of Symbolic Logic*, 56, 4 (1991): 1469–1485. DOI: 10.2307/2275487

[26] Kripke, S., "A completeness theorem in modal logic", *Journal of Symbolic Logic*, 24, 1 (1959): 1–14. DOI: 10.2307/2964568

[27] Kripke, S., "Semantical analysis of modal logic I. Normal modal propositional calculi", *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 9, 5–6 (1963): 67–96. DOI: 10.1002/malq.19630090502

[28] Kurucz, A., "Combining modal logics", pages 869–924 in P. Blackburn, J. van Benthem, and F. Wolter (eds.), *Handbook of Modal Logic*, Amsterdam: Elsevier, 2007. DOI: 10.1016/S1570-2464(07)80018-8

[29] Kvanvig, J., "The knowability paradox and the prospects for anti-realism", *Nous*, 29, 4 (1995): 481–499. DOI: 10.2307/2216283

[30] Łukowski, P., *Paradoxes*, Springer Netherlands, Dordrecht, 2011. DOI: 10.1007/978-94-007-1476-2

[31] Mackie, J. L., "Truth and knowability", *Analysis*, 40, 2 (1980): 90–92. DOI: 10.1093/analys/40.2.90

[32] Nozick, R., *Philosophical Explanations*, Harvard University Press, Cambridge, 1981.

[33] Percival, P., "Fitch and intuitionistic knowability", *Analysis*, 50, 3 (1990): 182–187. DOI: 10.1093/analys/50.3.182

[34] Priest, G., "Beyond the limits of knowledge", pages 93–104 in J. Salerno (ed.), *New Essays on the Knowability Paradox*, New York: Oxford University Press, 2009. DOI: 10.1093/acprof:oso/9780199285495.003.0008

[35] Quine, W. v. O., "The limits of knowledge", pages 59–67 in *The Ways of Paradox and Other Essays*, New York: Random House, 1976.

[36] Restall, G., "Not every truth can be known (at least, not all at once)", pages 339–354 in J. Salerno (ed.), *New Essays on the Knowability Paradox*, New York: Oxford University Press, 2009. DOI: 10.1093/acprof:oso/9780199285495.003.0022

[37] Sahlqvist, H., "Correspondence and completeness in the first and second order semantics for modal logic", pages 110–143 in S. Kanger (ed.), *Proceedings of the Third Scandinavian Logic Symposium. Uppsala 1973*, Amsterdam: North Holland Publishing, 1973. DOI: 10.1016/S0049-237X(08)70728-6

[38] Salerno, J., "Introduction", pages 1–10 in J. Salerno (ed.), *New Essays on the Knowability Paradox*, New York: Oxford University Press, 2009. DOI: 10.1093/acprof:oso/9780199285495.003.0001

[39] Salerno, J., "Knowability noir: 1945–1963", pages 29–48 in J. Salerno (ed.), *New Essays on the Knowability Paradox*, New York: Oxford University Press, 2009. DOI: 10.1093/acprof:oso/9780199285495.003.0004

[40] Tennant, N., *The Taming of the True*, Oxford University Press, Oxford, 2002. DOI: 10.1093/acprof:oso/9780199251605.001.0001

[41] Thomason, S. K., "Independent propositional modal logics", *Studia Logica*, 39 (1980): 143–144. DOI: 10.1007/BF00370317

[42] van Benthem, J., "Modal reduction principles", *Journal of Symbolic Logic*, 41, 2 (1976): 301–312. DOI: 10.2307/2272228

[43] van Benthem, J., *Modal Correspondence Theory*, PhD Thesis, Mathematisch Instituut & Instituut voor Grondslagenonderzoek, University of Amsterdam, 1976.

[44] van Benthem, J., "What one may come to know", *Analysis*, 64, 2 (2004): 95–105. DOI: 10.1093/analys/64.2.95

[45] van Ditmarsch, H., B. and Kooi, "The secret of my success", *Synthese*, 151 (2006): 201–232. DOI: 10.1007/s11229-005-3384-9

[46] van Ditmarsch, H., van W. der Hoek, W., and B. Kooi, *Dynamic Epistemic Logic*, Springer Netherlands, Dordrecht, 2007. DOI: 10.1007/978-1-4020-5839-4

[47] Wansing, H., "Diamonds are a philosopher's best friends", *Journal of Philosophical Logic*, 31, 6 (2002): 591–612. DOI: 10.1023/A:1021256513220

[48] Williamson, T., "Knowability and constructivism", *Philosophical Quarterly*, 38, 153 (1988): 422–432. DOI: 10.2307/2219707

[49] Williamson, T., "On intuitionistic modal epistemic logic", *Journal of Philosophical Logic*, 21, 1 (1992): 63–89. DOI: 10.1007/BF00126497

[50] Williamson, T., "Verificationism and non-distributive knowledge", *Australasian Journal of Philosophy*, 71, 1 (1993): 78–86. DOI: 10.1080/00048409312345072

[51] Wójcik, A., "The knowability paradox and unsuccessful updates", *Studies in Logic, Grammar and Rhetoric*, 62, 1 (2020): 53–71. DOI: 10.2478/slgr-2020-0013

Appendices

A. Very simple Sahlqvist formulas of $\mathcal{L}_{K,\Box}$

DEFINITION A.1. Let x be a first-order variable. The *standard translation* ST_x that converts formulas from the language $\mathcal{L}_{K,\Box}$ to first-order formulas is defined as follows:

$$\begin{aligned} ST_x(p) &= P(x), \\ ST_x(\neg\varphi) &= \neg ST_x(\varphi), \\ ST_x(\varphi \to \psi) &= ST_x(\varphi) \to ST_x(\psi), \\ ST_x(K\varphi) &= \forall y(xR_K y \to ST_y(\varphi)), \\ ST_x(\Box\varphi) &= \forall y(xR_\Box y \to ST_y(\varphi)), \end{aligned}$$

where $p \in Var$ and y is a new variable.

The *standard second-order translation* of the formula $\varphi \in \Gamma_{\mathcal{L}_{K,\Box}}$ is defined as the formula $\forall P_1 \ldots \forall P_k(ST_x(\varphi))$, where $P_1, \ldots, P_k$ are all unary predicates occurring in the formula $ST_x(\varphi)$.

To obtain the first-order formula corresponding to the axiom scheme (KP), we will use a simplified version of the Sahlqvist-van Benthem algorithm, characterized similarly to that in [4]. This algorithm enables us to obtain the first-order formula corresponding to the so-called very simple Sahlqvist formula, which for the language $\mathcal{L}_{K,\Box}$ is defined as follows:

DEFINITION A.2. A *very simple Sahlqvist antecedent* is a formula constructed from $\top$, $\bot$, and propositional variables, using only $\wedge$, $\Diamond$, and $\langle K\rangle$. A *very simple Sahlqvist formula* is an implication $\varphi \to \psi$, where φ is a very simple Sahlqvist antecedent and ψ is a positive formula.

It is clear that the formula $p \to \Diamond Kp$ is a very simple Sahlqvist formula and that every very simple Sahlqvist formula is a Sahlqvist formula.

ALGORITHM A.1 (Sahlqvist-van Benthem algorithm for very simple Sahlqvist formulas). Let $\nabla \in \{K, \Box\}$. If $\varphi \to \psi$ is a very simple Sahlqvist formula, then the corresponding first-order formula can be obtained by following these steps:

Step 1 (Standard second-order translation). The standard second-order translation of the formula $\varphi \to \psi$ results in a formula of the form

$$\forall P_1 \ldots \forall P_k(ST_x(\varphi) \to ST_x(\psi)).$$

The aim of the following steps is to eliminate the second-order quantifiers from the above formula to yield its first-order equivalent.

Step 2 (Formula transformation). Transform the second-order formula obtained in Step 1 using the equivalence

$$(\exists x_i \alpha(x_i) \wedge \beta) \leftrightarrow \exists x_i(\alpha(x_i) \wedge \beta),$$

and then the equivalence

$$(\exists x_i \alpha(x_i) \to \beta) \leftrightarrow \forall x_i(\alpha(x_i) \to \beta),$$

in order to move the first-order quantifiers in the antecedent $ST_x(\varphi)$ to the front of the implication. This procedure yields a formula of the form

$$\forall P_1 \ldots \forall P_k \forall x_1 \ldots \forall x_m((REL \wedge AT) \to ST_x(\psi)),$$

where REL is a conjunction of atomic formulas of the form $x_i R_\nabla x_j$, corresponding to occurrences of existential modal operators, and AT is

a conjunction of predicate symbols corresponding to propositional variables.

Step 3 (Establishing instances). Let P_i be a predicate occurring in the formula obtained in Step 2, and let $P_i(x_{i_1}), \ldots, P_i(x_{i_n})$ be all instances of the predicate variable P_i from AT. We define the substitution σ as follows:

$$\sigma(P_i) = \lambda u.(u = x_{i_1} \vee \ldots \vee u = x_{i_n}),$$

where the symbol λ denotes a lambda function as in [4].

Step 4 (Instantiating). Using the substitution defined in Step 3, we obtain the formula

$$[\sigma(P_1)/P_1, \ldots, \sigma(P_k)/P_k]\, \forall x_1 \ldots \forall x_m((REL \wedge AT) \rightarrow ST_x(\psi)).$$

The resulting formula is a first-order formula.

We will use the above algorithm to obtain a first-order formula corresponding to the formula $p \rightarrow \Diamond Kp$. In Step 1, by applying the standard second-order translation, we obtain the following formula:

$$\forall P(P(x) \rightarrow \exists y(xR_{\Box}y \wedge \forall z(yR_K z \rightarrow P(z)))).$$

In Step 2, we do not make any rearrangement. Note that AT is simply $P(x)$. Thus, in Step 3, we determine the following substitution:

$$\sigma(P) = \lambda u.u = x.$$

According to Step 4 of the algorithm, we carry out the substitution, obtaining the following formula:

$$x = x \rightarrow \exists y(xR_{\Box}y \wedge \forall z(yR_K z \rightarrow x = z)).$$

By eliminating the equality, we finally obtain the first-order formula:

$$\exists y(xR_{\Box}y \wedge \forall z(yR_K z \rightarrow x = z)).$$

PROPOSITION A.1 (4.1). *Let $\mathcal{F} = (W, R_K, R_{\Box})$ be an epistemic-alethic frame. Then, the following are equivalent*:

(a) $\mathcal{F} \models \varphi \rightarrow \Diamond K\varphi$,
(b) $\forall x \in W\ \exists y \in W(xR_{\Box}y \wedge \forall z \in W(yR_K z \rightarrow x = z))$.

PROOF. "(a) ⇒ (b)" We will prove this by contraposition. Let $\mathcal{F} = (W, R_K, R_\Box)$ be an epistemic-alethic frame such that

$$\exists x \in W\ \forall y \in W(xR_\Box y \to \exists z \in W(yR_K z \wedge x \neq z)). \tag{$*$}$$

We will show that for a certain formula φ and a certain model $\mathcal{M}$ based on $\mathcal{F}$, it holds that $\mathcal{M} \not\models \varphi \to \Diamond K\varphi$. Based on ($*$), choose $s \in W$ such that the following condition holds:

$$\forall y \in W(sR_\Box y \to \exists z \in W(yR_K z \wedge s \neq z)). \tag{$**$}$$

Consider the model $\mathcal{M} = (W, R_K, R_\Box, v)$ based on $\mathcal{F}$, such that $v(p) = \{s\}$. The goal is to show that $\mathcal{M}, s \not\models p \to \Diamond Kp$. Since $\mathcal{M}, s \models p$, it suffices to show that $\mathcal{M}, s \not\models \Diamond Kp$ (i.e., for all $t \in W$: if $sR_\Box t$, then $\mathcal{M}, t \not\models Kp$). Let $t \in W$ be such that $sR_\Box t$. We will prove that $\mathcal{M}, t \not\models Kp$. Since $sR_\Box t$, based on ($**$), choose $z \in W$ such that $tR_K z$ and $s \neq z$. From $s \neq z$ and $v(p) = \{s\}$, it follows that $\mathcal{M}, z \not\models p$. From $tR_K z$ and $\mathcal{M}, z \not\models p$, it follows that $\mathcal{M}, t \not\models Kp$. Thus, we have shown that $\mathcal{M}, s \not\models \Diamond Kp$. Consequently, we obtain $\mathcal{M}, s \not\models p \to \Diamond Kp$. Therefore, $\mathcal{M} \not\models p \to \Diamond Kp$. Since the model $\mathcal{M}$ is based on $\mathcal{F}$, this concludes the proof that $\mathcal{F} \not\models \varphi \to \Diamond K\varphi$.

"(b) ⇒ (a)" Let $\mathcal{F} = (W, R_K, R_\Box)$ be an epistemic-alethic frame such that

$$\forall x \in W\ \exists y \in W(xR_\Box y \wedge \forall z \in W(yR_K z \to x = z)). \tag{$\dagger$}$$

We will show that for any model $\mathcal{M}$ based on $\mathcal{F}$, it holds that $\mathcal{M} \models \varphi \to \Diamond K\varphi$. Let $\mathcal{M} = (W, R_K, R_\Box, v)$ be a model based on $\mathcal{F}$, and let $s \in W$ be arbitrary. Assume that $\mathcal{M}, s \models \varphi$. The goal is to show that $\mathcal{M}, s \models \Diamond K\varphi$. Based on ($\dagger$), choose $y \in W$ such that $sR_\Box y$ and the following condition holds:

$$\forall z \in W(yR_K z \to s = z). \tag{$\ddagger$}$$

We will prove that $\mathcal{M}, y \models K\varphi$. Let $z \in W$ be such that $yR_K z$. By ($\ddagger$), it follows that $s = z$. Since $s = z$ and $\mathcal{M}, s \models \varphi$, we obtain $\mathcal{M}, z \models \varphi$. Hence, $\mathcal{M}, y \models K\varphi$. From $\mathcal{M}, y \models K\varphi$ and $sR_\Box y$, it follows that $\mathcal{M}, s \models \Diamond K\varphi$. Consequently, we obtain $\mathcal{M}, s \models \varphi \to \Diamond K\varphi$. Therefore, $\mathcal{M} \models \varphi \to \Diamond K\varphi$. Since the model $\mathcal{M}$ is based on $\mathcal{F}$, this concludes the proof that $\mathcal{F} \models \varphi \to \Diamond K\varphi$. ⊣

B. Simple Sahlqvist formulas of $\mathcal{L}_{K,\Box}$

The difficulty in obtaining a first-order formula corresponding to the formula $(p \wedge \Diamond\Box p) \to \Diamond Kp$ lies in the fact that $p \wedge \Diamond\Box p$ is not a very simple Sahlqvist antecedent, and consequently, the entire formula is not a very simple Sahlqvist formula. Therefore, the algorithm presented in Appendix A cannot be applied in this case. We will define a new class of formulas — the class of simple Sahlqvist formulas — and extend Algorithm A.1. This extension will allow us to obtain a first-order formula corresponding to any simple Sahlqvist formula.

DEFINITION B.1. A *simple Sahlqvist antecedent* is a formula constructed from $\top$, $\bot$, and boxed atoms, using only $\wedge$, $\Diamond$, and $\langle K\rangle$. A *simple Sahlqvist formula* is an implication $\varphi \to \psi$, where φ is a simple Sahlqvist antecedent and ψ is a positive formula.

Note that the formula $(p\wedge\Diamond\Box p) \to \Diamond Kp$ is a simple Sahlqvist formula. This allows us to apply the following algorithm to it.

ALGORITHM B.1 (Sahlqvist-van Benthem algorithm for simple Sahlqvist formulas). Let $\nabla \in \{K, \Box\}$. If $\varphi \to \psi$ is a simple Sahlqvist formula, the corresponding first-order formula can be obtained as follows:

Step 1 and Step 2 proceed as in Algorithm A.1. The only difference is that as a result of these two steps, we obtain a formula of the form

$$\forall P_1 \ldots \forall P_k \forall x_1 \ldots \forall x_m((REL \wedge BOX\text{-}AT) \to ST_x(\psi)).$$

REL is a conjunction of atomic formulas of the form $x_i R_\nabla x_j$, corresponding to occurrences of existential modal operators. *BOX-AT* is a conjunction of formulas corresponding to each boxed atom. For a boxed atom of the form $\nabla_1 \ldots \nabla_n p$ (where $n \geqslant 0$, and $\nabla_i \in \{K, \Box\}$ for each $i = 1, \ldots, n$), this formula can be expressed as:

$$\forall y(x_i R^n y \to P(y)),$$

where $x_i R^n y$ abbreviates:

$$\exists z_1(x_i R_{\nabla_1} z_1 \wedge \exists z_2(z_1 R_{\nabla_2} z_2 \wedge \cdots \wedge \exists z_{n-1}(z_{n-2} R_{\nabla_{n-1}} z_{n-1} \wedge z_{n-1} R_{\nabla_n} y) \ldots)).$$

Note that if we consider the translation of a propositional variable (i.e., $n = 0$), the formula $x_i R^n y$ should be interpreted as $x_i = y$.

Step 3 (Establishing instances). Let P_i be a predicate occurring in the formula obtained in Step 2, and let $\pi_1(x_{i_1}), \ldots, \pi_k(x_{i_n})$ represent all the translations of boxed atoms in which the predicate P_i occurs. Every $\pi_j(x_{i_j})$ is therefore of the form $\forall y(x_{i_j} R^j y \rightarrow P(y))$. We define the substitution σ as follows:

$$\sigma(P_i) = \lambda u.(x_{i_1} R^1 u \vee \cdots \vee x_{i_n} R^n u).$$

Step 4 (Instantiating). Using the substitution defined in Step 3, we obtain the following formula:

$$[\sigma(P_1)/P_1, \ldots, \sigma(P_k)/P_k] \forall x_1 \ldots \forall x_m((REL \wedge BOX\text{-}AT) \rightarrow ST_x(\psi)).$$

The resulting formula is a first-order formula.

We will use the above algorithm to obtain a first-order formula corresponding to the formula $(p \wedge \Diamond\Box p) \rightarrow \Diamond Kp$. In Step 1, by applying the standard second-order translation, we obtain the following formula:

$$\forall P((P(x) \wedge \exists y(xR_\Box y \wedge \forall z(yR_\Box z \rightarrow P(z)))) \rightarrow \exists a(xR_\Box a \wedge \forall b(aR_K b \rightarrow P(b)))).$$

In Step 2, after performing the transformation, we have

$$\forall P \forall y((xR_\Box y \wedge \forall z(yR_\Box z \rightarrow P(z)) \wedge P(x)) \rightarrow \exists a(xR_\Box a \wedge \forall b(aR_K b \rightarrow P(b)))).$$

Note that *BOX-AT* is $\forall z(yR_\Box z \rightarrow P(z)) \wedge P(x)$. Thus, in Step 3, we determine the following substitution:

$$\sigma(P) \equiv \lambda u.(yR_\Box u \vee u = x).$$

According to Step 4 of the algorithm, we carry out the substitution and obtain

$$\forall y((xR_\Box y \wedge (yR_\Box x \vee x = x) \wedge \forall z(yR_\Box z \rightarrow (yR_\Box z \vee z = x))) \rightarrow \exists a(xR_\Box a \wedge \forall b(aR_K b \rightarrow (yR_\Box b \vee b = x)))).$$

By deleting the tautological parts from the antecedent, we finally obtain

$$\forall y(xR_\Box y \rightarrow \exists a(xR_\Box a \wedge \forall b(aR_K b \rightarrow (yR_\Box b \vee b = x)))).$$

PROPOSITION B.1 (6.1). *Let $\mathcal{F} = (W, R_K, R_\Box)$ be an epistemic-alethic frame. Then, the following are equivalent*:

(a) $\mathcal{F} \models (\varphi \wedge \Diamond\Box\varphi) \rightarrow \Diamond K\varphi$,
(b) $\forall x, y \in W(xR_\Box y \rightarrow \exists a \in W(xR_\Box a \wedge \forall b \in W(aR_K b \rightarrow (yR_\Box b \vee b = x))))$.

PROOF. "(a) $\Rightarrow$ (b)" We will prove this by contraposition. Let $\mathcal{F} = (W, R_K, R_\Box)$ be an epistemic-alethic frame such that

$$\exists x, y \in W(xR_\Box y \wedge \forall a \in W(xR_\Box a \rightarrow \exists b \in W(aR_K b \wedge \neg yR_\Box b \wedge b \neq x))). \tag{$*$}$$

We will show that for a certain formula φ and a certain model $\mathcal{M}$ based on $\mathcal{F}$, it holds that $\mathcal{M} \not\models (\varphi \wedge \Diamond\Box\varphi) \rightarrow \Diamond K\varphi$.

Based on ($*$), choose $s, t \in W$ such that $sR_\Box t$ and the following condition holds:

$$\forall a \in W(sR_\Box a \rightarrow \exists b \in W(aR_K b \wedge \neg tR_\Box b \wedge b \neq s)). \tag{$**$}$$

Consider the model $\mathcal{M} = (W, R_K, R_\Box, v)$ based on $\mathcal{F}$, such that $v(p) = \{x \in W : tR_\Box x\} \cup \{s\}$. The goal is to show that $\mathcal{M}, s \models p \wedge \Diamond\Box p$ and $\mathcal{M}, s \not\models \Diamond Kp$.

Given that $\mathcal{M}, s \models p$ and $sR_\Box t$, to establish $\mathcal{M}, s \models p \wedge \Diamond\Box p$, it suffices to demonstrate that $\mathcal{M}, t \models \Box p$. Let $u \in W$ such that $tR_\Box u$. Then, by the choice of $v(p)$, we obtain $\mathcal{M}, u \models p$. Hence, $\mathcal{M}, t \models \Box p$.

To complete the proof of the first implication, we need to demonstrate that $\mathcal{M}, s \not\models \Diamond Kp$, meaning $\forall a \in W(sR_\Box a \rightarrow \mathcal{M}, a \not\models Kp)$. Let $a \in W$ be such that $sR_\Box a$. Then, based on ($**$), choose $b \in W$ such that $aR_K b$, $\neg tR_\Box b$, and $b \neq s$. From $\neg tR_\Box b$ and $b \neq s$, according to the choice of $v(p)$, it follows that $\mathcal{M}, b \not\models p$. Since $aR_K b$ and $\mathcal{M}, b \not\models p$, we obtain $\mathcal{M}, a \not\models Kp$. Hence, $\mathcal{M}, s \not\models \Diamond Kp$.

We have thus shown that $\mathcal{M}, s \models p \wedge \Diamond\Box p$ and $\mathcal{M}, s \not\models \Diamond Kp$. Therefore, $\mathcal{M} \not\models (\varphi \wedge \Diamond\Box\varphi) \rightarrow \Diamond K\varphi$. Since the model $\mathcal{M}$ is based on $\mathcal{F}$, this concludes the proof that $\mathcal{F} \not\models (\varphi \wedge \Diamond\Box\varphi) \rightarrow \Diamond K\varphi$.

"(b) $\Rightarrow$ (a)" Let $\mathcal{F} = (W, R_K, R_\Box)$ be an epistemic-alethic frame such that

$$\forall x, y \in W(xR_\Box y \rightarrow \exists a \in W(xR_\Box a \wedge \forall b \in W(aR_K b \rightarrow (yR_\Box b \vee b = x)))). \tag{$\dagger$}$$

We will show that for any model $\mathcal{M}$ based on $\mathcal{F}$, it holds that $\mathcal{M} \models (\varphi \wedge \Diamond\Box\varphi) \rightarrow \Diamond K\varphi$.

Let $\mathcal{M} = (W, R_K, R_\Box, v)$ be a model based on $\mathcal{F}$, and let $s \in W$. Assume that $\mathcal{M}, s \models \varphi \wedge \Diamond\Box\varphi$. Our goal is to show that $\mathcal{M}, s \models \Diamond K\varphi$.

Given that $\mathcal{M}, s \models \Diamond\Box\varphi$, choose $t \in W$ such that $sR_\Box t$ and $\mathcal{M}, t \models \Box\varphi$. Since $sR_\Box t$, based on (†), choose $a \in W$ such that $sR_\Box a$ and the following condition holds:

$$\forall b \in W(aR_K b \to (tR_\Box b \vee b = s)). \tag{‡}$$

We shall prove that $\mathcal{M}, a \models K\varphi$. Let $b \in W$ be such that $aR_K b$. Since $aR_K b$, from (‡) we know that either $tR_\Box b$ or $b = s$.

If $tR_\Box b$, by $tR_\Box b$ and $\mathcal{M}, t \models \Box\varphi$, we obtain $\mathcal{M}, b \models \varphi$. If $b = s$, by $b = s$ and $\mathcal{M}, s \models \varphi \wedge \Diamond\Box\varphi$, we obtain $\mathcal{M}, b \models \varphi$. Thus, we have shown that $\mathcal{M}, b \models \varphi$. This completes the proof that $\mathcal{M}, a \models K\varphi$. From $\mathcal{M}, a \models K\varphi$ and $sR_\Box a$, it follows that $\mathcal{M}, s \models \Diamond K\varphi$. Consequently, we obtain $\mathcal{M}, s \models (\varphi \wedge \Diamond\Box\varphi) \to \Diamond K\varphi$. Therefore, $\mathcal{M} \models (\varphi \wedge \Diamond\Box\varphi) \to \Diamond K\varphi$. Since the model $\mathcal{M}$ is based on $\mathcal{F}$, this concludes the proof that $\mathcal{F} \models (\varphi \wedge \Diamond\Box\varphi) \to \Diamond K\varphi$. ⊣

Arkadiusz Wójcik
Faculty of Philosophy and Cognitive Science
University of Białystok, Poland
a.wojcik@uwb.edu.pl
https://orcid.org/0000-0002-9895-8449