Artificial Intelligence 349 (2025) 104423

Contents lists available at ScienceDirect

Artificial Intelligence

journal homepage: www.elsevier.com/locate/artint

The topology of surprise ☆

Alexandru Baltag [a], Nick Bezhanishvili [a], David Fernández-Duque [b,*]

[a] ILLC, University of Amsterdam, the Netherlands
[b] Department of Philosophy, University of Barcelona, Spain

ARTICLE INFO

Keywords:
Epistemic logic
Logical paradoxes
Topological semantics

ABSTRACT

In this paper we present a topological epistemic logic, with modalities for knowledge (modelled as the universal modality), knowability (represented by the topological interior operator), and unknowability of the actual world. The last notion has a non-self-referential reading (modelled by Cantor derivative: the set of limit points of a given set) and a self-referential one (modelled by Cantor's perfect core of a given set: its largest subset without isolated points, where x is isolated iff $\{x\}$ is open). We completely axiomatize this logic, showing that it is decidable and PSPACE-complete, and we apply it to the analysis of a famous epistemic puzzle: the Surprise Exam Paradox.

1. Introduction

Epistemic Logic has been formalized by Hintikka within the framework of possible-world semantics in relational (Kripke) models, and later rediscovered by game theorists [2] in the setting of partitional models (corresponding to the special case of S5 Kripke models, based on equivalence relations). In these forms, it has been used in Computer Science to reason about distributed systems, in AI to reason about agent-based knowledge representation, and in Philosophy to explore issues in formal epistemology.

While relational models can be used to model an agent's actual knowledge or beliefs, they do not represent the *potential evidence*: the observable properties of the world (which may be observed by the agent at some future stage). As such, these models cannot express propositional *knowability* (-what the agent may come to know), nor *secrecy* or "unknowability" of the actual world (-identified with some sensitive information that is crucial to the task at hand). Thus, Kripke-model-based Epistemic Logic cannot deal with the main issues in e.g., Learning Theory or Secure Communication, without the addition of other, rather ad-hoc ingredients (falling beyond the scope of relational Modal Logic).

An alternative interpretation of Modal Logic is based not on Kripke frames, but on topological spaces. This semantics can be traced back to McKinsey and Tarski [25]. When the modal $\Diamond$ is interpreted as topological closure Cl and the modal $\Box$ as topological interior Int, one obtains a semantics for the modal logic S4. McKinsey and Tarski also suggested a second topological semantics, obtained by interpreting the modal $\Diamond$ as Cantor derivative (recall that the derivative $d(A)$ of a set A consists of all limit points of A). The modal logic of Cantor derivative is semantically more expressive than the modal logic of the interior/closure operators: the latter can be defined in terms of derivative, but not vice-versa.

Since then, the usefulness of topological structures in Artificial Intelligence and Knowledge Representation has been well established. As noticed by Vickers [32] and Abramsky [1], the notion of observability and its logic require a topological setting. Abstract

☆ This paper is an invited revision of paper that first appeared at the 2022 International Conference on Principles of Knowledge Representation and Reasoning (KR-22).

* Corresponding author.
E-mail address: fernandez-duque@ub.edu (D. Fernández-Duque).

https://doi.org/10.1016/j.artint.2025.104423
Received 13 January 2024; Received in revised form 20 September 2025; Accepted 21 September 2025

Available online 29 September 2025

notions of computability also involve topological structures, with a famous example being the Scott topology. Research on spatial reasoning, in both topological and metric incarnations, is also of significant interest for AI. More recently, developments in Formal Learning Theory [8,11,22] and Distributed Computing [20] have taken a topological turn. Moreover, recent work in epistemic logic [5,7,27,6] uses topological structures.

In particular, a topological framework provides a natural way to model, and reason about, evidence, knowability and secrecy. The potential evidence (given by a designated family of *observable properties* of the world, interpreted as *sets of possible worlds*) naturally forms a basis (if we assume finite memory), or a sub-basis, for a topology (-the so-called "evidential topology").[1] In this setting, the interior operator explicitly captures the notion of "knowability" in the most natural and straightforward way, and other related topological notions play a key role in investigating various forms of learning and in developing generalizations of Computational Learning Theory [8,11].

These applications are based mostly on the notion of topological interior. Our paper builds on this existing work, but is the first to show the usefulness for Knowledge Representation of other topological notions, such as Cantor derivative. From a technical point of view, our formalism is obtained by adding to the logic of Cantor derivative a global modality (quantifying over all points), an operator capturing the perfect core $d^{\infty}(A)$ of a set A (defined as the largest subset of A that is equal to its own derivative) and a dynamic update modality (that goes from the original space to some definable subspace). Building on our previous work on topological μ-calculus [3,4], we give a complete axiomatization, as well as decidability and complexity results. Our proof is natural and not difficult to grasp, due in large part to subtle technical innovations which allow for a much more direct approach than that of related results in the literature (see e.g. [16,19]). The key is to observe that, whereas many standard techniques such as filtration (see e.g. [10]) do not mesh well with fixed point operators, Fine's well-established technique of maximal worlds [17] is surprisingly effective at dealing with the perfect core and other topological fixed points. In our presentation, maximal worlds become *final theories*, both as a reference to Fine and to avoid overloading the use of 'maximal'.

From a conceptual point of view, the key contribution of our paper is that we develop a logic of evidence-based knowledge, knowability, and (un)knowability of the actual world; and moreover, we apply it to the analysis of a famous epistemic paradox: the Surprise Examination paradox. We should stress that this is not just a theoretical concern or a purely philosophical game, but has consequences for real-life applications of CS and AI. In an applied context, the 'actual world' refers to the complete answer to the question(s) relevant for the current task: the full information pertinent for the problem at hand. The actual world's "unknowability" corresponds to the *secrecy* of this information from any intruders. Situations of this form are ubiquitous in CS and AI: for example, in a secure communication protocol, the 'actual world' may refer to a user's password, with the requirement that an eavesdropper should not be able to distinguish a true password from other possible options. Likewise, facial recognition software or DNA testing should identify a unique individual (rather than one that is 'close enough', especially when this may lead to conviction), but this sensitive information should never become public (to preserve the individual's right to privacy). Our logic may provide a stepping stone towards building a framework for the formalisation and verification of required specifications in such protocols.

We start by adopting the learning-theoretic reading of topology [8,7,22], in which a topological space is a way to represent the potential evidence that some (anonymous) agent may observe. The points of the space represent possible worlds (or possible states of the world): all the possibilities that are consistent with the agent's current knowledge. A proposition is known if it is true in all possible worlds. The potential evidence (that might be observed in the future) forms a topological basis $\mathcal{B}$: if a world x belongs to a basic open set $x \in U \in \mathcal{B}$, then the agent may observe proposition U in world x. The topological interior operator $\mathrm{Int}(A)$ captures the *knowability* of proposition A through observations $U \in \mathcal{B}$. When the agent gains more knowledge, some possibilities are eliminated (being ruled out by the new information), and thus the space shrinks to a subspace: this corresponds to performing a knowledge update (described in our logic by dynamic modalities).

While each of the above epistemic readings of standard topological notions are already known from the literature, the epistemic meaning of Cantor's derivative and the perfect core is not so obvious. In this paper, we propose a novel and very natural learning-theoretic interpretation of derivative. Essentially, the derivative $d(A)$ is the proposition asserting that *the actual world is unknowable (through observations), even if given (the additional information)* A.[2] Finally, the epistemic meaning of the perfect core $d^{\infty}(A)$ can be reconstructed from its fixed-point definition: essentially, $d^{\infty}(A)$ is the *self-referential version of Cantor's derivative*, i.e. the proposition asserting that "A is true, but the actual world is unknowable even given *this* information" (where 'this' refers to the very proposition that we are defining). Statements of this form are implicit in the foundation of modern cryptography through Kerckhoffs' principle: a cryptographic protocol should be secure even if the eavesdropper knows how the protocol functions, *including the fact that it is designed to hide crucial information.*

The main motivation for the introduction of the perfect core modality comes from our analysis of the Surprise Exam Paradox. The story says that a Student knows for sure that the date of the exam has been fixed in one of the five (working) days of next week. But he doesn't know in which day. The Teacher (who is known to be always truthful) announces that the *exam's date will be a surprise*:

[1] In this paper, we will adopt this interpretation. But we should note that this is *not* the only way to use topology in epistemic logic. E.g., in [5,7,6], open sets are interpreted as representing current (actual) evidence (rather than potential, observable ones), but in a context in which the evidence is 'soft' (i.e., fallible, not fully reliable).

[2] Indeed, prior to this paper, the dominant interpretation of derivative in the epistemological literature was Steinsvold's reading in terms of "belief" [31]. That interpretation has been criticized as not correctly reflecting the intuitive properties of belief and its relations to knowledge [6]. Though new, our interpretation is closer to an older work [28], based on a different framework: multi-agent $S5$ Kripke frames. In that setting, derivative is connected to ignorance (rather than to unknowability): the agent does not know the actual world.

even in the evening before the exam, the Student will still not know for sure that the exam is tomorrow. Intuitively, the Student can then prove (by backward induction, starting with Friday) that *the exam cannot take place in any day of the week*: first, if the exam were on Friday, then it wouldn't be a surprise (-since Friday is the last possible day, by Thursday evening the Student would know it); since the Teacher (truthfully) announced that the exam will be a surprise, it follows that the exam will not take place on Friday. But once Friday is eliminated, the Student can repeat the same argument, until all days are eliminated. But this is a contradiction (since we assumed the Student *knows* there will be an exam).

In some versions of the puzzle, there is an even more "paradoxical" follow-up: the assumption that the Teacher never lies is weakened, to allow the Student some way out. After deriving the above contradiction, he concludes that the Teacher lied: the exam will *not* be a surprise. Confident that, whenever the exams come, he will somehow get to know it an evening in advance (and thus be able to study in that last evening), the Student parties every day. Then, when the exam comes (say, on Wednesday), it *will* indeed be a complete surprise! So the Teacher told the truth after all?!

In this paper, we give a full analysis of the paradox using our topological epistemic logic. We distinguish between non-self-referential interpretations of Teacher's announcement (which can be formalized using Cantor derivative) and self-referential interpretations (which are captured using the perfect core modality). The first interpretation was pursued (in a non-topological, and less transparent, setting) in [18], and shown to be paradox-free: the only conclusion is that the exam cannot be on Friday, but the elimination process cannot be iterated. However, most logicians consider that the most natural (and intended) interpretation is the second, self-referential one. As in the above intuitive argumentation, this does lead to a contradiction. The correct conclusion is that a Teacher who is known to be always truthful *cannot make* such an announcement (since if she did, it would be a lie). In this, we agree with the verdict given in [29]. However, we also show that the above contradiction is only produced by the special evidential topology underlying the Surprise Exam Story. By changing the topology, we obtain "non-paradoxical" versions, in which the Teacher *can* truthfully make similar future-oriented self-referential "surprise" announcements. Our conclusion (against the opinions of many philosophical logicians) is that epistemic self-referentiality is *not* the cause of the apparent 'paradoxicality' of the Surprise Exam Paradox.

The paper is organized as follows: In Section 2 we recall the basic definitions of topology including that of the derived set and perfect core. We also recall the topological semantics of epistemic logic. In Section 3 we introduce the logical language and give the axiomatization of the logic of the perfect core. We also state there the main completeness, decidability and complexity results of the paper. Sections 4 and 5 provide discussion on our proposed solution to the surprise exam paradox via topological semantics. In Section 6 we summarize the results of the paper and discuss some open problems. Finally, all the technical details of the proofs are provided in the Appendix.

2. The evidential topology

As preliminaries, we recall here some notions from General Topology. In the view of our epistemic applications, we strengthen somewhat the standard notion of topological base, obtaining a concept that we call "strong base".

2.1. Topological preliminaries

Definition 2.1 *(Topology, strong base, open and closed sets, neighbourhoods)*. A **strong (topological) base** on a set X (called a *space*, and whose elements $x \in X$ are called *points*) is a family $\mathcal{B} \subseteq \mathcal{P}(X)$ of subsets of X (called *basic open* sets), with the property that it is *closed under finite intersections*: if $\mathcal{U} \subseteq \mathcal{B}$ is any finite subfamily, then $\bigcap \mathcal{U} \in \mathcal{B}$. This is in fact equivalent to requiring that a base is closed only under binary intersections (if $U, V \in \mathcal{B}$, then $U \cap V \in \mathcal{B}$) and contains the whole space (i.e. $X \in \mathcal{B}$).[3] A *basic neighbourhood* of a point $x \in X$ is a basic open set $U \in \mathcal{B}$ with $x \in U$.

A **topology** on a set X is a strong base $\mathcal{T} \subseteq \mathcal{P}(X)$, that satisfies the additional requirement that it is *closed under arbitrary (possibly infinite) unions*: if $\mathcal{U} \subseteq \mathcal{T}$ is any subfamily of $\mathcal{T}$, then $\bigcup \mathcal{U} \in \mathcal{T}$. The pair $(X, \mathcal{T})$ is a **topological space** and the sets $U \in \mathcal{T}$ are called *open* sets. Their complements $X - U$ (with $U \in \mathcal{T}$) are called *closed*, and have dual closure properties to the opens.[4] A *neighbourhood* of a point $x \in X$ is an open set $U \in \mathcal{T}$ with $x \in U$.

Operators in a Topological Space. [Interior, closure, derivative] An *interior point* of a set $A \subseteq X$ is a point $x \in X$ s.t. there exists a neighbourhood $U \in \mathcal{T}$ (of x) with $x \in U \subseteq A$. Given a strong base $\mathcal{B}$ for the topology $\mathcal{T}$, it is easy to see that x is an interior point of A iff there exists a *basic* neighbourhood $U \in \mathcal{B}$ (of x) s.t. $x \in U \subseteq A$. The ***interior*** $\mathrm{Int}(A)$ of a set $A \subseteq X$ is the set of all its interior points. A point $x \in X$ is *close* to a set $A \subseteq X$ if all its (basic) neighbourhoods intersect A: for all $U \in \mathcal{T}$ (or equivalently, for all $U \in \mathcal{B}$) s.t. $x \in U$ we have $U \cap A \neq \emptyset$. The ***closure*** $\mathrm{Cl}(A)$ of the set A is the set of all points that are close to A. A *limit point* of a set $A \subseteq X$ is a point $x \in X$ s.t. every (basic) neighbourhood U of x contains a point $y \in A$ with $y \neq x$; equivalently, x is a limit point of A iff $x \in \mathrm{Cl}(A - \{x\})$. The ***(Cantor) derivative*** $d(A)$ of a set A is the set of all the limit points of A. It is easy to see that $\mathrm{Cl}(A) = A \cup d(A)$. A non-limit point $x \in A - d(A)$ is called *isolated* in A.

[3] This last condition follows from applying closure under finite intersections to the empty family $\mathcal{U} = \emptyset \subseteq \mathcal{B}$, since $\bigcap \emptyset = X$.

[4] By applying closure under unions to the empty family of open sets $\mathcal{U} = \emptyset$, it is easy to see that $\emptyset$ is open (as well as closed, being the complement $X - X$ of the open set X).

It is important to note that operators Int, Cl and d are *monotonic* operators, e.g. in particular $A \subseteq B$ implies $d(A) \subseteq d(B)$.

Generated Topology. The *topology generated by* a strong base $\mathcal{B} \subseteq \mathcal{P}(X)$ is the smallest topology $\mathcal{T} \subseteq \mathcal{P}(X)$ s.t. $\mathcal{B} \subseteq \mathcal{T}$. We then say that $\mathcal{B}$ is a *base for* $\mathcal{T}$. The generated topology can be explicitly characterized as consisting of all possible unions of basic opens: $\mathcal{T} = \{\bigcup \mathcal{U} : \mathcal{U} \subseteq \mathcal{B}\}$.

Subspace Topology. Every subset $A \subseteq X$ of a topological space $(X, \mathcal{T})$ is a ***subspace*** of the original space, when endowed with the subspace topology $\mathcal{T}_A = \{A \cap U : U \in \mathcal{T}\}$. Every strong base $\mathcal{B}$ for $\mathcal{T}$ induces a corresponding strong base for $\mathcal{T}_A$, obtained by taking $\mathcal{B}_A = \{A \cap U : U \in \mathcal{B}\}$. All the above topological notions can be relativized to a subspace: e.g. for any subset $P \subseteq A$, we can define its relative interior $\mathrm{Int}_A(P)$ in A, closure $\mathrm{Cl}_A(P)$ in A and derivative $d_A(P)$ in A, by applying the above definitions in the subspace A. It is easy to see that $\mathrm{Int}_A(P) = A \cap \mathrm{Int}(P \cup (X - A))$, $\mathrm{Cl}_A(P) = A \cap \mathrm{Cl}(P \cap A)$, and $d_A(P) = A \cap d(P \cap A)$.

Perfect Sets and Perfect Core. We say that a set $A \subseteq X$ is ***perfect*** if $A = d(A)$. The ***perfect core*** of a set A is a subset of A denoted by $d^\infty(A)$, and defined as the largest perfect subset of A.[5] The perfect core $d^\infty(A)$ is the largest fixed point of the relative derivative operator $d_A : \mathcal{P}(A) \to \mathcal{P}(A)$, that takes subsets $P \subseteq A$ into their relative derivative $d_A(P) = A \cap d(P)$ in A.[6] This fixed point exists (by the Knaster-Tarski fixed point theorem) because of the *monotonicity* of the relative derivative operator d_A (itself a consequence of the monotonicity of derivative and intersection). Using standard μ-calculus notation for this largest fixed point, we can thus write

$$d^\infty(A) = \nu P. A \cap d(P \cap A).$$

Remark 2.2. The perfect core $d^\infty(X)$ of a topological space X is a closed subset of X. However, for $A \subseteq X$, $d^\infty(A)$ need not be closed; for example, $d^\infty((0,1)) = (0,1)$, which is not closed as a subset of $\mathbb{R}$. But note that $d^\infty(A)$ is always closed **with respect to the subspace topology** on A.

Cantor-Bendixson Rank. For any set $A \subseteq X$, we define a transfinite sequence of subsets of A, by putting:

$$d^0(A) = A, \quad d^{\alpha+1}(A) = d_A(d^\alpha(A)) = A \cap d(d^\alpha(A)),$$
$$d^\lambda(A) = \bigcap_{\alpha<\lambda} d^\alpha(A) \text{ for limit ordinals } \lambda.$$

It is easy to check that this is a descending sequence

$$A = d^0(A) \supseteq d(A) = d^1(A) \supseteq \ldots \supseteq d^\alpha(A) \supseteq \ldots,$$

which thus must reach a *fixed point*; i.e. there must exist an ordinal α s.t. $d^{\alpha+1}(A) = d^\alpha(A)$. The smallest such ordinal is called the ***(Cantor-Bendixson) rank*** of A, denoted by $\mathrm{rank}(A)$. Moreover, *the fixed point of the above iterative process $d^\alpha(A)$ is the perfect core*:

$$d^{\mathrm{rank}(A)}(A) = d^\infty(A).$$

2.2. The epistemic interpretation of topology

We proceed now to explain the epistemic interpretation of the above topological notions, in terms of observable evidence and information updates.

Possible Worlds, Knowledge, Observable Evidence, and Evidential Topology. We think of the points $x \in X$ as representing *possible worlds* (or possible states of the world): all the possibilities that are consistent with some (anonymous) agent's information. Only one of these points represents the *actual world* (the true state of affairs), but the agent may not know which one: all she knows for certain is that it belongs to the set X. Every subset $P \subseteq X$ represents a "proposition", which in a given world $x \in X$ may be "true" (i.e., hold) if $x \in P$, otherwise P does not hold in x. A proposition P is "known" for certain only if it is true in all possible worlds that are consistent with the agent's information, i.e. if $P = X$. A strong basis $\mathcal{B} \subseteq \mathcal{P}(X)$ represents our agent's *potential evidence*: the properties of the world that can in principle be *directly observed* by the agent. When $x \in U \in \mathcal{B}$, the agent may observe the truth of proposition U in world x. Note that only the observable properties that are *true* in a world x will be observed in x (i.e. we assume observations to be sound or "correct"). So in world x the observable evidence corresponds to basic neighbourhoods of the point x. Note also that this is *not* yet "evidence in hand" (that the agent already possesses), but "evidence out there" (that might be observed in the future). For example, these may represent information contained in the literature that the agent has access to (but may have not yet read), or the possible outcomes of performing a measurement or experiment. The two conditions that underlie our definition of strong basis have a clear epistemic meaning: closure under binary intersections says that our agent is *able to accumulate observations*: after observing propositions U and V, the agent will in effect have observed the truth of the conjunction $U \cap V$ (coming to know that $x \in U \cap V$); while the condition $X \in \mathcal{B}$ says that the agent can directly *observe the truth of a tautology*.

[5] Here, "perfect" should be understood with respect to the subset topology of A and "largest" is used in the sense of inclusion: so the perfect core $d^\infty(A)$ is the unique set B satisfying the following three conditions: (1) $B \subseteq A$; (2) $B = d_A(B)$; (3) every set B' satisfying conditions (1) and (2) is included in B.

[6] Once again, "largest" is taken here in the sense of inclusion.

Knowability and Conditional Knowability. Interior points $x \in \mathrm{Int}(P)$ represent worlds in which proposition P is *knowable at x* based on direct observations: P is true at x, and this fact can be known after some more evidence is observed. This interpretation follows directly from the definition: $x \in \mathrm{Int}(P)$ holds iff there exists some observable evidence that entails P (i.e. $U \in \mathcal{B}$ with $x \in U \subseteq P$). So, as an epistemic proposition, $\mathrm{Int}(P)$ says that *the proposition P can be known from observations*. More generally, the proposition $\mathrm{Int}(A \Rightarrow P) = \mathrm{Int}((X - A) \cup P)$ captures *conditional knowability*: it says that *P can be known (from observations) given A*.

Unknowability and Falsifiability. The complement $X - \mathrm{Int}(P)$ thus corresponds to "unknowability" of P, while the closure $\mathrm{Cl}(P) = X - \mathrm{Int}(X - P)$ corresponds to unfalsifiability of P: $x \in \mathrm{Cl}(P)$ means that, no matter what more evidence about x will be observed, P will never be known to be false. Note though that *our notion of unknowability is not an absolute barrier to knowledge*: it only expresses the fact that P cannot be known by direct observations (of evidence observable by the agent). Such an 'unknowable' P may still become known based on information received from another source (e.g. another agent), or more generally by *knowledge updates,* discussed below.

Verifiable and Falsifiable Propositions. The open sets $U \in \mathcal{T}$ represent *(inherently) verifiable propositions*: the ones having the property that they are knowable/verifiable whenever they are true (cf. [22,32]). This interpretation is backed by the following equivalence:

$$P \in \mathcal{T} \text{ iff } P \subseteq \mathrm{Int}(P).$$

Similarly, the closed sets represent *(inherently) falsifiable propositions*: whenever they are false, they can become known to be false after some more evidence is observed.

Knowledge Updates. The move from the original topology on X to the subspace topology on some subset $A \subseteq X$ corresponds to performing an *update* of the agent's knowledge base with the proposition A: the possible worlds not satisfying A are eliminated, so the agent comes to know A after that. The update can be the result of a direct observation $A \in \mathcal{B}$; but it can also be the result of some communication from some outside source of information (e.g. an announcement from some other agent), in which case it is quite possible that $A \notin \mathcal{B}$ (i.e. A is *not* observable by our agent). However, for this update-by-elimination to be justified, it is essential that our agent *knows for certain that the source of the new information is absolutely reliable* (e.g. the other, informing agent is telling the truth).[7] The relativized interior $\mathrm{Int}_A(P) = \{x \in X : \exists B \in \mathcal{B}(x \in B \cap A \subseteq P)\}$ in the subspace topology will then capture a notion of *updated knowability* (after updating with A, the agent can come to know P based on further observations).

Examples of Evidential Topologies.

- *Complete ignorance*: the **trivial topology** $\mathcal{T} = \{\varnothing, X\}$ on a set X;
- *Omniscience (God's topology)*: the **discrete topology** $\mathcal{T} = \mathcal{P}(X) = \{Y | Y \subseteq X\}$ on X;
- Knowledge based on *measurements of a point on a line*: the **standard topology of real numbers** $X = \mathbb{R}$, with the topology $\mathcal{T}$ generated by the strong basis $\mathcal{B} = \{(a, b) : a, b \in \mathbb{Q}, a < b\}$ (open intervals with rational endpoints)[8];
- Knowledge based on *measurements in space*: the **standard topology on** $\mathbb{R}^n$, with state space $X = \mathbb{R}^n$ and topology $\mathcal{T}$ whose elements are all countable unions of rational open balls, i.e. $A \subseteq \mathbb{R}^n$ is open iff it is of the form $A = \bigcup_{i=1}^{\infty} \{x \in \mathbb{R}^n | d(x, a_i) < b_i\}$, where $a_i, b_i \in \mathbb{Q}^n$ and d is the Euclidean distance in n-dimensional space $\mathbb{R}^n$. Rational open balls do not form a strong basis; instead, a strong basis for this topology consists of all finite intersections of rational open balls.

Concrete Example: The Policeman and the Speeding Car. A policeman uses radars with varying accuracy to determine whether a car is speeding in a 50 mph speed-limit zone. The *set of possible worlds* is $X = (0, \infty)$ (since we assume the car is known to be *moving*). The strong base

$$\mathcal{B} = \{(a, b) : a, b \in \mathbb{Q}, 0 < a < b < \infty\}$$

consists of *all possible measurement results by arbitrarily accurate radars*. The topology $\mathcal{T}$ generated by $\mathcal{B}$ is the *standard topology on real numbers* (restricted to X). "*Speeding*" is the proposition $S = (50, \infty)$.

Suppose now that a radar with accuracy 2 mph shows 51 ± 2 mph. This induces an *update*: the original space X shrinks to the subspace $A = (49, 53)$. In this updated space, "*Speeding*" becomes $S_A = (50, 53)$. Still, even now (in the subspace A, i.e. after the radar reading), the policeman does *not know* that the car is speeding (since $S_A \neq A$). However, the property "the car is speeding" *is in principle verifiable* (by the policeman): *if* the car is indeed speeding, then its velocity must be some $x \in S_A = (50, 53)$. Given a more accurate radar, the policeman can obtain a better measurement (a, b) with $x \in (a, b) \subseteq S_A$. This is reflected in the fact that $S_A = (50, 53)$ is *open* in the standard topology.

In contrast, *Not-Speeding* $NS = (0, 50]$ is *in general not verifiable* (not open). This means that *whether NS is knowable or not depends on the actual speed*! For instance, NS is knowable in the world in which the speed is $x = 49$. But it is *not* knowable in the world $x = 50$.

[7] When this is not the case, other forms of updating are to be considered (in which the non-A worlds are *not* eliminated, but only considered in some sense less plausible, or less probable, than the A-worlds).

[8] We may also allow arbitrary real-valued endpoints, but rational endpoints have the technical advantage that the basic elements can be enumerated.

On the other hand, not-speeding NS is in general falsifiable (closed in X): whenever it is false, it can be disproved by a sufficiently accurate measurement of the speed.

The Epistemic Interpretation of Cantor Derivative. To understand the derivative, recall the equivalence:

$$x \in d(A) \text{ iff } x \in \mathrm{Cl}(A - \{x\}).$$

But note that $\mathrm{Cl}(A - \{x\}) = X - \mathrm{Int}(X - (A - \{x\})) = X - \mathrm{Int}((X - A) \cup \{x\}) = X - Int(A \Rightarrow \{x\})$, where $\Rightarrow$ is material implication, i.e. $A \Rightarrow B := (X - A) \cup B$. Using our interpretation of $X - P$ as negation of the proposition P, and of $\mathrm{Int}(A \Rightarrow P)$ as conditional knowability (of P given A), we conclude that

$$x \in d(A) \text{ iff } \ x \text{ is not knowable given } A.$$

So, as an epistemic proposition, Cantor's derivative $d(A)$ says that "*the actual world is unknowable given A*".

The Epistemic Meaning of the Perfect Core. Looking now at the perfect core $d^{\infty}(A)$, we can infer its epistemic meaning from the above fixed-point identity:

$$d^{\infty}(A) = \nu P.A \cap d(P).$$

The perfect core can thus be understood as the *self-referential version of Cantor's derivative*: $d^{\infty}(A)$ captures the epistemic proposition "A is true, but the actual world is unknowable given *this* information" (where 'this' refers to the very proposition that we are defining). As we'll see, this is precisely the kind of self-referential statement that plays a key role in the Surprise Examination Paradox.

3. The logic of derivative and perfect core

In this section we introduce the formal syntax and semantics of our logic.

Syntax. The language $\mathcal{L}_{\langle\cdot\rangle}$ of dynamic-epistemic logic of derivative and perfect core consists of formulas recursively defined by the following BNF:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Diamond\varphi \mid \odot\varphi \mid \widehat{K}\varphi \mid \langle\varphi\rangle\varphi$$

where letters p come from a fixed (finite or infinite) set $Prop$ of *atomic sentences*. We will henceforth adhere to the standard conventions of omission of parentheses. The language $\mathcal{L}$ of (static) epistemic logic of derivative and perfect core is the fragment of $\mathcal{L}_{\langle\cdot\rangle}$ obtained by eliminating all dynamic modalities $\langle\varphi\rangle$.

Semantics. We interpret this language on **epistemic topo-models** $\mathbf{M} = (X, \mathcal{T}, \|\cdot\|)$: topological spaces $(X, \mathcal{T})$ with a valuation function (mapping every atomic sentence p into a subset $\|p\| \subseteq X$). The semantics is given by extending this valuation recursively to all of $\mathcal{L}_{\langle\cdot\rangle}$, defining $\|\varphi\|_{\mathbf{M}}$ using the usual clauses for Booleans, while

$$\|\Diamond\varphi\|_{\mathbf{M}} = d(\|\varphi\|_{\mathbf{M}})$$

is the Cantor derivative of $\|\varphi\|_{\mathbf{M}}$ with respect to the topology $\mathcal{T}$, and

$$\|\odot\varphi\|_{\mathbf{M}} = d^{\infty}\|\varphi\|_{\mathbf{M}} = \nu P.\big(\|\varphi\|_{\mathbf{M}} \cap d(P)\big)$$

is the perfect core of $\|\varphi\|_{\mathbf{M}}$. The operator $\widehat{K}$ is the global existential modality, quantifying existentially over all possible worlds: $\|\widehat{K}\varphi\|_{\mathbf{M}} = X$ if $\|\varphi\|_{\mathbf{M}} \neq \emptyset$, otherwise $\|\widehat{K}\varphi\|_{\mathbf{M}} = \emptyset$. Finally, $\langle\varphi\rangle\psi$ is the dynamic modality for epistemic updates, whose semantics is given by evaluating ψ in the updated model: if, for any subset $A \subseteq X$, we put $\mathbf{M}_A = (A, \mathcal{T}_A, \|\cdot\|_A)$ for the updated model, with the subspace topology $\mathcal{T}_A = \{U \cap A : U \in \mathcal{T}\}$ and relativized valuation $\|p\|_{\mathbf{M}_A} = \|p\| \cap \mathbf{M}_A$, then we set

$$\|\langle\varphi\rangle\psi\|_{\mathbf{M}} = \|\psi\|_{\mathbf{M}_{\|\varphi\|}},$$

where $\|\varphi\| = \|\varphi\|_{\mathbf{M}}$ is the valuation of φ in the original model. As usual, we may write $(\mathbf{M}, x) \vDash \varphi$ iff $x \in \|\varphi\|_{\mathbf{M}}$. When the model $\mathbf{M}$ is clear from the context, we may skip it, writing e.g. $\|\varphi\|$ and $x \vDash \varphi$, and similarly we may write $\|\cdot\|_A$ instead of $\|\cdot\|_{\mathbf{M}_A}$.

In an epistemic context, we read $\widehat{K}$ as *epistemic possibility*: $\widehat{K}\varphi$ says that "as far our agent knows, φ *may* be true", in the sense that φ is consistent with the agent's information. We read $\Diamond\varphi$ as saying that "the actual world is unknowable (through observations) given φ"; we read $\odot\varphi$ as a self-referential statement, saying that "φ is true, but the actual world is unknowable (through observation) given *this* information" (where 'this information' refers to the very proposition we are defining); finally, we read $\langle\varphi\rangle\psi$ as saying that "φ holds, and ψ will also hold after updating with φ".

Abbreviations: We will use the standard abbreviations for propositional connectives $\varphi \vee \psi$, $\varphi \Rightarrow \psi$, $\varphi \Leftrightarrow \psi$, $\top$ and $\bot$, as well as the following additional ones: $\Box\varphi := \neg\Diamond\neg\varphi$, $K\varphi := \neg\widehat{K}\neg\varphi$, $[\varphi]\psi := \neg\langle\varphi\rangle\neg\psi$, $\widehat{\mathcal{K}}\varphi := \varphi \vee \Diamond\varphi$, and $\mathcal{K}\varphi := \neg\widehat{\mathcal{K}}\neg\varphi$. To justify these notations, note that K is the universal modality (quantifying universally over all worlds that are possible according to our agent) and $\mathcal{K}$ is the interior modality: $\|K\varphi\| = X$ iff $\|\varphi\| = X$, and $\|K\varphi\| = \emptyset$ otherwise; $\|\widehat{\mathcal{K}}\varphi\| = \mathrm{Cl}(\|\varphi\|)$; and $\|\mathcal{K}\varphi\| = \|\varphi \wedge \Box\varphi\| = \mathrm{Int}(\|\varphi\|)$. So, given our interpretation of possible worlds, closure and interior, we can read $K\varphi$ as "φ is known" (to our agent), $\mathcal{K}\varphi$ as "φ is knowable" (through observations by our agent). We read $\widehat{\mathcal{K}}\varphi$ as "φ cannot be falsified" (by any observations by the agent). Meanwhile, $\Diamond$ is the Cantor derivative d and $\Box$ is its dual $\widehat{d}$, where $x \in \widehat{d}(A)$ if there is some $U \in \mathcal{T}$ such that $(U \setminus \{x\}) \subseteq A$, sometimes called the *punctured interior* of A. This means that every point near x, except for possibly x, belongs to A. Finally, the dual dynamic modality $[\varphi]\psi$ is the *conditional version* of $\langle\varphi\rangle\psi$: it says that "if φ holds, then ψ will also hold after updating with φ".

Deductive calculus: We will work with the following deductive calculus.

- Axioms and Rules of Propositional Logic.
- Necessitation Rule, and Distribution (= Kripke's Axiom), for the modalities K, $\Box$ and $[\varphi]$.[9]
- Duality Axioms: $\Diamond\varphi \Leftrightarrow \neg\Box\neg\varphi$, $\widehat{K}\varphi \Leftrightarrow \neg K\neg\varphi$, and $\langle\psi\rangle\varphi \Leftrightarrow \neg[\psi]\neg\varphi$.
- Positive and Negative Introspection for Knowledge:

$$K\varphi \Rightarrow KK\varphi \qquad \neg K\varphi \Rightarrow K\neg K\varphi$$

- Positive Introspection of Knowability *(if φ is knowable, then it is knowable to be knowable)*: $\mathcal{K}\varphi \Rightarrow \mathcal{K}\mathcal{K}\varphi$
- Knowledge Implies Knowability: $K\varphi \Rightarrow \mathcal{K}\varphi$
- Monotonicity Rule for the Perfect Core: $\dfrac{\varphi \to \psi}{\odot\varphi \to \odot\psi}$
- Fixed Point Axiom: $\odot\varphi \Rightarrow (\varphi \wedge \Diamond\odot\varphi)$
- Induction Axiom: $\mathcal{K}(\varphi \Rightarrow \Diamond\varphi) \Rightarrow (\varphi \Rightarrow \odot\varphi)$
- Reduction Axioms for Update Modalities:

$$\langle\varphi\rangle p \Leftrightarrow (\varphi \wedge p)$$

$$\langle\varphi\rangle\neg\theta \Leftrightarrow (\varphi \wedge \neg\langle\varphi\rangle\theta)$$

$$\langle\varphi\rangle\widehat{K}\theta \Leftrightarrow (\varphi \wedge \widehat{K}\langle\varphi\rangle\theta)$$

$$\langle\varphi\rangle\Diamond\theta \Leftrightarrow (\varphi \wedge \Diamond\langle\varphi\rangle\theta)$$

$$\langle\varphi\rangle\odot\theta \Leftrightarrow \odot\langle\varphi\rangle\theta$$

As usual, $\vdash \varphi$ means that φ is derivable using these axioms and rules. For a set of formulas Γ, $\Gamma \vdash \varphi$ means that there are $\gamma_1, \ldots, \gamma_n \in \Gamma$ such that $\vdash (\gamma_1 \wedge \ldots \wedge \gamma_n) \to \varphi$. The set Γ is inconsistent if $\Gamma \vdash \bot$, else it is consistent.

Theorem 3.1 *(Completeness for $\mathcal{L}_{\langle\cdot\rangle}$). The above system is a sound and complete axiomatization of the dynamic $\mathcal{L}_{\langle\cdot\rangle}$-logic of Cantor derivative and perfect core; in other words, given a formula φ of $\mathcal{L}_{\langle\cdot\rangle}$, we have that $\vdash \varphi$ (φ is* derivable*) if and only if $\models \varphi$ (φ is* valid *over the class of epistemic topo-models).*

Proving soundness is an easy verification. Completeness follows immediately from putting together the following two results:

Theorem 3.2 *(Provable Co-expressivity of $\mathcal{L}_{\langle\cdot\rangle}$ and $\mathcal{L}$). Every formula in the language $\mathcal{L}_{\langle\cdot\rangle}$ is provably equivalent*[10] *to some formula in the static fragment $\mathcal{L}$. Hence, the languages $\mathcal{L}_{\langle\cdot\rangle}$ and $\mathcal{L}$ have the same expressivity.*[11]

Theorem 3.3 *(Completeness for $\mathcal{L}$). The system obtained from the above axiomatic system for $\mathcal{L}_{\langle\cdot\rangle}$ by eliminating all axioms and rules that refer to dynamic modalities (specifically: eliminating Necessitation and Distribution for $[\varphi]$, as well as the reduction axioms) is a sound and complete axiomatization of the static $\mathcal{L}$-logic of Cantor derivative and perfect core.*

Proof Summary While the proof of Theorem 3.2 is an easy induction (using the reduction axioms to gradually push the dynamic modalities past other operators and then eliminate them), the proof of Theorem 3.3 is highly non-trivial, and uses methods that we developed in our recent work on topological μ-calculus [3,4]. The full proof is found in the Appendix; here we provide a bird's eye overview. Essentially, we start from the canonical model Ω (comprising all maximally consistent theories accessible from some fixed theory), a standard construction in modal logic. But we should stress that Ω is *not* our intended model.[12] Indeed, the usual Truth Lemma fails for our logic $\mathcal{L}$ in the canonical model: formulas are not necessarily satisfied in Ω by the theories that contain them. Next, for any given finite set of formulas Σ, we select a special submodel of the canonical model Ω^Σ (called the Σ-final model), which consists of "Σ-final theories": essentially, these are the ones whose cluster is locally definable by some formula in Σ. Our goal will

[9] In fact, Necessitation for $\Box$ follows from Necessitation for K and the axiom "Knowledge Implies Knowability".

[10] This means that the equivalence is provable in the above axiomatic system for $\mathcal{L}_{\langle\cdot\rangle}$.

[11] But they differ in succinctness: formulas in $\mathcal{L}_{\langle\cdot\rangle}$ can be in general exponentially more succinct than their translations in $\mathcal{L}$ [24]. In addition, they can capture the desired dynamic-epistemic scenarios in a much more transparent and direct way than their translations. This makes dynamic modalities very useful for applications, and justifies our choice of the larger language $\mathcal{L}_{\langle\cdot\rangle}$.

[12] In fact, the notion of truth in the canonical model will play no role in this paper: we never evaluate our formulas in it, but we only use a few basic syntactic properties of this model.

be to show that the Truth Lemma does hold in Ω^Σ for Σ-formulas. It is easy to show that Ω^Σ satisfies the usual Existential Witness Lemma for modalities $\Diamond$ and $\widehat{K}$ (and formulas in Σ), but extending this to the perfect core modality $\odot$ requires some work.

Another key ingredient in our proof is the fact that Ω^Σ is "essentially" a finite object: though possibly infinite in size, it has finite 'depth', and moreover it contains only finitely many bisimilarity classes. As a consequence, the largest fixed points of the operators $P \mapsto d_{\|\varphi\|}(P)$ (that define $\|\odot\varphi\|$) are all attained in Ω^Σ below some fixed *finite* stage of the Cantor-Bendixson process. We then use these ingredients to prove our Truth Lemma for the final model Ω^Σ.

The details are in Appendix B. In Appendix C, we use the selection method to obtain a finite submodel of Ω^Σ that satisfies the same relevant formulas, thus obtaining the finite model property (and hence decidability). In Appendix D, we analyse the complexity of the selection algorithm, proving:

Theorem 3.4 *(FMP, Decidability and Complexity). The (static and dynamic) logics of Cantor derivative and perfect core have the strong finite model property (FMP), and hence they are decidable. Moreover, the satisfiability problem for the static $\mathcal{L}$-logic is PSPACE-complete.*

Some Technical-Historical Connections. As mentioned in the Introduction, McKinsey and Tarski [25] were the first to look at the modal logic of topological closure and topological interior. In our notations, these are captured by the knowability modalities $\widehat{\mathcal{K}}$ and $\mathcal{K}$. They showed that this is the same as the modal logic of reflexive-transitive frames, more precisely the modal logic S4. In our formalism, the axiom 4 corresponds to our axiom of Positive Introspection for knowability: $\mathcal{K}\varphi \Rightarrow \mathcal{K}\mathcal{K}\varphi$. We refer to [9] for an overview of results on topological completeness of modal logics above S4.

As also mentioned in the Introduction, McKinsey and Tarski also considered the modal logic of Cantor derivative. Esakia [14,15] showed that the derivative logic of all topological spaces is the same as the logic of weakly transitive frames, namely the modal logic $\mathsf{wK4} = \mathsf{K} + \mathsf{w4}$, where w4 is the weak transitivity axiom: $\Diamond\Diamond p \to \Diamond p \vee p$. In our formalism, this is easily seen to be *equivalent* to the above-mentioned axiom of Positive Introspection for knowability. Indeed, given our definition of knowability, the axiom $\mathcal{K}\varphi \Rightarrow \mathcal{K}\mathcal{K}\varphi$ can be unfolded into

$$(\varphi \wedge \Box\varphi) \Rightarrow \Box\Box\varphi.$$

This is a Sahlqvist formula (see e.g. [12]) corresponding to the weak transitivity condition on relational models, whose equivalent dual form is Esakia's weak transitivity axiom w4. Relational models are reviewed in more detail in Appendix A.

4. Surprise: non-self-referential version

There are many 'solutions' to the Surprise Exam Paradox in the literature [13,21,18,23,29,26,33,30]. Some of them concern different versions of the puzzle, in which some of the assumptions are suspended, e.g. the Student may *not know for sure* (but only believe) that there will be an exam next week, or that the Teacher always tells the truth. Though interesting, these provide "easy" ways to avoid the contradiction, so we will ignore these weakened versions, focusing on the version in which these assumptions are granted. Even so, most of the solutions proposed in the literature are unfortunately informal, or only half formalized. Gerbrandy's approach [18] is one of the few exceptions. We hereby briefly summarize his approach.

Gerbrandy's Solution. The setting uses the older, non-topological version of our dynamic-epistemic logic, more precisely the so-called Public Announcement Logic: an epistemic model $\mathbf{M} = (X, \|\cdot\|)$ is simply given by a set of possible worlds X together with a valuation map; the logic is restricted to the fragment generated by atomic sentences, Boolean connectives, the knowledge operator $K\varphi$ (modelled as universal modality) and the dynamic update operators $[\varphi]\theta$ (also called 'public announcement', and modelled by relativization to the *subset* $\|\varphi\|$, with no subspace topological structure). Like our logic, this logic is single-agent: the Teacher is only treated as an infallible *source* of truthful information, not as an agent. So the knowledge operator K refers to the Student's knowledge. Knowability $\mathcal{K}\varphi$, derivative modality $\Diamond\varphi$, and perfect core $\odot\varphi$ do not belong to this language. But the update modalities are still eliminable, via the reduction laws for Booleans and knowledge.

In the model, the set $X = \{x_1, x_2, x_3, x_4, x_5\}$ consists of five possible worlds, with the obvious meaning: for each $1 \leq i \leq 5$, x_i is the world in which the exam will come in the corresponding i^{th} day of the week. The language has 5 atomic sentences $\{p_i : 1 \leq i \leq 5\}$, where p_i means "the exam will be in the i^{th} day". The valuation is again obvious: $\|p_i\| = \{x_i\}$. Clearly, this model satisfies $K(\bigvee_{i=1}^5 p_i)$, which captures one of the main assumptions of the puzzle: the Student knows for sure there will be an exam in the next week. Furthermore, for each $1 \leq i \leq 5$, the passage of the previous days without any exam can be 'simulated' in this logic by an update with the sentence $\bigwedge_{j=1}^{i-1} \neg p_j$: indeed, this is the information gained by the Student by the evening of day $i-1$. Hence, Gerbrandy formalizes Teacher's announcement as the sentence

$$\text{SURPRISE} \;:=\; \bigwedge_{i=1}^{5} [\bigwedge_{j=1}^{i-1} \neg p_j] \neg K p_i.$$

This sentence says that, no matter in which day i will the exam come, by the evening of day $i-1$ the Student will not know for sure that the exam will be the next day. Using the reduction axioms and the definition of $[\varphi]\psi := \neg\langle\varphi\rangle\neg\psi$, this formula can be simplified to

$$\text{SURPRISE} \;\Leftrightarrow\; \bigwedge_{i=1}^{5} \left((\bigwedge_{j=1}^{i-1} \neg p_j) \to \neg K(\bigvee_{j=1}^{i} p_j) \right).$$

Finally, the assumption that the Student knows for sure that the Teacher never lies is implemented by performing an update with the above sentence SURPRISE: all worlds in which the sentence is false are eliminated, and the model shrinks to $\|\text{SURPRISE}\|$. But, using the above static equivalent, it is easy to see that, in the model X, the sentence SURPRISE is false only in world w_5 (in which the exam is on Friday) and true in all the others. Hence, the model shrinks to $\|\text{SURPRISE}\| = \{x_1, x_2, x_3, x_4\}$.

Thus, according to Gerbrandy, *the only valid conclusion is that the exam cannot be on Friday*: the first elimination step in the informal reasoning underlying the 'paradox' is the only correct one. All further elimination steps are *not* justified: e.g., the second step (eliminating Thursday) would require performing *a second update* with the sentence SURPRISE. But the Teacher only announced the sentence once! The sentence SURPRISE was true before being announced (assuming the exam won't be on Friday), but *nothing guarantees that the sentence will still be true after this announcement*: the Teacher did not claim *that!* If say, the exam will be on Thursday, then the sentence SURPRISE changes its truth value (from true to false) after the Teacher's announcement: this does not in any way contradict the truthfulness of Teacher's announcement (since it *was true* at the moment when it was announced). So the apparent 'paradox' only points to the existence of sentences that change their truth value after being announced.[13]

A first objection to the above approach is that it gives a very "low level" formalization of the sentence SURPRISE, that is highly dependent on irrelevant details (such as the number of days in the week, the linear temporal order of the observable evidence in the form of day-passing, etc). If we change the story to cover 2 weeks, the sentence SURPRISE changes. Even worse: we can build similar stories, to which the above approach simply cannot be applied, since e.g. the number of worlds is infinite, the potential observations are also infinitely many, and they cannot be arranged in any salient linear order. Let us look now at such an example.

Infinite Surprise. Let us denote the set of positive integers by $\mathbb{N}$. It is known that the Teacher chose a point x belonging to the set

$$A = \{0\} \cup \{1/n : n \in \mathbb{N}, n \geq 2\} \cup \{1/n + 1/n^m : n, m \in \mathbb{N}, n, m \geq 2\}$$

and marked it on the real line drawn on a board. The Student can perform observations, measuring the position of the point, with any arbitrary precision $\varepsilon > 0$ (by building better measurement devices); but obviously, he can never measure the position with infinite precision ($\varepsilon = 0$)! But the Teacher (who is known to be always truthful) tells the Student: "*No matter how good your measurement is, you will never know the exact position of the point*!"

Intuitively, the Student can reproduce the Surprise Exam argument to conclude that $x \notin A$, obtaining a contradiction (since he *knows* that $x \in A$). First, if the point is of the form $x = \frac{1}{n} + \frac{1}{n^m}$ for some $n, m \in \mathbb{N}$ with $n, m \geq 2$, then he will eventually be able to know its location exactly, if he continues increasing the precision of his measurements: indeed, whenever he will reach a precision $\varepsilon < |\frac{1}{n^m} - \frac{1}{n^{m+1}}|$, then his measurement will yield an open interval of the form $(a - \varepsilon, a + \varepsilon) \ni x$, whose only intersection with A is the singleton $\{x\} = \{\frac{1}{n} + \frac{1}{n^m}\}$ consisting of the exact position. But this contradicts the Teacher's announcement (that he will never know the exact position); this contradiction rules out all points of the form $\frac{1}{n} + \frac{1}{n^m}$ (with $n, m \geq 2$), so x must belong to the set $\{0\} \cup \{1/n : n \in \mathbb{N}, n \geq 2\}$. By repeating the argument, the Student can rule out next all points of the form $x = \frac{1}{n}$ (with $n \geq 2$), since in any such case he will eventually be able to know its location exactly (whenever he reaches a precision $\varepsilon < |\frac{1}{n} - \frac{1}{n+1}| = \frac{1}{n(n+1)}$), concluding that x *must belong to the singleton set* $\{0\}$. So now the Student *knows* the exact location $x = 0$ (without even having had to do any measurement), again contradicting Teacher's announcement!

Though the argument is essentially identical to the Surprise Exam, it cannot be treated using the above approach, since the possible worlds and the possible observations are infinitely many.

This is where the topological approach comes to the rescue. By abstracting away from day-passing or measurements, and considering them to be just special cases of families of observable evidence, given in the form of strong topological bases, we can see the sentence SURPRISE simply says that "*the actual world is not knowable through observations*". Using our semantics, this is captured by the formula

$$\text{SURPRISE} \ := \ \Diamond\top,$$

where $\Diamond$ is the derivative modality with respect to the evidential topology (generated by the basis $\mathcal{B}$). In the case of our Infinite Surprise, it is clear what the evidential topology is: the *standard* topology on the set A, generated by the family $\mathcal{B} = \{(a, b) \cap A : a, b \in \mathbb{Q}, a < b\}$ of (relativized) open intervals with rational endpoints. Applying Gerbrandy's analysis to this topological version, we see that

$$\|\text{SURPRISE}\|_A = \|\Diamond\top\|_A = d_A(A) = d(A) = \{0\} \cup \{1/n : n \geq 2\}$$

(since all other points are isolated in A), and we can thus conclude that *only this first elimination step is correct*: the only information that the Student is justified in extracting from the Teacher's announcement is the fact that $x \in \{0\} \cup \{\frac{1}{n} : n \geq 2\}$. Further elimination steps are not justified: though true when it was announced, the sentence $\Diamond\top$ may have changed its truth value after the announcement.

The topology of the original Surprise Paradox Going back to the original Surprise Exam story, what is the evidential topology in that case? Since "observations" correspond there to the *passing of days without exams*, the relevant strong base is

[13] Such examples are called 'Moore sentences' and are by now well-understood as non-paradoxical utterings, easily dealt with in the framework of Dynamic Epistemic Logic.

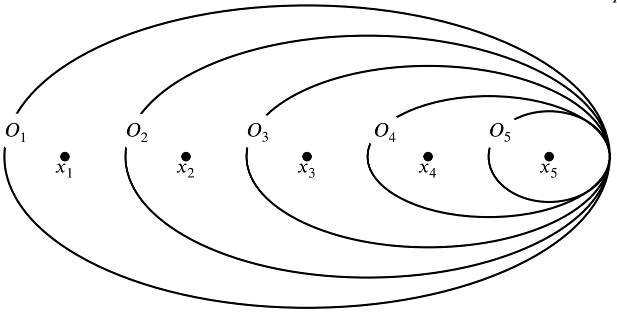

Fig. 1. The initial space X.

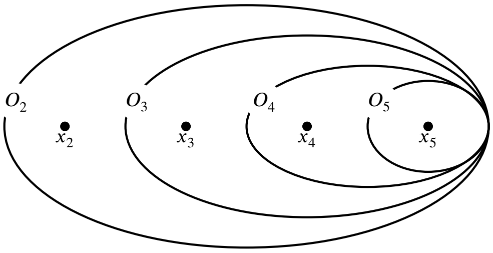

Fig. 2. The subspace O_2, obtained after Monday morning.

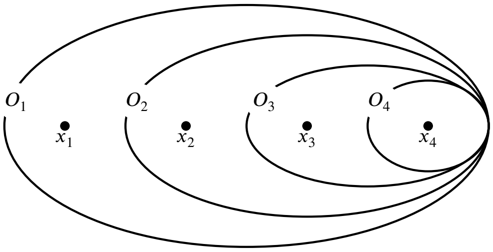

Fig. 3. The subspace $d(X)$, obtained after Teacher's announcement.

$$\mathcal{B} = \{O_1, O_2, O_3, O_4, O_5\},$$

where $O_i = X - \{x_j : j < i\} = \{x_j : i \leq j\}$. Here, $O_1 = X$ corresponds to the trivial tautological observation before Monday (that the exam will be in one of the 5 days of next week; O_2 corresponds to the negative observation after Monday morning: that the exam was *not* on Monday; etc.

So the generated evidential topology $\mathcal{T} = \{\emptyset\} \cup \mathcal{B}$ is *the topology of advance (negative) observations* (of observing non-exams in the days *before* the exam). Fig. 1 gives a visualization of this space.

The passing of each day corresponds to an update e.g., when Monday morning passes with no exam, our student observes $O_2 = X - \{x_1\}$ ("no Monday exam"), which thus becomes actual, hard evidence (rather than a potential observation). This update eliminates x_1, reducing the space to the subspace O_2 in Fig. 2.

Once again, as in Gerbrandy's analysis, $\|\Diamond\top\| = X - \{x_5\} = \{x_1, x_2, x_3, x_4\}$ (since x_5 is the only isolated point in this topology). So after updating the initial space X from Picture 1 with the Teacher's (non-referential) surprise announcement $\Diamond\top$, we are left with the first derivative $d(X) = \{x_1, x_2, x_3, x_4\}$, as pictured in Fig. 3.

The exam is known not to take place on Friday. So we have obtained a uniform treatment of the puzzle, that simplifies and generalizes Gerbrandy's solution.

5. Surprise: self-referential version

While the above formalization of the sentence SURPRISE seems natural at first sight, there is something profoundly odd about it. The teacher announced that the *exam's date **will** be a surprise*: this seems to point to the *actual future*, as it will unfold *after* this announcement is made. However, the above formalization allows for the possibility that the announcement was meant to be true only before the announcement (or counterfactually: if no such announcement was made), but to change its truth value to false after the announcement is made. In that case, in what sense can one still claim that the Teacher was truthful in her announcement about what "will" happen?

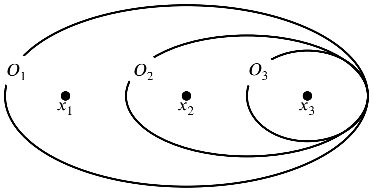

Fig. 4. The subspace $d^2(X) = d(d(X))$, after Teacher's 2nd announcement.

Looking at the sentence $\Diamond\top$ (or at Gerbrandy's more complicated non-topological counterpart), we can see that the best way to describe it in natural language is a counterfactual statement of the type: "*the exam's date* ***would*** *have been a surprise, if I didn't make this very announcement*". Moreover, this interpretation in terms of a counterfactual (instead of the actual) future seems to be crucial for Gerbrandy's 'solution' of the paradox.

However, this is *not* what the Teacher said, and it does *not* sound like the most natural interpretation of her statement. When referring to the future in an announcement, it is typically implicitly assumed that the speaker factors in her own announcement action: thus, she is expected to use the word "will" to refer to what will happen after she makes the announcement. "It will be a surprise" means that it *will* be so, not that it would have been so in some other possible future.

Thus, to understand the Teacher's statement, we need to make explicit its implicit self-referentiality, reading it as "*You will not know in advance the exam day (i.e. after hearing* ***this*** *very announcement)*". Most authors agree that *this self-referential interpretation is the intended one*.

Gerbrandy was aware of this interpretation (without formalizing it), but like many other logicians he thought that it leads to a genuine, Liar-like paradox, because of its circularity. In contrast, other logicians, such as Quine, argued in older work [29] that there is no real paradox, but only an impossible assumption: the conclusion should only be that a *source who is known to always tell the truth cannot make such a (future-oriented, implicitly self-referential) announcement* (since that would be a lie).

Using our derivative and dynamic modalities, we can formalize the self-referential announcement as a 'circular' proposition P satisfying the equation

$$P = \langle P\rangle\Diamond\top.$$

Moreover, this is *all* that is claimed in the Teacher's announcement: there is no other implicit information in it. This means that we are looking at the *most general statement* satisfying the equation, i.e. the *largest fixed point* of the operator $P \mapsto \langle P\rangle\Diamond\top$. Using standard μ-calculus notation, we can write the statement as

$$\text{SURPRISE}^{\infty} \;:=\; \nu P.\langle P\rangle\Diamond\top,$$

and call it the *self-referential surprise announcement*. Although the above formalization is not in our language $\mathcal{L}_{\langle\cdot\rangle}$ (but only in its fixed-point extension), it can be given an equivalent formulation, via the following sequence of logical equivalencies[14]:

$$\text{SURPRISE}^{\infty} \equiv \nu P.P \wedge \Diamond P \equiv \nu P.\Diamond P \equiv \odot\top.$$

Thus, the formula $\odot\top$, denoting the perfect core of our space $\|\odot\top\|_X = d^{\infty}(X)$, captures the full self-referential meaning of the surprise announcement SURPRISE^{∞}. There is nothing paradoxical with this type of self-referentiality: the monotonicity of the derivative operator ensures the existence of the fixed point. If a Teacher who is known never to lie made this announcement, that would induce an update that shrinks the original space X to its perfect core X^{∞}.

We can now recognize the successive eliminative steps in the Student's reasoning as corresponding to the Cantor-Bendixson process of calculating the perfect core: the first step eliminates the isolated point x_5, calculating the Cantor derivative $d^1(X) = X - \{x_5\}$ (as pictured above in Fig. 3); the next step calculates $d^2(X) = X - \{x_4, x_5\}$, whose topology is as in Fig. 4.

After five steps, we reach a fixed point $d^5(X) = d^{\infty}(X) = \emptyset$. A similar remark applies to our above Infinite Surprise example: the first step yields $d^1(A) = \{0\} \cup \{\frac{1}{n} : n \in \mathbb{N}\}$; the next step yields $d^2(A) = \{0\}$; finally, the third step reaches the fixed point $d^3(A) = d^{\infty}(A) = \emptyset$. And since in both cases the perfect core is empty, a contradiction is actually reached!

But, in this interpretation, *all the elimination steps are justified* (unlike in Gerbrandy's counterfactual interpretation). The conclusion is that, in the self-referential version of the story, *the Student's entire inductive eliminative reasoning is entirely correct*! The contradiction obtained in the end ($\|\text{SURPRISE}^{\infty}\| = d^{\infty}(X) = \emptyset$) only shows that *the update with* SURPRISE^{∞} *cannot be truthfully performed in this*

[14] To show the first equivalence, we use our reduction laws to show that $\langle P\rangle\Diamond\top \equiv (P \wedge \Diamond\langle P\rangle\top) \equiv (P \wedge \Diamond P)$, from which the first equivalence follows. For the second equivalence, put $A = \|\nu P.P \wedge \Diamond P\|$ and $B = \|\nu P.\Diamond P\|$, and note that by definition A is the largest set s.t. $A = A \cap d(A)$, and B is the largest set s.t. $B = d(B)$. So, to show $B \subseteq A$, it is enough to show that $B = B \cap d(B)$, which follows trivially from the fact that $B = d(B)$. Similarly, to prove the converse inclusion $A \subseteq B$, it is enough to prove that $A = d(A)$. The direction $A \subseteq d(A)$ follows trivially from the fact that $A = A \cap d(A)$. To show the other direction $d(A) \subseteq A$: we first use the direction ($A \subseteq d(A)$) and the monotonicity of d to obtain $d(A) \subseteq d(d(A))$, hence $d(A) = d(A) \cap d(d(A))$. This means that $d(A)$ satisfies the equation $P = P \cap d(P)$. But A is by definition the largest solution of this equation, hence we have $d(A) \subseteq A$, as desired. Finally, for the third equivalence above, note that by definition $\|\odot\top\| = d^{\infty}(\|\top\|) = \nu P.(\|\top\| \cap d(P)) = \nu P.d(P) = \|\nu P.\Diamond P\|$.

case: if it is known that the Teacher never lies, then *the statement* $\textsc{surprise}^\infty$ *is false, and in fact known to be false*, regardless of the day of the exam.

Liar-like paradox? Not really. The sentence $\textsc{surprise}^\infty$ has in any case some definite truth value, unlike the Liar sentences. As already mentioned, one of the assumptions of the story must simply be false: either it is *not known for sure that the Teacher always tells the truth*, or else the Teacher *never makes this self-referential announcement* in these particular situations (since it would be a lie). Note that our formalization of the sentence $\textsc{surprise}^\infty$ makes the first option impossible: by using an *update* with Teacher's announcement, the formula *assumes that the Teacher cannot lie*. Thus, the sentence is false, but *the Teacher cannot utter it*. The *appearance* of paradox is due to the fact in this specific example, the only fixed point is the empty set. However, a proposition with empty extension is by definition *not* paradoxical, but just false (in all possible worlds).

But this doesn't validate the Students' ultimate conclusion (in the follow-up story): partying every day is *not* justified. *That last follow-up step is the Student's only mistake*. If the Student gives up the first assumption (that he knew that the Teacher never lies), then the whole iterative elimination reasoning is *blocked*: even the first step is no longer justified! So, in that case, the Student can no longer be sure that the Teacher lies: she may be lying, or she may be telling the truth. All bets are off, the exam might come any day. Studying every day, instead of partying, is now the only safe option.[15]

Our diagnosis agrees with Quine's: a Teacher who is known not to lie cannot make the announcement $\textsc{surprise}^\infty$ in our two examples. But this impossibility result is *not* due to the self-referential character of the announcement. Self-referentiality is only dangerous when applied to non-monotonic operators (e.g., negation, as in the Liar). But derivative is monotonic, so *the self-referentiality involved in the Surprise story is innocuous*.[16]

In fact, the sentence $\textsc{surprise}^\infty$ can even be *true* in some situations! To see this, let us consider a modified version of the last example.

Infinite Surprise with a Twist. Everything goes as in the Infinite Surprise story, except that this time the Teacher choses a point x belonging to the set $B = A \cup [1, 2]$, where $A = \{0\} \cup \{1/n : n \in \mathbb{N}, n \geq 2\} \cup \{1/n + 1/n^m : n, m \in \mathbb{N}, n, m \geq 2\}$ is the set in the previous ('untwisted') version of Infinite Surprise. The same Cantor-Bendixson inductive process of elimination can be now used to show that the perfect core is $d^4(B) = d^\infty(B) = [1, 2]$. In this situation, an update with the same self-referential sentence $\textsc{surprise}^\infty$ shrinks the set of possible points to the subspace $[1, 2]$. In other words, an announcement of this sentence by a Teacher known to never lie simply conveys the information that the actual points satisfy $x \in [1, 2]$. A smart Student should be able to correctly infer this information, by applying the same type of "paradoxical" reasoning as in the above examples. But no contradiction is reached now: this scenario *can* happen, and if the point really is in $[1, 2]$ then the Teacher told the truth!

The conclusion of our analysis is that, in any Surprise-like paradox, *the appearance of "paradoxicality" is not due to self-referentiality, but only to the fact that the perfect core happens to be empty*. The existence of non-empty perfect sets is a topological fact, that has important epistemic consequences: the self-referential 'Surprise' sentence *can* in fact be *true* (even if in the standard story it turns out to be false). So the 'paradox' is not a paradox at all, and the Students' inductive process of elimination is a correct logical argument[17]: just a special case of the inductive Cantor-Bendixson process of calculating the perfect core! Thus, our analysis reveals deep connections between the 'paradox' and classical work in Analysis and Topology.

6. Concluding remarks

In this paper, we developed a unified topological interpretation of knowledge, observable evidence, knowability and knowledge updates, and studied a notion of "epistemic surprise" (expressing the unknowability of the actual world), that comes in two flavors: a non-self-referential version (described by Cantor derivative) and a self-referential one (described by the perfect core). We applied these notions to the analysis of the Surprise Exam Paradox, gave an axiomatisation of the associated logic, and in the Appendix prove that it is decidable and that its static fragment is PSPACE-complete.

Note two key features of the Surprise Exam Paradox: first, that it involves a unique 'correct' answer to a question (-the day of the exam); second, that the answer remains impossible to identify even with the knowledge that it is impossible to identify. Both of these features are involved in cryptographic protocols, where a unique message is sent in such a way that even an eavesdropper knowing that such specific information has been hidden should not be able to retrieve the original message. Our logic provides a framework in which to formally verify the required specifications in such contexts and e.g. ensure that an eavesdropper is not able to obtain protected information.

[15] To formalize such a relaxation of the assumption that the Teacher is known not to lie would lead to a different formalization of the sentence (one that avoids the use of updates, permitting a false announcement by the Teacher). But this would be a *different* formula than our $\textsc{surprise}^\infty$; and *that* formula would no longer be guaranteed to be false in all possible worlds!

[16] In contrast, the Liar sentence requires a fixed point for negation/complementation, which doesn't exist in a Boolean algebra. Another possible source of the feeling of paradox given by the Surprise Exam story might be the *negative* form of the Surprise sentence, as expressed in natural language, which makes it superficially similar to the Liar sentence. Thus, its self-referentiality may *look* dangerous at first sight. But looks are deceiving: in the expression "the actual world can be known, given P", the proposition P appears conditionally, and thus in a negative position; hence, when we negate this expression (saying "the world cannot be known, given P"), P reverts to a positive position. This explains the monotonicity of Cantor's derivative (and relative derivative), and thus the non-paradoxical nature of $\textsc{surprise}^\infty$.

[17] With the obvious exception of the follow-up story: as we explained above, going to party every day (after giving up on the initial assumption that it was known that the Teacher never lies) is the Student's only mistake.

Some outstanding open questions still remain. First, what is the *complexity of our dynamic logic* $\mathcal{L}_{\langle\cdot\rangle}$? Although the reduction to $\mathcal{L}$ is exponential, we conjecture that $\mathcal{L}_{\langle\cdot\rangle}$ is still PSPACE-complete (as is the case for standard Public Announcement Logic [24]). Such a result should be obtainable by avoiding a direct reduction to the static fragment (which, as mentioned, is exponential). Second, developing a *multi-agent version* of our logic would be of great value for studying epistemic dialogues, security protocols and other multi-agent epistemic scenarios and puzzles. In future work, we plan to tackle these open problems and their applications.

CRediT authorship contribution statement

Alexandru Baltag: Writing – review & editing, Writing – original draft, Investigation, Conceptualization. **Nick Bezhanishvili:** Writing – review & editing, Writing – original draft, Investigation. **David Fernández-Duque:** Writing – review & editing, Writing – original draft, Formal analysis.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: David Fernandez-Duque reports financial support was provided by Research Foundation Flanders. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We are grateful to the anonymous referees for many useful suggestions which have improved the paper.

The first and second named authors acknowledge the support of the MSCA-RISE-Marie Skłodowska-Curie Research and Innovation Staff Exchange (RISE) project MOSAIC 101007627 funded by Horizon 2020 of the European Union. The third author was partially supported by the Spanish Ministry of Science and Innovation grant PID2023-149556NB-I00. He is also an external collaborator of the Department of Mathematics WE16 of Ghent University and was supported by the FWO-FWF Lead Agency grant G030620N (FWO)/ I4513N (FWF).

Appendix A. Relational semantics

This Appendix relies heavily on relational semantics for modal logic. Although topological models provide the 'right' model for evidence in a general setting [6], a relational Kripke-style approach is often more convenient for a technical development. The two are linked via *Alexandroff spaces,* topological spaces $(X, \mathcal{T})$ in which the topology $\mathcal{T}$ is closed under *arbitrary* intersections. It is well-known such spaces admit a presentation as irreflexive and weakly transitive relational structures, or *Kripke frames*, e.g., [9]. We will briefly discuss this correspondence below.

Special Case: Standard Relational Models. If we restrict to the class of *Alexandroff spaces*, then we obtain as a special case a *relational semantics* for our logics. Alexandroff spaces are topological spaces which have the additional property that *arbitrary* intersections of open sets are open [9]. It is well-known that Alexandroff spaces are the same as standard *relational models* $(X, R, \|\cdot\|)$, with R *irreflexive and weakly transitive*: i.e. if $wRsRv$, then either $w = v$ or wRv. To see this, we note that the R-closed sets (i.e., $A \subseteq X$ such that $w \in A$ and wRv implies $v \in A$) of an irrelfexive and weakly transitive frame form an Alexandroff topology and, conversely, if X is an Alexandroff space, we define xRy iff $x \in d(\{y\})$. In other words, xRy iff $x \neq y$ and y belongs to every neighbourhood of x. Then R is irreflexive and weakly transitive, see e.g., [9].

In relational semantics, the framework of Section 3 is modified so that $\Diamond\varphi$ means that there is an accessible world where φ holds, i.e. $x \in \|\Diamond\varphi\|$ iff there is $y \in X$ such that xRy and $y \in \|\varphi\|$. It is then easy to see that the topological semantics of the modal $\Diamond$ on Aleksandorff spaces coincides with the relational semantics, when Alexandroff spaces, as outlined in the previous paragraph, are viewed as relational structures.

Note also that the topological semantics for $\odot$ on Alexandroff spaces, in the relational terms, amounts to putting $w \vDash \odot\varphi$ iff there is an infinite chain of (not necessarily distinct) worlds

$$w = w_0 \; R \; w_1 \; R \; w_2 \; R \dots R \; w_n \; R \dots$$

with $w_n \vDash \varphi$ for all n. Moreover, one can easily see that $\mathcal{K}$ is in this case the (universal) Kripke modality for the reflexive closure $\mathrm{Id} \cup R$ of R, which (due to weak transitivity) coincides with its reflexive-transitive closure R^*, where Id is the identity relation on X.

Non-standard Relational Models. The above relational clauses can give an interpretation of our syntax in any relational model $(X, R, \|\cdot\|)$ (not necessarily associated to an Alexandroff topo-model). In particular, we will be interested in dropping the irreflexivity condition, and thus interpreting our syntax in models in which R is only required to be weakly transitive. Such models are also called wK4 *models,* and their underlying frames wK4 *frames.*

Lemma A.1. *The logic of weakly transitive relational models (for our syntax) is the same as the logic of irreflexive and weakly transitive models.*

Proof. Given any weakly transitive model $\mathbf{M} = (X, R, \|\cdot\|)$, we associate to it an irreflexive and weakly transitive model $\tilde{\mathbf{M}} = (\tilde{X}, \tilde{R}, \|\cdot\|^{\sim})$, by letting W^{i} and W^{r} be the set of irreflexive and reflexive points of $\mathbf{M}$, respectively, and setting

$$\tilde{X} := \left(W^{\mathrm{i}} \times \{0\}\right) \cup \left(W^{\mathrm{r}} \times \{0,1\}\right).$$

It is useful to consider a map $\pi : \tilde{X} \to X$, given by $\pi(x,i) := x$. Using this, we can define the accessibility relation on $\tilde{X}$ by putting

$$\tilde{x} \, R \, \tilde{y} \quad \text{if} \quad \pi(\tilde{x}) \, R \, \pi(\tilde{y}) \text{ and } \tilde{x} \neq \tilde{y}$$

for all $\tilde{x}, \tilde{y} \in \tilde{X}$; and we define the valuation on $\tilde{X}$ by

$$\|p\|^{\sim} := \{\tilde{x} \in \tilde{X} : \pi(\tilde{x}) \in \|p\|\}.$$

It is easy to see that $\tilde{\mathbf{M}}$ is an irreflexive and weakly transitive relational model, and that the map $\pi : \tilde{X} \to X$ is a p-morphism[18] with respect to both modalities $\Diamond, \widehat{K}$ of our syntax. Since p-morphisms preserve the truth not only of modal formulas, but of μ-calculus formulas (see [4] for further discussion) and public announcements (as they can be eliminated), the two models are modally equivalent with respect to our syntax. □

Models for closure and interior. If we are instead interested in relational models where the Kripke modalities correspond to closure and interior in an Alexandroff space, we should consider frames where R is transitive and reflexive, also known as S4 *frames*. A model based on an S4 *frame* is an S4 *model.* The closure-interior logic of all topological spaces is equivalent to the logic of all S4 models (hence of all Alexandroff spaces), and also to the logic of all *finite* S4 models [25].

Appendix B. Proof of completeness

We prove here our main completeness result (Theorem 3.1). For this, we need to prove Theorem 3.2 (on the fact that $\mathcal{L}_{\langle\cdot\rangle}$ and $\mathcal{L}$ are provably co-expressive) and Theorem 3.3 (completeness for the static logic $\mathcal{L}$).

To establish Theorem 3.2, we first need a preliminary result:

Lemma B.1. *Let φ be any formula in $\mathcal{L}_{\langle\cdot\rangle}$, with the property that there exists some 'static' formula $\varphi' \in \mathcal{L}$ such that $\vdash \varphi \Leftrightarrow \varphi'$ is a theorem in our axiom system. Then, for every static formula $\theta \in \mathcal{L}$, there exists some other static formula $\theta' \in \mathcal{L}$, such that*

$$\vdash \langle\varphi\rangle\theta \Leftrightarrow \theta'$$

is also a theorem in our axiom system.

Proof. Induction on the standard subformula-complexity of θ. *For $\theta := p$ atomic*, the reduction axiom for atoms gives us that $\vdash \langle\varphi\rangle p \Leftrightarrow (\varphi \wedge p)$. Using our Lemma's assumption and propositional logic, we obtain $\vdash \langle\varphi\rangle p \Leftrightarrow (\varphi' \wedge p)$, thus obtaining a static equivalent.

The case $\theta := \neg\psi$. By the induction hypothesis (applied to ψ), there exists some $\psi' \in \mathcal{L}$, s.t. $\vdash \langle\varphi\rangle\psi \Leftrightarrow \psi'$ is provable in our system. This, together with our Lemma's assumption and the reduction axiom $\vdash \langle\varphi\rangle\neg\psi \Leftrightarrow (\varphi \wedge \neg\langle\varphi\rangle\psi)$, gives us that $\vdash \langle\varphi\rangle\theta \Leftrightarrow (\varphi' \wedge \neg\psi')$ is provable in our system, thus yielding the desired static equivalent.

The case $\theta := \psi \wedge \rho$ is similar, using the following derivable reduction law (which follows from Necessitation and Distribution for $[\varphi]$, using propositional logic):

$$\vdash \ \langle\varphi\rangle(\psi \wedge \theta) \ \Leftrightarrow \ (\langle\varphi\rangle\psi \wedge \langle\varphi\rangle\theta).$$

The case $\theta := \widehat{K}\psi$. By the induction hypothesis, we obtain that $\langle\varphi\rangle\psi \Leftrightarrow \psi'$ is provable in our system for some $\psi' \in \mathcal{L}$. Using our Lemma's assumption, Necessitation and Kripke's axiom for K as well as the reduction axiom for $\widehat{K}$, we obtain that $\vdash \langle\varphi\rangle\theta \Leftrightarrow (\varphi' \wedge \widehat{K}\psi')$.

The cases $\theta := \Diamond\psi$ and $\theta := \odot\psi$ are similar (using the corresponding reduction axioms). □

Proof of Theorem 3.2. We need to show that, for every formula $\varphi \in \mathcal{L}_{\langle\cdot\rangle}$, there exists some 'static' formula $\varphi' \in \mathcal{L}$ such that $\vdash \varphi \Leftrightarrow \varphi'$ is a theorem in our axiom system.

The proof is by induction on the usual subformula-complexity of φ. The atomic case $\varphi := p$ is trivial (since $p \in \mathcal{L}$). The Boolean cases $\varphi := \neg\psi$ and $\varphi := \psi \wedge \theta$, and the modal cases for $\hat{K}$, $\Diamond$ and $\odot$, are straightforward.

The case $\varphi := \langle\psi\rangle\theta$. By the induction hypothesis, there exist some $\psi', \theta' \in \mathcal{L}$ s.t. $\vdash \psi \Leftrightarrow \psi'$ and $\vdash \theta \Leftrightarrow \theta'$ are provable in our system. Using Kripke's axiom and Necessitation for $[\psi]$, we obtain that $\vdash \varphi \Leftrightarrow \langle\psi\rangle\theta'$ is provable in our system. On the other hand, by Lemma B.1 (applied to ψ, ψ' and θ'), there exists some $\theta'' \in \mathcal{L}$, s.t. $\vdash \langle\psi\rangle\theta' \Leftrightarrow \theta''$. Putting these together and using propositional logic, we conclude that $\vdash \varphi \Leftrightarrow \theta''$ is provable in our system. □

[18] We recall that a *p-morphism* or a *bounded morphism* is a functional bisimulation. In other words, a p-morphism between frames (X, R) and (X', R) is a map $f : X \to X'$ such that for each $x \in X$ we have $f[R[x]] = R'[f(x)]$, e.g., [10, Chapter 2] and [12, Chapter 3].

The rest of this section will be dedicated to proving Theorem 3.3 (completeness for the static logic $\mathcal{L}$).

Canonical Model. The standard 'canonical model' construction provides an (infinite) weakly transitive model. This is a non-standard relational model (since irreflexivity is not guaranteed), but this is not an issue in view of Lemma A.1. A *theory* is a maximally consistent set of formulas (i.e. a set T that is consistent and has no proper consistent extension). We can define a *canonical equivalence relation* between theories, by putting

$$T \sim T' \;\text{ iff }\; \forall\varphi \left(\text{ if } K\varphi \in T \text{ then } \varphi \in T' \right).$$

Similarly, the *canonical accessibility relation* $\longrightarrow$ between two theories T, T' is given as usual, by putting

$$T \longrightarrow T' \;\text{ iff }\; \forall\varphi \left(\text{ if } \Box\varphi \in T \text{ then } \varphi \in T' \right).$$

The axioms for K are Sahlquist [10], so that $\sim$ is an equivalence relation and $T \longrightarrow S$ implies that $T \sim S$ (given the Knowledge Implies Knowability axiom). To ensure that the knowledge modality really quantifies over all possible worlds, we need to restrict our model so that the relation $\sim$ becomes the universal relation. For this, we now *fix a theory* T_0, and we will restrict our canonical construction to the generated submodel.[19] Let Ω be the family of all theories T s.t. $T_0 \sim T$. The *canonical model for* T_0 is the structure $\mathbf{\Omega} = (\Omega, \longrightarrow, \|\cdot\|)$, where the canonical accessibility relation $\longrightarrow$ is restricted here to Ω, and $\|\cdot\|$ is the *canonical valuation* on Ω, given by

$$\|p\| \;:=\; \{T \in \Omega : p \in T\}.$$

Since the Positive Introspection of Knowability axiom is Sahlquist, it immediately follows that *the canonical model* $(\Omega, \longrightarrow, \|\cdot\|)$ *is indeed weakly transitive* (though not irreflexive). As a consequence, *the reflexive-transitive closure* $\longrightarrow^*$ *of the canonical relation coincides with its reflexive closure* $\longrightarrow \cup \, \mathrm{Id}_\Omega$.

9We will make use of two other well-known properties of the canonical model, given by the next two lemmas.

Lemma B.2 *(Lindenbaum Lemma). Every consistent set Φ of formulas has a maximally consistent extension ($T \in \Omega$ s.t. $\Phi \subseteq T$).*

Lemma B.3 *(Canonical Witness Lemma). For every theory $T \in \Omega$ and formula φ, we have:*

1. *$\Diamond\varphi \in T$ iff there exists some theory $T' \in \Omega$ s.t. $T \longrightarrow T' \ni \varphi$.*
   We also have an equivalent statement in $\Box$-form:

   $$\Box\varphi \in T \;\text{ iff }\; \forall T' \in \Omega \left(\text{ if } T \longrightarrow T' \text{ then } \varphi \in T' \right).$$

2. *$\widehat{K}\varphi \in T$ iff there exists some theory $T' \in \Omega$ s.t. $T \sim T' \ni \varphi$.*
   The statement in K-form is:

   $$K\varphi \in T \;\text{ iff }\; \forall T' \in \Omega \left(\text{ if } T \sim T' \text{ then } \varphi \in T' \right).$$

The left-to-right implication in the first statement above is known as the (Canonical) $\Diamond$-Existence Lemma, see e.g., [10, Lemma 4.20]. The proofs are well-known, and these results imply that the so-called Truth Lemma holds in the canonical model for the $\odot$-free fragment of our logic. Recall that the Truth Lemma states that a formula belongs to a maximal consistent set of the canonical model iff it is true at this maximal consistent set [10, Lemma 4.21].

We similarly obtain a Canonical $\widehat{\mathcal{K}}$-Witness Lemma, using the following result.

Lemma B.4. *For theories $T, T' \in \Omega$, we have:*

$$T \longrightarrow^* T' \;\text{ iff }\; \forall\varphi (\text{if } \mathcal{K}\varphi \in T \text{ then } \varphi \in T'),$$

where $\longrightarrow^* = \longrightarrow \cup \, \mathrm{Id}_\Omega$ is the reflexive closure of $\longrightarrow$.

Proof. The *left-to-right implication*: Assume that $T \longrightarrow^* T'$. If $T = T'$, then $\mathcal{K}\varphi \in T$ implies by definition that $\varphi \in T = T'$, as desired. If $T \neq T'$, then we must have $T \longrightarrow T'$, and then $\mathcal{K}\varphi \in T$ implies by definition that $\Box\varphi \in T$, which implies that $\varphi \in T'$ (by the Canonical $\Diamond$-Witness Lemma), as desired.

The *right-to-left implication*: Assume $\forall\varphi(\mathcal{K}\varphi \in T \Longrightarrow \varphi \in T')$. To show that $T \longrightarrow^* T'$, we assume that $T \neq T'$, and we need to prove that $T \longrightarrow T'$. Since $T \neq T'$, there exists some formula $\theta \in T$ with $\theta \notin T'$. To show the desired conclusion, let φ be any arbitrary formula such that $\Box\varphi \in T$; we need to prove that $\varphi \in T'$. From $\theta \in T$ we infer that $(\varphi \vee \theta) \in T$; similarly, from $\Box\varphi \in T$,

[19] Recall that a generated submodel of a model $\mathcal{M} = (W, R, V)$ is a model $\mathcal{M}' = (W, R', V')$, where $W' \subseteq W$, the relation $R' = R \cap (W' \times W')$ and $V'(p) = V(p) \cap W'$ for each propositional variable p, and $w \in W'$ and wRw' imply that $w' \in W'$, see e.g., [10, Section 2.1].

using Necessitation and Distribution we infer that $\Box(\varphi \vee \theta) \in T$. Putting these together, we obtain $\mathcal{K}(\varphi \vee \theta) \in T$. By our assumption, this implies that $(\varphi \vee \theta) \in T'$, and since $\theta \notin T'$, we conclude that $\varphi \in T'$, as desired. □

As a consequence, we immediately get:

Lemma B.5 *(Canonical $\widehat{\mathcal{K}}$-Witness Lemma). For every formula φ and theory $T \in \Omega$, we have $\widehat{\mathcal{K}}\varphi \in T$ iff there exists some theory $T' \in \Omega$ s.t. $T \longrightarrow^* T' \ni \varphi$.*

Proof. If $T \longrightarrow^* T' \ni \varphi$, then by Lemma B.4 we cannot have $\mathcal{K}\neg\varphi \in T$, so by maximal consistency of T, $\widehat{\mathcal{K}}\varphi \in T$. Conversely, if $\widehat{\mathcal{K}}\varphi \in T$ then by definition this means that $\varphi \vee \Diamond\varphi \in T$ and thus either $\varphi \in T$ or $\Diamond\varphi \in T$. In the first case we may set $T' = T$ and in the second case we use the Canonical Witness Lemma B.3 to find T' with $T \longrightarrow T' \ni \varphi$. □

Interestingly enough, the analogue of the Existence Lemma for $\odot$ also holds in the canonical model:

If $\odot\varphi \in T \in \Omega$, then there is an infinite chain $T = T_0 \longrightarrow T_1 \longrightarrow \ldots \longrightarrow T_n \longrightarrow \ldots$, with $\varphi \in T_n \in \Omega$ (and hence $\varphi \in T_n$) for all n.

To see this, use the Fixed Point Axiom to obtain $\Diamond\odot\varphi \in T$, from which the $\Diamond$-Witness Lemma B.3 yields T_1 with $T \longrightarrow T_1 \ni \odot\varphi$; then, repeating this reasoning inductively, we find T_2 with $T_1 \longrightarrow T_2 \ni \odot\varphi$, and so on. Unfortunately, *the converse fails*: there exist theories T which are part of an infinite φ-chain as above, but $\odot\varphi \notin T$.

Example B.6. Consider a sequence $(p_n)_{n<\infty}$ of pairwise distinct atoms and note that for every n, the set $\Phi_n := \{p_0, \neg\odot\top\} \cup \{\mathcal{K}(p_i \Rightarrow \Diamond p_{i+1}) : i < n\}$ is consistent (since all finite subsets are satisfiable). Use the Canonical Truth Lemma for Basic Modal Logic (and the fact that $\mathcal{K}$ is definable in it) to construct $(T_n)_{n<\infty}$ with $\Phi_n \subseteq T_n$ and $T_0 \to T_1 \to \ldots \to T_n \to \ldots$. Thus, $T_0 \models \odot\top$, although $(\neg\odot\top) \in T_0$.

So we don't have a full Canonical $\odot$-Witness Lemma, and as a consequence the Truth Lemma fails in the canonical model for the full language (with the perfect core modality $\odot$). Moreover, the filtration method (standardly used to deal with this problem in the case of PDL) does not seem to work here either. Surprisingly though, the older and simpler 'selection' method works: we will look at submodels of the canonical model, obtained by selecting only a special kind of theories, called 'final' theories.

Canonical Submodels. Any subset $X' \subseteq X$ of the set of worlds of a relational model $M = (X, \longrightarrow, \|\cdot\|)$ determines a unique *submodel*, obtained by taking: X' as its set of worlds; the restriction of $\longrightarrow$ to X' as its accessibility relation; and the valuation given by $\|p\| \cap X'$. A *canonical submodel* is a submodel of the canonical model.

Final Theories. Given a formula θ, a theory $T \in \Omega$ is *θ-final* if we have: $\theta \in T$, and for all theories $S \in \Omega$, if $T \longrightarrow S$ and $\theta \in S$ then $S \longrightarrow T$ (hence $T \longleftrightarrow S$). Given a set Σ of formulas, a theory $T \in \Omega$ is *Σ-final* if it is θ-final for some formula $\theta \in \Sigma$.

Final Model. Let Σ be any set of formulas. The *Σ-final model* is the canonical submodel determined by the set $\Omega^\Sigma := \{T \in \Omega : T \text{ is } \Sigma\text{-final}\}$ of all Σ-final theories. We denote this submodel by $\mathbf{\Omega}^\Sigma$.

It follows from Lemma B.9 below that the final model is never empty, and in fact it may be infinite. However, we can show that it has finite 'depth' whenever Σ is finite. For this, we need the following definition:

Depth of a Point in a Model. Given a (weakly transitive, not necessarily irreflexive) relational model $\mathbf{M} = (X, \longrightarrow, \sim, \|\cdot\|)$, and a point $x \in X$, a *strict (finite) x-chain* is a finite sequence of points of the form $x = x_0 \longrightarrow x_1 \longrightarrow \ldots x_n$ with $x_{i+1} \not\longrightarrow x_i$ for all $i < n$. The number n is called the *length* of our finite chain. The *depth* $\mathrm{dpt}(x)$ *of the point* $x \in X$ is the supremum of the lengths of all x-chains:

$$\mathrm{dpt}(x) := \sup\{n \in \mathbb{N} : \exists \text{ a strict } x\text{-chain of length } n\}.$$

In general, we have $\mathrm{dpt}(x) \geq 0$, with $\mathrm{dpt}(x) = 0$ iff for every $y \in X$, $x \longrightarrow y$ implies $y \longrightarrow x$; and $\mathrm{dpt}(x) = \infty$ iff there exist x-chains of every length $n \in \mathbb{N}$. The *depth* $\mathrm{dpt}(\mathbf{M})$ *of the model* $\mathbf{M}$ is the supremum of the depths of all points of the model:

$$\mathrm{dpt}(\mathbf{M}) := \sup\{\mathrm{dpt}(x) : x \in X\}.$$

Lemma B.7. *Let $\mathbf{M} = (X, \longrightarrow, \sim, \|\cdot\|)$ be a relational model of finite depth, and $x, y \in X$ be two points. Then we have the following:*

1. *if $x \longrightarrow^* y$, then $\mathrm{dpt}(x) \geq \mathrm{dpt}(y)$;*
2. *if $x \longleftrightarrow y$, then $\mathrm{dpt}(x) = \mathrm{dpt}(y)$;*
3. *if $x \longrightarrow y$ and $\mathrm{dpt}(x) = \mathrm{dpt}(y)$, then $x \longleftrightarrow y$;*
4. *if $x \longrightarrow y$ and $y \not\longrightarrow x$, then $\mathrm{dpt}(x) > \mathrm{dpt}(y)$.*

Proof. Easy verification. □

Lemma B.8 *(Finite Depth Lemma). Assume that Σ is a finite set of formulas of size $|\Sigma|$. Then the Σ-final model $\mathbf{\Omega}^{\Sigma}$ has a finite depth bounded by $|\Sigma| - 1$. In other words: for every strict chain of Σ-final theories $T_0 \longrightarrow T_1 \longrightarrow \ldots T_n$ (satisfying $T_{i+1} \not\longrightarrow T_i$ for all $i < n$), we have that $n \leq |\Sigma| - 1$.*

Proof. Suppose, towards a contradiction, that $T_0 \longrightarrow T_1 \longrightarrow \ldots T_n$ is a strict chain of Σ-final theories of length $n \geq |\Sigma|$. Since all T_i are Σ-final, there exist formulas $\theta_0, \ldots, \theta_n \in \Sigma$ s.t. T_i is θ_i-final (and hence $\theta_i \in T_i$) for all $i \leq n$. But this is a sequence of $n+1 \geq |\Sigma|+1 > |\Sigma|$ formulas in Σ, so some formula θ must be repeated. Let θ be such a repeating formula in the enumeration, and let i and j be indices such that $i < j$ and $\theta_i = \theta_j = \theta$.

So we have $T_i \longrightarrow T_{i+1} \longrightarrow^* T_j$, with both T_i and T_j being θ-final, and so also $T_i \longrightarrow^* T_j$. We have two cases: either $T_i \longrightarrow T_j$ or $T_i = T_j$. We claim that in both cases we have $T_{i+1} \longrightarrow^* T_i$. To show this, consider first the case $T_i \longrightarrow T_j$. By θ-finality we get $T_i \longleftrightarrow T_j$, hence $T_i \longrightarrow T_{i+1} \longrightarrow^* T_j \longleftrightarrow T_i$, and thus $T_i \longrightarrow T_{i+1} \longrightarrow^* T_i$, as desired. In the second case, we assume $T_i = T_j$, so we immediately obtain $T_{i+1} \longrightarrow^* T_j = T_i$, as desired.

So we showed that we have $T_i \longrightarrow T_{i+1} \longrightarrow^* T_i$. There are again two cases: either $T_i \longrightarrow T_{i+1} \longrightarrow T_i$, or $T_i \longrightarrow T_{i+1} = T_i$. In the first case, we immediately conclude that $T_i \longleftrightarrow T_{i+1}$, which contradicts the 'strictness' of our chain. In the second case, we have $T_{i+1} = T_i \longrightarrow T_{i+1} = T_i$, so we again conclude that $T_i \longleftrightarrow T_{i+1}$, in contradiction with our 'strictness' assumption. □

In order to prove completeness with respect to the final model, we first need to show that every consistent formula belongs to some final theory. This is achieved by combining the Lindenbaum Lemma with the following

Lemma B.9 *(Final Lemma). If $\varphi \in T \in \Omega$, then there exists some φ-final theory $T^* \in \Omega$ such that $T \longrightarrow^* T^*$ (and obviously, $\varphi \in T^*$, by finality).*

Proof. We will use a well-known variant of Zorn's Lemma, stated for preorders, i.e. sets equipped with a transitive, reflexive relation: a preordered set $(\mathcal{S}, \leq)$ has a maximal element if every chain has an upper bound. (Here, being maximal in a preordered set means that there is no strictly larger element.)

Let $\varphi \in T \in \Omega$. Take $\mathcal{S} := \{T' \in \Omega : T \longrightarrow^* T' \ni \varphi\}$, with the relation $\longrightarrow^*$ as its preorder. Let $\mathcal{S}' \subseteq \mathcal{S}$ be a chain of theories in $\mathcal{S}$. To show that it has an upper bound, take the set

$$\Phi := \{\varphi\} \cup \{\mathcal{K}\theta : \mathcal{K}\theta \in T' \text{ for some } T' \in \mathcal{S}'\}.$$

We show that Φ *is consistent*: suppose this is not the case. Then there exists some *finite* such inconsistent subset $\Phi' = \{\varphi\} \cup \{\mathcal{K}\theta_1, \ldots, \mathcal{K}\theta_n\}$, with $\mathcal{K}\theta_1 \in T_1, \ldots, \mathcal{K}\theta_n \in T_n$ for some theories $T_1, T_2, \ldots, T_n \in \mathcal{S}'$. Since $\mathcal{S}'$ is a chain, we can assume that $T_1, T_2, \ldots T_{n-1} \longrightarrow^* T_n$, and thus $\mathcal{K}\theta_1, \ldots, \mathcal{K}\theta_n \in T_n$. Since $T_n \in \mathcal{S}$, we also have $\varphi \in T_n$, so $\Phi' \subseteq T_n$, which contradicts the consistency of T_n.

Applying now Lindenbaum's Lemma, there exists some maximally consistent extension $S \in \Omega$ with $\Phi \subseteq S$. By construction (and using Lemma B.4), we have $T' \longrightarrow^* S$ for all $T' \in \mathcal{S}'$, so S is an upper bound for the chain $\mathcal{S}$. Applying Zorn's lemma, we obtain a $\longrightarrow^*$-maximal element $T^* \in \mathcal{S}$. In particular, this means that $\varphi \in T^*$ and $T \longrightarrow^* T^*$, as desired. To prove that T^* is φ-final, suppose that $T^* \longrightarrow S \ni \varphi$; we have to show that $S \longrightarrow T^*$. By the $\longrightarrow^*$-maximality of T^*, we must have $S \longrightarrow^* T^*$, i.e. either $S \longrightarrow T^*$ or $S = T^*$. If $S \longrightarrow T^*$, then we are done. If $S = T^*$, then $S = T^* \longrightarrow S = T^*$, so we get again $S \longrightarrow T^*$, as desired. □

The next step is to establish an analogue of the $\Diamond$-Witness Lemma for final theories:

Lemma B.10 *(Final Witness Lemma). For any theory $T \in \Omega$ and formula φ, we have:*

1. *$\Diamond\varphi \in T$ iff there exists some φ-final theory T' such that $T \longrightarrow T'$.*
2. *$\widehat{K}\varphi \in T$ iff there exists some φ-final theory T' such that $T \sim T'$.*

(Obviously, we have $\varphi \in T'$ in both cases, by finality.)

Proof. We prove the first claim, as the second is analogous. The *left-to-right implication*: by the Canonical $\Diamond$-Witness Lemma B.3, $\Diamond\varphi \in T$ implies the existence of some theory S with $T \longrightarrow S$ and $\varphi \in S$. By the Final Lemma B.9, there exists some φ-final theory S^* with $S \longrightarrow^* S^*$ and $\varphi \in S^*$. If $T \longrightarrow S^*$, then we can take $T' := S^*$ and we are done (since S^* is φ-final and $T \longrightarrow S^* \ni \varphi$, as desired). Note that this covers the case where $S = S^*$, so we may henceforth assume that $S \longrightarrow S^*$.

If $T \not\longrightarrow S^*$, then from this and $T \longrightarrow S \longrightarrow S^*$ we get by weak transitivity that $T = S^*$, and so $T \longrightarrow S \longrightarrow S^* = T$. In this case, we can take $T' := S$. Indeed, since we already know that $T \longrightarrow S \ni \varphi$, to finish the proof we only need to check that S is φ-final. For this, let $U \in \Omega$ be any theory with $S \longrightarrow U \ni \varphi$; we need to show that $U \longrightarrow S$. From $S^* = T \longrightarrow S \longrightarrow U$, we obtain by weak transitivity that either $U = S^* = T \longrightarrow S$ (and we are done), or $S^* \longrightarrow U \ni \varphi$. In the second case, by the φ-finality of S^*, we have $U \longrightarrow S^* = T \longrightarrow S$; by weak transitivity, we obtain either $U \longrightarrow S$ (and we are done) or $U = S \longrightarrow U = S$. So, in all cases, we concluded that $U \longrightarrow S$, as desired.

The converse follows directly from the Canonical $\Diamond$-Witness Lemma B.3, as a special case. □

Next, we will show that an analogue of the Witness Lemma for $\odot$ does hold in the Σ-final model (unlike in the canonical model). In fact, for finite Σ, we will prove a strong version of this lemma, in which we replace the infinite chain of ϕ-theories witnessing a formula of the form $\odot\phi \in T$ (according to the semantic clause for $\odot$) with a *very special kind of infinite chain*: a "witnessing cluster" $T \longrightarrow T' \longleftrightarrow T''$ with $\varphi \in T \cap T' \cap T''$. Our goal is to prove a $\odot$-Witness Lemma for final theories, that uses the witnessing-cluster condition. For this, we first need two following preliminary results.

Lemma B.11. *If $T \in \Omega$ is θ-final, then it is also $\widehat{\mathcal{K}}\theta$-final.*

Proof. Assume T is θ-final. To show that it is also $\widehat{\mathcal{K}}\theta$-final, observe that we have $\widehat{\mathcal{K}}\theta \in T$ (since $\theta \Longrightarrow \widehat{\mathcal{K}}\theta$ is a theorem in our logic). Second, let $S \in \Omega$ be s.t. $T \longrightarrow S$ and $\widehat{\mathcal{K}}\theta \in S$, and we need to prove that $S \longrightarrow T$. Since $\widehat{\mathcal{K}}\theta \in S$, we have either $\theta \in S$ or $\Diamond\theta \in S$. In the first case, from $T \longrightarrow S \ni \theta$ and the fact that T is θ-final, we conclude that $S \longrightarrow T$, as desired. In the second case, from $\Diamond\theta \in S$ we infer (by the Canonical $\Diamond$-Witness Lemma) that there exists $S' \in \Omega$, with $S \longrightarrow S' \ni \theta$. Since $T \longrightarrow S \longrightarrow S'$, by weak transitivity we have either $T = S'$ or $T \longrightarrow S'$. If $T = S'$, then we conclude $S \longrightarrow S' = T$, and we are done. If $T \longrightarrow S'$, then since T is θ-final and $\theta \in S'$, we get $S' \longrightarrow T$. Thus we have $S \longrightarrow S' \longrightarrow T$, hence by weak transitivity we get that either $S \longrightarrow T$ (and we are done) or $S = T$ (in which case $S = T \longrightarrow S = T$, so we again obtain $S \longrightarrow T$, as desired. □

Lemma B.12. *Let T, T', T'' be theories, and φ, θ be formulas, such that: $T \longrightarrow T' \longleftrightarrow T''$, T' is θ-final, and $\varphi \in T \cap T' \cap T''$. Then $\odot\varphi \in T$.*

Proof. Since T' is θ-final, we have $\theta \in T'$, and so also $\widehat{\mathcal{K}}\theta \in T'$. Note also that, by the Canonical $\Diamond$-Witness Lemma B.3, $T'' \longrightarrow T' \ni \theta$ implies that $\Diamond\theta \in T''$, hence also $\widehat{\mathcal{K}}\theta \in T''$. Putting these facts together with $\varphi \in T \cap T' \cap T''$, we conclude that $(\widehat{\mathcal{K}}\theta \wedge \varphi) \in T', T''$.

To prove our lemma, we first show the following

Claim: $\mathcal{K}((\widehat{\mathcal{K}}\theta \wedge \varphi) \Rightarrow \Diamond(\widehat{\mathcal{K}}\theta \wedge \varphi)) \in T'$.

The Claim is immediate from the following two facts:

(1) $((\widehat{\mathcal{K}}\theta \wedge \varphi) \Rightarrow \Diamond(\widehat{\mathcal{K}}\theta \wedge \varphi)) \in T'$, and
(2) $\Box((\widehat{\mathcal{K}}\theta \wedge \varphi) \Rightarrow \Diamond(\widehat{\mathcal{K}}\theta \wedge \varphi)) \in T'$.

Proof of fact (1): From $(\widehat{\mathcal{K}}\theta \wedge \varphi) \in T''$ and $T' \longrightarrow T''$, we obtain $\Diamond(\widehat{\mathcal{K}}\theta \wedge \varphi) \in T'$ (by Lemma B.3), and the desired conclusion follows by basic laws of propositional logic.

Proof of fact (2): by the Canonical $\Diamond$-Witness Lemma B.3, it is enough to show that $\forall S \in \Omega$, if $T' \longrightarrow S$ and $\widehat{\mathcal{K}}(\theta \wedge \varphi) \in S$, then $\Diamond(\widehat{\mathcal{K}}\theta \wedge \varphi) \in S$. To check this, let S be such that $T' \longrightarrow S \ni (\widehat{\mathcal{K}}\theta \wedge \varphi)$. Since T' is θ-final, by Lemma B.11 it is also $\widehat{\mathcal{K}}\theta$-final; from this, together with $T' \longrightarrow S \ni \widehat{\mathcal{K}}\theta$, we obtain that $S \longrightarrow T'$. Using this together with the fact that $(\widehat{\mathcal{K}}\theta \wedge \varphi) \in T'$, and applying again the Canonical $\Diamond$-Witness Lemma B.3, we conclude that $\Diamond(\widehat{\mathcal{K}}\theta \wedge \varphi) \in S$, as desired.

Using the Claim and the Induction Axiom, we conclude that $\odot(\widehat{\mathcal{K}}\theta \wedge \varphi) \in T'$. Using the Montonicity rule we see that $\odot\varphi \in T'$. Since $T \longrightarrow T'$, we get $\Diamond\odot\varphi \in T$ (again by Lemma B.3), and since $\varphi \in T$, we have $(\varphi \wedge \Diamond\odot\varphi) \in T$. Finally, using Induction and Monotonicity we see that $\vdash (\varphi \wedge \Diamond\odot\varphi) \Rightarrow \odot\varphi$ is a theorem in our system. We conclude that $\odot\varphi \in T$, as desired. □

Now we can prove the following (strong version of a) $\odot$-Witness Lemma:

Lemma B.13 *(Final $\odot$-Witness Lemma). Let Σ be a finite set of formulas, φ be a formula with $\odot\varphi \in \Sigma$, and T be a Σ-final theory such that $\varphi \in T$. The following are equivalent:*

(1) *$\odot\varphi \in T$;*
(2) *there exist $\odot\varphi$-final theories T', T'', with $T \longrightarrow T' \longleftrightarrow T''$;*
(3) *there exist Σ-final theories T', T'', with $T \longrightarrow T' \longleftrightarrow T''$ and $\varphi \in T', T''$;*
(4) *there exists an infinite chain of Σ-final theories $T = T_0 \longrightarrow T_1 \longrightarrow \dots T_n \longrightarrow \dots$, such that $\varphi \in T_n$ for all n.*

Proof. $(1) \Rightarrow (2)$**:** Assume $\odot\varphi \in T$. By the Final Lemma B.9, there exists some $\odot\varphi$-final theory T' such that $T \longrightarrow T'$ and $\odot\varphi \in T'$. Since $\vdash \odot\varphi \Longrightarrow \Diamond\odot\varphi$ is a theorem of our logic, we must have $\Diamond\odot\varphi \in T'$. By the Final $\Diamond$-Witness Lemma B.10, there exist some $\odot\varphi$-final theory T'' such that $T' \longrightarrow T''$ and $\odot\varphi \in T''$. The fact that T' is $\odot\varphi$-final ensures that $T'' \longleftrightarrow T'$, as desired.

$(2) \Rightarrow (3)$**:** It is obvious that (3) is a weaker statement.

$(3) \Rightarrow (1)$**:** Assume that Σ-final theories T', T'', with $T \longrightarrow T' \longleftrightarrow T''$ and $\varphi \in T', T''$ are given. Then there exists some $\theta \in \Sigma$ such that T' is θ-final. Apply Lemma B.12 to obtain the desired conclusion.

$(3) \Rightarrow (4)$**:** Obvious again. For all $n \geq 1$, just take $T_{2n-1} := T'$ and $T_{2n} := T''$.

(4) ⇒ (3): Let $T = T_0 \longrightarrow T_1 \longrightarrow \ldots T_n \longrightarrow \ldots$ be an infinite chain of Σ-final theories, such that $\varphi \in T_n$ for all n. By the Finite Depth Lemma B.8, this cannot be a strict chain (–in fact even its initial segment of length $|\Sigma| - 1$ must be non-strict): so there exist an index k such that $T \longrightarrow^* T_k \longleftrightarrow T_{k+1}$. We need to prove now the stronger statement (3). If we have $T = T_k$, then we get $T = T_k \longrightarrow T_{k+1} \longrightarrow T_k = T$, so by taking $n := k+1$ and $m := k$, we obtain $T \longrightarrow T_n \longleftrightarrow T_m$, as desired. If however we have $T \neq T_k$, then we get $T \longrightarrow T_k \longleftrightarrow T_{k+1}$, so by taking $n := k$ and $m := k+1$, we reach again the desired conclusion. □

We have now all the ingredients to immediately prove a Truth Lemma for the final model (and thus our completeness result). But, for later use in the decidability proof, it is convenient to state a more general form of this Truth Lemma, by abstracting the relevant properties of the final model into a definition: we consider submodels of the final model satisfying closure properties that are (the syntactic counterpart of the existential parts of) the above Final $\Diamond$- and $\odot$-Witness Lemmas.

Definition B.14 *(Perfect Submodels).* A submodel of the Σ-final model $\mathbf{\Omega}^\Sigma$ is *perfect* if the underlying set $M \subseteq \Omega^\Sigma$ satisfies the following two conditions:

(1) for every theory $T \in M$ and every formula $\Diamond\varphi \in T \cap \Sigma$, there exists some φ-final theory $T' \in M$ with $T \longrightarrow T'$;
(2) for every theory $T \in M$ and every formula $\odot\varphi \in T \cap \Sigma$, there exist $\odot\varphi$-final theories $T', T'' \in M$ with $T \longrightarrow T' \longleftrightarrow T''$, and
(3) for every theory $T \in M$ and every formula $\widehat{K}\varphi \in T \cap \Sigma$, there exists some φ-final theory $T' \in M$ with $T \sim T'$.

Example B.15. Lemmas B.10 and B.13 show that *the Σ-final model is a perfect submodel* (of itself). Later, for our decidability proof, we will see examples of *finite* perfect models.

The key result underlying our completeness and decidability proofs is the following.

Lemma B.16 *(Truth Lemma). Let Σ be a finite set of formulas, closed under subformulas, and let $\mathbf{M} = (M, \longrightarrow, \|\cdot\|)$ be a perfect submodel of the Σ-final model $\mathbf{\Omega}^\Sigma$. Then for all formulas $\varphi \in \Sigma$, we have:*

$$\|\varphi\|_{\mathbf{M}} = \{T \in M : \varphi \in T\}$$

Proof. By structural induction on φ. The atomic case and Boolean cases are standard, so we consider only the modal cases.

The case $\varphi := \Diamond\psi$: For one direction, assume that $\Diamond\psi \in T$. By condition (1) in the definition of perfect submodels, there exists some theory $T' \in M$ with $T \longrightarrow T'$ and $\psi \in T'$. By the induction hypothesis, we get $T' \models_{\mathbf{M}} \psi$, and hence $T \models_{\mathbf{M}} \Diamond\psi$, as desired.

For the converse, assume that $T \models_{\mathbf{M}} \Diamond\psi$. By the semantics, there must exist $T' \in M$ with $T \longrightarrow T'$ and $T' \models_{\mathbf{M}} \psi$. By the induction hypothesis, we get $\psi \in T'$, and so we conclude that $\Diamond\psi \in T$ (by the Canonical $\Diamond$-Witness Lemma B.3), as desired.

The case $\varphi := \widehat{K}\psi$: This case is analogous, but using the $\widehat{K}$-Witness Lemma.

The case $\varphi := \odot\psi$: For one direction, assume that $\odot\psi \in T$. By condition (2) in the definition of perfect submodels, there exist theories $T', T'' \in M$ with $T \longrightarrow T' \longleftrightarrow T''$ and $\odot\psi \in T', T''$. From $\odot\psi \in T, T', T''$, we obtain $\psi \in T, T', T''$ (by the Fixed Point Axiom), and hence (by the induction hypothesis) we have that T, T' and T'' satisfy ψ in the model $\mathbf{M}$. But then the infinite sequence $T \longrightarrow T' \longrightarrow T'' \longrightarrow T' \longrightarrow T'' \longrightarrow \ldots$ shows that $T \models_{\mathbf{M}} \odot\psi$.

For the converse, assume that $T \models_{\mathbf{M}} \odot\psi$. By definition, there must exist an infinite chain $T = T_0 \longrightarrow T_1 \longrightarrow \ldots \longrightarrow T_n \longrightarrow \ldots$, with $T_n \in M$ (hence, T_n is Σ-final) and $T_n \models_{\mathbf{M}} \psi$ for all n. By the induction hypothesis, we get $\psi \in T_n$ for all n. Applying the Final $\odot$-Witness Lemma B.13, we conclude that $\odot\psi \in T$, as desired. □

We can now finish our completeness proof.

Proof of Theorem 3.3 (Weak Completeness). Fix a consistent formula θ, and let Σ be the (finite) set consisting of θ as well as all subformulas of θ. Fix a Σ-final theory $T_0 \in \Omega^\Sigma$ with $\theta \in T_0$ (-such a theory exists by the Lindenbaum Lemma combined with the Final Lemma B.9), and consider the canonical model $\mathbf{\Omega} = (\Omega, \longrightarrow, \|\cdot\|)$ for T_0. Since $\theta \in T_0 \in \Omega^\Sigma$ and the Σ-final model $\mathbf{\Omega}^\Sigma$ is perfect (Example B.15), we can apply to it the Final Truth Lemma B.16 to conclude that $T_0 \models \theta$ in Ω^Σ. Hence, our axiomatic system is complete for the class of weakly transitive relational models. By Lemma A.1, we can add irreflexivity, so the system is also complete for the class of irreflexive and weakly transitive relational models. But, as already mentioned, this class coincides with the class of Alexandroff topo-models, so the system is also complete for topo-models. □

For decidability and FMP, we need to do a bit more work.

Appendix C. Proof of decidability

In this section, we show that the validity problems for $\mathcal{L}_{\langle\cdot\rangle}$ and $\mathcal{L}$ are decidable. Given a formula φ of $\mathcal{L}_{\langle\cdot\rangle}$, the proof of Theorem 3.2 allows us to recursively find a provably equivalent formula $\varphi' \in \mathcal{L}$, and we can reduce the validity of φ to that of φ'. Thus it suffices to prove the decidability of the static logic $\mathcal{L}$.

The key element in this proof is a small model property, which is obtained by finding an effectively bounded perfect submodel of $\mathbf{\Omega}^{\Sigma}$. Clause (3), which ensures a witnessing lemma for $\widehat{K}$, can easily be dealt with in step zero by just adding all required witnesses. Thus our main concern will be with 'defects' involving the other modalities.

Definition C.1 *(Defects and defect-depth).* If a submodel (determined by a set) $M \subseteq \Omega^{\Sigma}$ is *not* perfect, then every pair $(T, \Diamond\varphi) \in M \times \Sigma$ providing a counterexample to the clause (1) of the definition of perfect models is called a *$\Diamond$-defect* of M. A *correction* of the defect $(T, \Diamond\varphi)$ is a φ-final theory $T' \in \Omega^{\Sigma}$ with $T \longrightarrow T'$. Similarly, every counterexample $(T, \odot\varphi) \in M \times \Sigma$ to the clause (2) of the same definition is called a *$\odot$-defect* of M. A *correction-pair* of the defect $(T, \odot\varphi)$ is a pair (T', T'') of $\odot$-final theories with $T \longrightarrow T' \longleftrightarrow T''$.

The *defect-depth* $\mathrm{ddpt}(M)$ *of the submodel* (determined by) M is the maximum depth of the defects of M, defined as

$$\max\{\mathrm{dpt}(T) : \ (T, \varphi) \text{ is a defect of } M \text{ for some } \varphi \in \Sigma\}.$$

By Lemma B.8, we have $0 \le \mathrm{ddpt}(M) \le |\Sigma| - 1$ (for all final submodels M).

Obviously, M is perfect iff it has no defects (of any of the two kinds). To prove FMP, it is clear that it is enough to show the following:

Lemma C.2. *Let Σ be a finite set of formulas closed under subformulas and such that if $\odot\varphi \in \Sigma$ then $\Diamond\odot\varphi \in \Sigma$. For any Σ-final theory T_0, there exists a finite perfect submodel $M \subseteq \Omega^{\Sigma}$ with $T_0 \in M$.*

Proof. We recursively construct an infinite sequence of finite submodels

$$M_0, M_1, \ldots, M_n, \ldots$$

of the Σ-final model $\mathbf{\Omega}^{\Sigma}$:

- For each $\widehat{K}\varphi \in T_0 \cap \Sigma$, choose $T_{\varphi} \in \Omega^{\Sigma}$ such that $\varphi \in T_{\varphi}$. Put $M_0 := \{T_0\} \cup \{T_{\varphi} : \widehat{K}\varphi \in T_0 \cap \Sigma\}$.
- Given M_n, put $M_{n+1} := M_n$ if M_n is perfect. Otherwise, for each defect $(T, \Diamond\varphi)$, choose a correction $T' \in \Omega^{\Sigma}$; we'll refer to T' as the *designated correction* of that defect. Similarly, for each defect $(T, \odot\varphi)$, choose a correction-pair (T', T''); we'll refer to T' and T'' as the *designated corrections* of that defect. Define Δ_n to be the set of all $S \in \Omega^{\Sigma}$ which are a designated correction of some defect of M_n. Then, let $M_{n+1} = M_n \cup \Delta_n$.

By induction, it is clear that *all models M_n are finite* (–the induction step uses the fact that a finite model has only finitely many defects, since Σ is finite).

Claim 1. If (S, φ) is a defect of M_{n+1}, then $\varphi \in S \in M_{n+1} \setminus M_n$, and S is a designated correction of some defect (T, ψ) of M_n, with $T \longrightarrow S$.

Proof of Claim 1. Let (S, φ) be a defect of M_{n+1}. By definition, we then have $\varphi \in S \in M_{n+1}$. Suppose that we also have $S \in M_n$. We consider two cases. *Case 1*: assume that (S, φ) is *also a defect of* M_n. Then, by the construction of M_{n+1}, this defect has a designated correction $S' \in M_{n+1}$, or else a correction-pair $(S', S'') \in M_{n+1}$. But then the theory S', or the pair (S', S''), testify that (S, φ) is *not* a defect of M_{n+1}, contradicting our initial premise. *Case 2:* assume that (S, φ) is *not* a defect of M_n. In this case, there must exist witnesses $T' \in M_n$ or $(T', T'') \in M_n \times M_n$ attesting that (S, φ) is not a defect of M_n. But since $M_n \subseteq M_{n+1}$, the same witnesses also attest that (S, φ) is not a defect of M_{n+1}, again contradicting our initial premise.

So we must have $S \in M_{n+1} \setminus M_n$. But then (by the construction of M_{n+1}) S must be a designated correction to some defect (T, ψ) of M_n. If (T, ψ) is a $\Diamond$-defect, then (by the definition of its corrections) we have $T \longrightarrow S$, and we are done. If (T, ψ) is a $\odot$-defect, then S is part of a correction pair (S, S') or (S', S). In the first case (by the definition of correction pairs) we have $T \longrightarrow S$, and we are done; in the second case, we have $T \longrightarrow S' \longleftrightarrow S$, which together with the fact that $T \neq S$ (since $T \in M_n$ and $S \in M_{n+1} \setminus M_n$) gives again $T \longrightarrow S$ (by weak transitivity), as desired.

Claim 2. For a formula $\varphi = \Diamond\theta$ define $\varphi' = \theta$, and for a formula $\varphi = \odot\theta$ define $\varphi' = \odot\theta$ (φ' is undefined otherwise).

Then, if (T, φ) is a defect of M_n, there is $S \in M_{n+1}$ such that $T \longrightarrow S$ and $\varphi' \in S$.

Proof of Claim 2. This follows immediately from inspecting the definitions of 'correction' and 'correction-pair', and the fact that M_{n+1} contains a designated correction for M_n.

Claim 3. If M_{n+2} is not perfect, then $\mathrm{ddpt}(M_{n+2}) < \mathrm{ddpt}(M_n)$.

Proof of Claim 3. Let (T_{n+2}, φ_{n+2}) be a defect of M_{n+2}. Then, by applying Claim 1 twice, there exists some defect (T_n, φ_n) of M_n, as well as some defect (T_{n+1}, φ_{n+1}) of M_{n+1}, s.t. $T_{n+2} \in M_{n+2} \setminus M_{n+1}$ is a designated correction of (T_{n+1}, φ_{n+1}) (and hence $T_{n+1} \longrightarrow T_{n+2}$), and similarly $T_{n+1} \in M_{n+1} \setminus M_n$ is a designated correction of (T_n, φ_n) (hence $T_n \longrightarrow T_{n+1}$)).

By Lemma B.7, from $T_n \longrightarrow T_{n+1} \longrightarrow T_{n+2}$ we obtain $\mathrm{dpt}(T_n) \ge \mathrm{dpt}(T_{n+1}) \ge \mathrm{dpt}(T_{n+2})$, and one of the two inequalities is strict unless we also have $T_n \longleftrightarrow T_{n+1} \longleftrightarrow T_{n+2}$, so we need only consider this case.

Define $\theta := \varphi'_{n+2}$ as in Claim 2. Note that $T_n \neq T_{n+2}$ (since $T_n \in M_n$ but $T_{n+2} \notin M_{n+1} \supseteq M_n$), so we must have $T_n \longleftrightarrow T_{n+2}$ by weak transitivity. Since $\varphi_{n+2} \in T_{n+2}$, we have $\Diamond\varphi_{n+2} \in T_n$. By our axioms, we have either $\theta \in T_n$, or $\Diamond\theta \in T_n$; if $\varphi_{n+2} = \odot\psi$ this follows from $\Diamond\odot\psi \to \odot\psi$, otherwise $\varphi_{n+2} = \Diamond\psi$ and this follows from $\Diamond\Diamond\psi \to \psi \vee \Diamond\psi$.

Case 1: (T_{n+2}, φ_{n+2}) *is a* $\Diamond$*-defect*, so that $\varphi_{n+2} = \Diamond\theta \in T_{n+2}$. If $\theta \in T_n$, we have that $T_{n+2} \longrightarrow T_n \ni \theta$ and $T_n \in M_n \subseteq M_{n+2}$, which gives a witness in M_{n+2} for clause (1) applied to $\Diamond\theta \in T_{n+2}$, contradicting the assumption that (T_{n+2}, φ_{n+2}) is a defect of M_{n+2}. Otherwise, $\Diamond\theta \in T_n \in M_n$. Then by our construction there must exist some $S_{n+1} \in M_{n+1}$ with $T_n \longrightarrow S_{n+1} \ni \theta$ (either because $(T_n, \Diamond\theta)$ was not a defect of M_n hence such a theory S_{n+1} already existed in M_n, or else because the defect $(T_n, \Diamond\theta)$ has a designated correction in M_{n+1}). But $T_{n+2} \longleftrightarrow T_n \longrightarrow S_{n+1}$ implies that $T_{n+2} \longrightarrow^* S_{n+1}$, which together with the fact that $T_{n+2} \neq S_{n+1}$ (since $S_{n+1} \in M_{n+1}$ while $T_{n+2} \in M_{n+2} \setminus T_{n+1}$) gives us that $T_{n+2} \longrightarrow S_{n+1}$ (by weak transitivity). But then $\theta \in S_{n+1} \in M_{n+1} \subseteq M_{n+2}$ is a witness in M_{n+2} for clause (1) applied to $\Diamond\theta \in T_{n+2}$, again contradicting the assumption that (T_{n+2}, φ_{n+2}) is a defect of M_{n+2}.

Case 2: (T_{n+2}, φ_{n+2}) *is a* $\odot$*-defect*, say $\varphi_{n+2} = \odot\psi \in T_{n+2}$. Since $T_n \longleftrightarrow T_{n+2}$, we have $\Diamond\odot\psi \in T_n \in M_n$. Since $\odot\psi \in \Sigma$ (because this is a defect of M_{n+2}), we also have $\Diamond\odot\psi \in \Sigma$ (by the additional closure requirement of our lemma). By construction, there must exist some $S_{n+1} \in M_{n+1}$ with $T_n \longrightarrow S_{n+1} \ni \odot\psi$ (again either because $(T_n, \Diamond\odot\psi)$ was not a defect so that such a theory S_{n+1} already existed in M_n, or because the defect $(T_n, \Diamond\odot\psi)$ has a designated correction in M_{n+1}). But then we can repeat this argument on $(S_{n+1}, \Diamond\psi)$; by construction, there must exist $\odot$-final $S'_{n+2}, S''_{n+2} \in M_{n+2}$, with $S_{n+1} \longrightarrow S'_{n+2} \longleftrightarrow S''_{n+2}$ (again either because such theories already existed in M_{n+1}, or because the defect $(S_{n+1}, \odot\psi)$ has designated corrections in M_{n+2}). But then we have $T_{n+2} \longleftrightarrow T_n \longrightarrow S_{n+1} \longrightarrow S'_{n+2}$, so $T_{n+2} \longrightarrow^* S'_{n+1}$, and hence by weak transitivity we have $T_{n+2} \longrightarrow S'_{n+1}$ (since $T_{n+2} \neq S'_{n+1}$, because $T_{n+2} \in M_{n+2} \setminus M_{n+1}$, while $S'_{n+1} \in M_{n+1}$). So $T_{n+2} \longrightarrow S'_{n+1} \longleftrightarrow S''_{n+1}$ are witness in M_{n+2} for clause (ii) applied to $\odot\psi \in T_{n+2}$, contradicting the assumption that (T_{n+2}, φ_{n+2}) is a defect of M_{n+2}.

Given Claim 2, let now $N := 2 \cdot |\Sigma|$, where $|\Sigma|$ is the size of Σ.[20] We claim that M_N *is a perfect submodel*.

To show this, assume that this is not the case. Then of course none of the submodels M_n with $n \leq N$ are perfect. By repeatedly applying Claim 2, we have

$$\mathrm{ddpt}(M_0) < \mathrm{ddpt}(M_2) < \ldots < \mathrm{ddpt}(M_N).$$

This contradicts the fact that $0 \leq \mathrm{ddpt}(M_n) \leq \mathrm{dpt}(M_n) \leq |\Sigma| - 1$ (by the Finite Depth Lemma B.8): the set $\{0, 1, \ldots, |\Sigma| - 1\}$ has cardinality $|\Sigma|$, so it cannot contain $\frac{N}{2} + 1 = |\Sigma| + 1$ distinct natural numbers. □

Proof of FMP and Decidability. Fix a consistent formula θ, and let Σ be a finite set containing θ, and closed under subformulas and under the additional clause in the previous lemma (if $\odot\psi \in \Sigma$ then $\Diamond\odot\psi \in \Sigma$). In particular, we may set Σ to be the set of subformulas of φ together with all formulas $\Diamond\psi$, where ψ is a subformula of φ; such a Σ will have at most $2n$ formulas, where n is the length of φ. Fix as before a Σ-final theory $T_0 \in \Omega^\Sigma$ with $\theta \in T$, and let $\mathbf{M}$ be the finite perfect submodel constructed in the above Lemma. Since $\theta \in T_0 \in M$ and $\mathbf{M}$ is perfect, we can apply the Final Truth Lemma B.16 to conclude that $T_0 \vDash \theta$ in $\mathbf{M}$. Our submodel $\mathbf{M}$ is a finite weakly transitive relational model, but by the technique in the proof of Lemma A.1, we can convert it into an equivalent model, that is finite, irreflexive and weakly transitive. But this is nothing but a finite topo-model, so we have proved (strong) FMP for the topological semantics. Decidability immediately follows. □

To finish the proof of Theorem 3.4, we need to look at the complexity of the decision problem for the static logic.

Remark C.3. As it follows from the proof given in Appendix B, the FMP is obtained with respect to finite relational models, which can be represented as finite Alexandroff spaces. Therefore, in Theorem 3.4, the finite topo-models can be assumed to be finite Aleksandorff spaces.

Appendix D. PSPACE completeness

We may obtain a PSPACE complexity bound for the static logic from our decidability proof. First note that the validity problem for $\mathcal{L}$ is PSPACE-hard, as it embeds wK4, which is PSPACE-complete [12]. So we focus on the upper bound.

We begin with a PSPACE algorithm for satisfiability in the $\widehat{K}$-free fragment $\mathcal{L}^{\odot}_{\Diamond}$. First, some preliminary definitions. We work with a set of formulas Σ closed under subformulas, single negations and such that if $\odot\varphi \in \Sigma$, then $\Diamond\odot\varphi \in \Sigma$. As mentioned above, for any formula φ, the size of such a Σ containing φ is linear on the length of φ. A Σ*-type* is a subset $\Phi \subseteq \Sigma$ such that $\psi \wedge \theta \in \Phi$ implies that $\psi \in \Phi$ and $\theta \in \Phi$, $\neg(\psi \wedge \theta) \in \Phi$ implies that $\neg\psi \in \Phi$ or $\neg\theta \in \Phi$, and $\neg\psi \in \Phi$ if and only if $\psi \notin \Phi$ (all modulo double negations). A Σ *cluster-type* is a multiset C of Σ-types with $|C| \leq 2|\Sigma|$ and such that if $\Phi, \Psi \in C$ are such that $\Phi \neq \Psi$, or they are equal but occur at least more than once:

1. if $\psi \in \Phi$ then $\Diamond\psi \in \Psi$,

[20] We can lower this bound somewhat, taking instead the size of the set $\{\varphi : \Diamond\varphi \in \Sigma \text{ or } \odot\varphi \in \Sigma\}$.

2. if $\Diamond\psi \in \Phi$ and $\psi \notin \Psi$ then $\Diamond\psi \in \Psi$,
3. if $\odot\psi \in \Phi$ and $\psi \in \Psi$ then $\odot\psi \in \Psi$, and
4. if $\psi, \neg\odot\psi \in \Phi$ then $\neg\Diamond\odot\psi \in \Phi$.

A *defect* of C is either any formula $\Diamond\psi \in \Phi \in C$ such that $\psi \notin \Phi'$ for any $\Phi' \in C$ with $\Phi' \neq \Phi$, or a formula $\odot\psi \in \Phi \in C$ such that there is at most one $\Phi' \in C$ with $\psi \in \Phi'$.

The idea is that types represent points in a model and cluster-types represent *irreflexive* clusters, in the following sense. First, we note by Theorem 3.4 that we may restrict our attention to finite models, which can be regarded as irreflexive, weakly transitive relational models. Consider such a model, $\mathbf{M} = (X, \longrightarrow, \|\cdot\|)$. Each $w \in X$ can then be assigned a Σ-type $\mathrm{tp}_\Sigma(w)$ consisting of the set of formulas of Σ that are true on w. Thus every point in a model gives rise to a Σ-type. The converse direction is only partially true: if we are given a Σ-type Φ, it is not immediately evident whether or not we can find a model $\mathbf{M}$ with a point w such that $\mathrm{tp}_\Sigma(w) = \Phi$. This boils down to asking whether the conjunction $\bigwedge\Phi$ is satisfiable. Thus types potentially (but not always) represent the formulas that are true on some point of some model.

As for cluster-types, recall that a *cluster* is a set $C \subseteq X$ such that $w \longleftrightarrow^* v$ for all $w, v \in C$. Observe that we may restrict the size of each cluster C to have at most $2|\Sigma|$ elements. This is because each formula of Σ needs to occur at most twice, and removing additional points from C will not create new defects. A cluster C can then be assigned a cluster-type by $\mathrm{tp}_\Sigma(C) = \{\mathrm{tp}_\Sigma(w) : w \in C\}$. Note however that given irreflexivity, it is important to record whether a type appears more than once, as it may affect the truth of formulas $\Diamond\theta$. Thus we let $\mathrm{tp}_\Sigma(C)$ be a multiset instead of a set.

Next we define a mild generalization of satisfiability whose use of space is easier to control, which we call the **controlled satisfiability problem**. An instance of the controlled satisfiability problem is a sequence $(\varphi, b, \Sigma, \mathcal{K}\psi_1, \ldots, \mathcal{K}\psi_m, n)$, where $\varphi, \Box\psi_i \in \Sigma$, $b \in \{1, 2\}$, and $n \in \mathbb{N}$. We **accept** this instance if $\varphi \wedge \mathcal{K}\psi_1 \wedge \ldots \wedge \mathcal{K}\psi_m$ is satisfiable in a model of depth at most n, whose root cluster has b instances of φ (so that if $b = 2$, every point of the root cluster satisfies $\odot\varphi$). Here, a *root cluster* of a model $\mathbf{M} = (X, \longrightarrow, \|\cdot\|)$ is a cluster $C \subseteq X$ such that $w \longrightarrow^* v$ whenever $w \in C$ and $v \in X$. Not every model has a root cluster, but when working in $\mathcal{L}^{\odot}_{\Diamond}$ we can restrict our attention to such models, since the generated model from any point w will have a root cluster (namely, the cluster of w).

We then have that $\varphi \in \mathcal{L}^{\odot}_{\Diamond}$ is satisfiable if and only if $(\varphi, 1, \Sigma, |\Sigma| - 1)$ is accepted, where Σ is the least set containing φ with the required closure properties. To see this, note that by Lemma B.8, $\mathbf{\Omega}^\Sigma$ has depth bounded by $|\Sigma| - 1$, and we know that φ is satisfiable iff it is satisfied on $\mathbf{\Omega}^\Sigma$. Moreover, it is satisfiable on this model iff it occurs at least once in some cluster since every point is contained in a cluster, even if it is a singleton, and thus we set the second parameter to 1. As noted above, we can then take a generated submodel to ensure that φ is satisfied in the root model.

To solve an instance of $(\varphi, b, \Sigma, \mathcal{K}\psi_1, \ldots, \mathcal{K}\psi_m, n)$, we perform the following steps:

1. Choose a cluster type C such that φ occurs at least b times in C, and every $\Psi \in C$ has $\psi_i, \Box\psi_i \in \Psi$ for all $i = 1, \ldots, m$. Accept if C has no defects, and reject if no such C exists.
2. If $n = 0$, reject. Otherwise, let $\psi'_1, \ldots, \psi'_{m'}$ be all formulas such that $\Box\psi'_i \in \bigcup C$.

   (a) For every defect of C of the form $\Diamond\theta$, solve the instance

   $$(\theta, 1, \Sigma, \mathcal{K}\psi'_1, \ldots, \mathcal{K}\psi'_{m'}, n - 1).$$

   (b) For every defect of C of the form $\odot\theta$, solve the instance

   $$(\theta, 2, \Sigma, \mathcal{K}\psi'_1, \ldots, \mathcal{K}\psi'_{m'}, n - 1).$$

3. Reject if any of the above instances rejects, otherwise accept.

It can be verified by induction on n that $(\varphi, b, \Sigma, \mathcal{K}\psi_1, \ldots, \mathcal{K}\psi_m, n)$ has an accepting computation iff $\varphi \wedge \mathcal{K}\psi_1 \wedge \ldots \wedge \mathcal{K}\psi_m$ is satisfiable on a model of depth $\leq n$ where φ occurs at least b times on the root; indeed, we are simply building a model step-by-step. So it remains to check that the algorithm can be implemented in polynomial space.

Since each Σ-type is $O(|\Sigma|)$ in size and each Σ cluster-type has at most $2|\Sigma|$ elements, we need $O(2|\Sigma| \cdot |\Sigma|) = O(|\Sigma|^2)$ space to store C. With this in mind, we may prove by induction on n that the algorithm requires $O(n|\Sigma|^2)$: each recursive call in Step 2 uses $O\big((n-1)|\Sigma|^2\big)$, and we may reuse the same space so we do not need additional space for the multiple calls. In addition, we need to store C, which takes $O(|\Sigma|^2)$ space. We may also store the list of defects that have been processed (this takes $O(|\Sigma|)$ space, but can be avoided if we simply treat defects in some pre-established order). Thus we need $O(|\Sigma|^2 + (n-1)|\Sigma|^2) = O(n|\Sigma|^2)$ space, as claimed.

Finally, we extend the algorithm to the full static language $\mathcal{L}$. This is done as follows: first, let Σ_K be the set of formulas ψ such that $K\psi \in \Sigma$. We non-deterministically choose a set $\Pi = \{\psi_1, \ldots, \psi_m\} \subseteq \Sigma_K$ such that $\neg\varphi \notin \Pi$; these will be the formulas of the form $K\psi$ true in our target model. For each $\theta \in \Sigma \setminus \Pi$, we solve the instance $(\theta, 1, \Sigma, \mathcal{K}\psi_1, \ldots, \mathcal{K}\psi_m, |\Sigma|)$, and accept if all such instances are accepted, otherwise reject. This algorithm is correct since we can amalgamate each model of $\theta \wedge \mathcal{K}\psi_1 \wedge \ldots \wedge \mathcal{K}\psi_m$ to obtain a model of $\big(\bigwedge_{\psi \notin \Pi} \hat{K}\psi\big) \wedge \big(\bigwedge_{\psi \in \Pi} K\psi\big)$; this amalgamated model will be a model of φ. It is a PSPACE algorithm, since we only need to store the set Π and the list of $\theta \in \Sigma \setminus \Pi$ that have to be treated, which takes $O(|\Sigma|)$ space, in addition to the space already required to solve each instance $(\theta, 1, \Sigma, \mathcal{K}\psi_1, \ldots, \mathcal{K}\psi_m, |\Sigma|)$, which as we have seen is polynomial.

This finishes the proof of Theorem 3.4.

Data availability

No data was used for the research described in the article.

References

[1] S. Abramsky, Domain theory in logical form, Ann. Pure Appl. Log. 51 (1991) 1–77.
[2] R. Aumann, Backward induction and common knowledge of rationality, Games Econ. Behav. 8 (1995) 6–19.
[3] A. Baltag, N. Bezhanishvili, D. Fernández-Duque, The topological mu-calculus: completeness and decidability, in: Proc. of LICS 36, IEEE Press, 2021, pp. 1–13.
[4] A. Baltag, N. Bezhanishvili, D. Fernández-Duque, The topological mu-calculus: completeness and decidability, J. ACM 70 (2023) 33:1–33:38.
[5] A. Baltag, N. Bezhanishvili, A. Özgün, S. Smets, Justified belief and the topology of evidence, in: Proc. of WoLLIC 2016, Springer, 2016, pp. 83–103.
[6] A. Baltag, N. Bezhanishvili, A. Özgün, S. Smets, A topological approach to full belief, J. Philos. Log. 48 (2019) 205–244.
[7] A. Baltag, N. Gierasimczuk, A. Özgün, A.L.V. Sandoval, S. Smets, A dynamic logic for learning theory, J. Log. Algebr. Methods Program. 109 (2019) 100485.
[8] A. Baltag, N. Gierasimczuk, S. Smets, On the solvability of inductive problems: a study in epistemic topology, in: Proc. of TARK 2015, ENTCS, 2015, pp. 81–98.
[9] J. van Benthem, G. Bezhanishvili, Modal logics of space, in: Handbook of Spatial Logics, Springer, Dordrecht, 2007, pp. 217–298.
[10] P. Blackburn, M. de Rijke, Y. Venema, Modal Logic, Cambridge University Press, 2001.
[11] M. Brecht, A. Yamamoto, Topological properties of concept spaces, Inf. Comput. 208 (2010) 327–340.
[12] A. Chagrov, M. Zakharyaschev, Modal Logic, The Clarendon Press, New York, 1997.
[13] T.Y. Chow, The surprise examination or unexpected hanging paradox, Am. Math. Mon. 105 (1998) 41–51.
[14] L. Esakia, Weak transitivity—a restitution, in: Logical Investigations, No. 8, Moscow, 2001, "Nauka", Moscow, 2001, pp. 244–255 (in Russian).
[15] L. Esakia, Intuitionistic logic and modality via topology, Ann. Pure Appl. Log. 127 (2004) 155–170, Provinces of logic determined.
[16] D. Fernández-Duque, Tangled modal logic for spatial reasoning, in: T. Walsh (Ed.), IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16–22, 2011, IJCAI/AAAI, 2011, pp. 857–862.
[17] K. Fine, Logics containing $K4$. I, J. Symb. Log. 39 (1974) 31–42.
[18] J. Gerbrandy, The surprise examination in dynamic epistemic logic, Synthese 155 (2007) 21–33.
[19] R. Goldblatt, I. Hodkinson, Spatial logic of tangled closure operators and modal mu-calculus, Ann. Pure Appl. Log. 168 (2017) 1032–1090.
[20] E. Goubault, J. Ledent, S. Rajsbaum, A simplicial complex model for dynamic epistemic logic to study distributed task computability, Inf. Comput. (2020).
[21] N. Hall, How to set a surprise exam, Mind 108 (1999) 647–703.
[22] K.T. Kelly, The Logic of Reliable Inquiry, Oxford University Press, 1996.
[23] K. Levi, The solution to the surprise exam paradox, South. J. Philos. 47 (2009) 131–158.
[24] C. Lutz, Complexity and Succinctness of Public Announcement Logic, Association for Computing Machinery, New York, NY, USA, 2006, pp. 137–143.
[25] J.C.C. McKinsey, A. Tarski, The algebra of topology, Ann. Math. 45 (1944) 141–191.
[26] J. McLelland, C. Chihara, The surprise examination paradox, J. Philos. Log. 4 (1975) 71–89.
[27] A. Özgün, Evidence in Epistemic Logic: a Topological Perspective, Ph.D. thesis, ILLC, Univ. of Amsterdam and Univ. of Lorraine, 2017.
[28] R. Parikh, Finite and infinite dialogues, in: Logic from Computer Science, Springer, 1992, pp. 481–497.
[29] M.V. Quine, On a so-called paradox, Mind 62 (1953) 65–67.
[30] R. Sorensen, Recalcitrant variations of the prediction paradox, Australas. J. Philos. 69 (1984) 355–362.
[31] C. Steinsvold, Topological models of belief logics, Ph.D. thesis, City University of New, York, 2007.
[32] S. Vickers, Topology via Logic, Cambridge University Press, Cambridge, 1989.
[33] C. Wright, A. Sudbury, The paradox of the unexpected examination, Australas. J. Philos. 55 (1977) 41–58.