THE REVIEW OF SYMBOLIC LOGIC
Volume 16, Number 2, June 2023

CARNAP'S PROBLEM FOR MODAL LOGIC

DENIS BONNAY
Université Paris Nanterre
and
DAG WESTERSTÅHL
Stockholm University and Tsinghua University, Beijing

Abstract. We take *Carnap's problem* to be to what extent standard consequence relations in various formal languages fix the meaning of their logical vocabulary, alone or together with additional constraints on the form of the semantics. This paper studies Carnap's problem for basic modal logic. Setting the stage, we show that neighborhood semantics is the most general form of compositional possible worlds semantics, and proceed to ask which standard modal logics (if any) constrain the box operator to be interpreted as in relational Kripke semantics. Except when restricted to finite domains, no modal logic characterizes exactly the Kripkean interpretations of $\Box$. Moreover, we show that, in contrast with the case of first-order logic, the obvious requirement of permutation invariance is not adequate in the modal case. After pointing out some known facts about modal logics that nevertheless force the Kripkean interpretation, we focus on another feature often taken to embody the gist of modal logic: locality. We show that invariance under point-generated subframes (properly defined) does single out the Kripkean interpretations, but only among topological interpretations, not in general. Finally, we define a notion of bisimulation invariance—another aspect of locality—that, together with a reasonable closure condition, gives the desired general result. Along the way, we propose a new perspective on normal neighborhood frames as filter frames, consisting of a set of worlds equipped with an accessibility relation, and a free filter at every world.

§1. Carnap's problem and modal logic. The formal apparatus of logics as we know them is twofold: a semantics, which defines a notion of truth in a structure, determining under which conditions something is true, and a syntax, or proof-theory, which defines a relation of logical consequence, determining what follows from what. Semantics and syntax for a given logic share the same language, but, obviously, there is more than that for a pair of them to make a logic. We think of semantics and syntax as complementary perspectives on the same thing. What does this exactly mean? On the one hand, syntax is to match semantics, and the match is typically established by proving correctness and completeness theorems. Those are part of the core results that belong to the metalogical study of a logic. However, semantics should also match syntax. The way we interpret the logical vocabulary does not to come out of nowhere; it should be determined, in some sense, by the relation of logical consequence. This side of the match may be

Received: September 7, 2020.
2020 *Mathematics Subject Classification*: 00A30, 03A05, 03B45.
Key words and phrases: Carnap's problems, modal logic, neighborhood semantics, Kripke semantics, compositionality, locality, permutation invariance, bisimulation.

doi:10.1017/S1755020321000083

established by proving categoricity theorems, showing that the standard interpretation of the logical vocabulary is the unique interpretation guaranteeing correctness.

Early on, as the distinction between semantic and syntactic methods became prominent, Carnap considered such categoricity results as core metalogical results, on a par with correctness and completeness [3]. However, they somehow got forgotten along the way. Logicians usually take interpretations of the logical vocabulary for granted, implicitly vindicating their choice of semantics by its expressive fruitfulness and by the metalogical results it allows for, rather than bothering to inquire whether it is forced by the syntax under certain assumptions.[1] The present paper is devoted to Carnap's problem for modal logic. It is thus part of a more general project aimed at proving categoricity results establishing that standard semantics for given logics are a good match for their syntax.

To get a better sense of the general project, and to assess the specifics of Carnap's problem for modal logic, let us first review how things stand for first-order logic, as studied in previous work [2]. For Carnap's problem to be well-defined, one needs to agree on two things: (i) a semantic framework, which determines the range of possible interpretations, and (ii) a relation of logical consequence, with respect to which correctness is required. In the case of first-order logic, possible interpretations are reasonably determined by the standards of model-theoretic semantics (e.g., quantifiers are considered as second-order predicates) and logical consequence is simply classical consequence for first-order logic. But this does not suffice to ensure categoricity. In the propositional case, Carnap had worried about the fact that some simple non-standard assignments of truth-values to formulas, such as making true all and only tautologies, are compatible with classical consequence. We pointed out in [2] that as long as the semantics is required to be *compositional*, only the standard interpretation of the connectives fits the consequence relation; here we do have categoricity. Moving to predicate logic, however, compositional non-standard interpretations of quantifiers are easy to find, for example, by interpreting quantification as restricted quantification.

Should one conclude from this that the project fails? This would mean that there is an ineliminable arbitrariness at the heart of classical semantics. Rather than embracing this pessimistic conclusion, one may look for supplementary principles aspiring (like compositionality) to the status of semantic universals. Indeed, a categoricity result for first-order logic was proved in [2]: the standard interpretation of connectives and quantifiers is the only model-theoretic interpretation which is compositional, respects classical consequence, and furthermore satisfies logicality in the guise of permutation invariance.

The situation with modal logic is quite different both on the semantic and the syntactic side. First, there is a profusion of semantics for modal logic which may hold equal rights to being good matches to modal consequence relations. Kripke semantics may be the most commonly used semantics for exploring modal logics from K onward. But topological semantics, which interprets S4 and extensions thereof, is a well-studied

[1] There are exceptions. Building on [25], Feferman [7] endeavors to show that, given a certain second-order metatheory in which the inferential behavior of a (generalized) quantifier can be described, the only quantifiers implicitly defined by such a description are those definable in first-order logic. The assumptions in that metatheory can be debated, however; see [1, 6]. Our approach avoids most of these issues since we start from the resulting consequence relation itself, independently of how it is defined.

https://doi.org/10.1017/S1755020321000083 Published online by Cambridge University Press

semantic framework starting with Tarski's work in the late 1930s. In another direction, neighborhood semantics relates each world to a set of sets of worlds, directly specifying which boxed formulas are true there. And just like some topological models, but not all, may be viewed as Kripke models in disguise, neighborhood models are a genuine generalization of Kripke semantics. If needed, they may also be used to provide matches for logics below K. Moreover, even in the case of Kripke semantics, what we get is not a fixed interpretation for $\Box$ like the one we obtain for $\forall$, but rather *parametric* models, where $\Box$ is universal quantification along the accessibility relation.

On the syntactic side as well, things look much more intricate than they did for first-order logic. In the latter case, there is a clear distinction between the logic and theories formalized within the logic. A Carnapian categoricity result for first-order logic should clearly target interpretations of the logical vocabulary which are admissible with respect to classical consequence, disregarding altogether the theories which the first-order language with non-logical symbols may be used to express. By contrast, modal logic is a family of logics rather than a single logic, with systems such as K, KB, or S5 being just as much as modal logics as they are theories in the language of pure modal logic.

Judging from what we have just said, Carnap's problem for modal logic may seem hopeless, like the quest for a unique Grail in a world of many. But this would be conceding too much too soon to the relativity of logical systems. Kripke semantics may fairly be considered as expressing the gist of modal logic as a logic for possible worlds, which is designed to express propositions in an intensional context with a local perspective on the variety of worlds. Kripke semantics interprets modal truth as truth at a world and enforces locality by interpreting $\Box$ as quantification restricted along an accessibility relation. Granted, one may think of modal logic in different ways, say as a logic for space, making topological semantics the most natural interpretation, or as a hyperintensional logic for belief and knowledge, with neighborhood semantics as the handiest tool. But a very natural question is whether Kripkean semantics is forced upon us by a specific understanding of modal logic as the logic of possible worlds, with this understanding embodied by a commitment to specific modal axioms or specific semantic constraints.

All this is perfectly compatible with both the variety of modal logics and the parametric nature of Kripke models. Regarding the first point, the question is how modal axioms contribute to forcing Kripkean interpretations, without questioning the interest of modal logics not strong enough, or stronger than necessary, to do so. As to the second point, standardness of $\Box$ is taken to consist not in its denotation being unique but in it being interpretable as bounded quantification with respect to an accessibility relation on worlds.

Thus, Carnap's problem gets a clearer, if not fully definite, meaning as a question for Kripke semantics and modal logic. What makes Kripke semantics special as a semantics for modal logic? Is it possible to turn common parlance about what is congenial to the Kripkean perspective on modal logic into precise axiomatic or semantic constraints? The aim is not the illusory one of promoting Kripke semantics as the true semantics of modal logic, but rather to reverse engineer Kripke semantics in order to understand how and when it originates from modal consequence relations.

We will proceed with this endeavor in the following way. First, we make the setting precise (Section 2), defining the set of possible interpretations among which Kripkean interpretations are to be singled out, and explaining more carefully how different

https://doi.org/10.1017/S1755020321000083 Published online by Cambridge University Press

from the first-order case the situation of modal logic is (e.g., why requiring invariance under permutation does not close the deal the way it does for predicate logic). We also establish a strengthened version of the known fact that, as a framework for the semantics of $\Box$, possible worlds semantics, properly defined, is essentially the same as neighborhood semantics,

Then we explore two different strategies to get to Kripkean interpretations. The first one (Section 3) consists in looking above K for modal logics that do not admit of neighborhood or topological models, except for those interpretable as Kripke models. The second (Sections 4 and 5) consists in sticking to K as the basic logic for modal languages and looking for semantic constraints which would express what is congenial to the modal perspective on possible worlds. Here the key idea is to explore the characteristic locality of modal logic.

§2. Background notions and results.

2.1. Language, consequence relations, and logics. Throughout this paper we work with the basic modal language $\mathcal{L}$, given by:

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

p belongs to a fixed set *Prop* of propositional letters. Other connectives ($\bot, \top, \rightarrow, ...$) and modal operators ($\Diamond, ...$) are defined as usual. We may identify $\mathcal{L}$ with the set of $\mathcal{L}$-formulas.

A *consequence relation* is a subset $\models$ of $\mathcal{P}(\mathcal{L}) \times \mathcal{L}$ (assumed to have certain closure properties like reflexivity, monotonicity, etc.).

A *modal logic* is a set $\mathsf{L} \subseteq \mathcal{L}$ containing all classical propositional tautologies and closed under the rules of Modus Ponens ($\varphi, \varphi \rightarrow \psi \in \mathsf{L}$ implies $\psi \in \mathsf{L}$) and Uniform Substitution (if $\varphi \in \mathsf{L}$, then $\varphi' \in \mathsf{L}$, where φ' is obtained by uniformly substituting formulas for propositional letters in φ). The letters $\mathsf{L}, \mathsf{L}', ...$ will always stand for modal logics. L is *normal* if $\Box(p \rightarrow q) \rightarrow (\Box p \rightarrow \Box q) \in \mathsf{L}$ and L is closed under Necessitation ($\varphi \in \mathsf{L}$ implies $\Box\varphi \in \mathsf{L}$).

We usually write $\models_{\mathsf{L}} \varphi$ instead of $\varphi \in \mathsf{L}$. Each modal logic L has an associated consequence relation $\models_{\mathsf{L}}$ defined by

$$\Gamma \models_{\mathsf{L}} \varphi \text{ iff there is a finite } \Gamma_0 \subseteq \Gamma \text{ such that } \bigwedge \Gamma_0 \rightarrow \varphi \in \mathsf{L}.$$

With the understanding that $\bigwedge \emptyset = \top$, we have that $\models_{\mathsf{L}} \varphi$ iff $\emptyset \models_{\mathsf{L}} \varphi$, so there is no ambiguity in writing $\models_{\mathsf{L}} \varphi$.

K is the smallest normal modal logic. If ψ is a formula, $\mathsf{K}\psi$ is the smallest normal modal logic containing ψ. With the usual naming of axioms, $\mathsf{S4} = \mathsf{KT4}$, $\mathsf{S5} = \mathsf{KT4B} = \mathsf{KT5}$, etc.

2.2. Compositionality. Abstractly, an *interpretation* assigns a *semantic value* to each *expression* of a language. In the case of $\mathcal{L}$, interpretable expressions are of two categories: *formulas* and *operators* ($\neg, \wedge, \Box$). Usually, the interpretation of the operators is regarded as fixed, but here we are interested precisely in the range of possible interpretations of these, in particular of $\Box$.

Compositionality relies on the fact that complex expressions are generated from *atomic* expressions by syntactic rules. Accordingly, only atomic expressions need to be interpreted; in the case of $\mathcal{L}$, propositional letters and operators. Let a *model* be

https://doi.org/10.1017/S1755020321000083 Published online by Cambridge University Press

a pair $\mathcal{M} = (I, V)$ of an interpretation I of the operators and a *valuation* V of the propositional letters, where the range of V is a set of semantic values, and functions in I are in keeping with the syntactic type of the operators they interpret. We can then say that a *compositional semantics* for $\mathcal{L}$ is a binary function $[\![\cdot]\!].$, recursively assigning to each pair of a formula φ and a model $\mathcal{M}$ a unique value $[\![\varphi]\!]_{\mathcal{M}}$, as follows:

$$\begin{aligned} &[\![p]\!]_{\mathcal{M}} = V(p), \\ &[\![\neg\varphi]\!]_{\mathcal{M}} = I(\neg)([\![\varphi]\!]_{\mathcal{M}}), \\ &[\![\varphi \wedge \psi]\!]_{\mathcal{M}} = I(\wedge)([\![\varphi]\!]_{\mathcal{M}}, [\![\psi]\!]_{\mathcal{M}}), \\ &[\![\Box\varphi]\!]_{\mathcal{M}} = I(\Box)([\![\varphi]\!]_{\mathcal{M}}). \end{aligned} \tag{1}$$

Thus, compositionality tells us that the interpretations of $\neg$, $\wedge$, and $\Box$ must be functions (of the appropriate arity) from formula values to formula values. Note that so far we have said nothing about what these values are.

2.3. Possible worlds semantics. Several compositional semantics have been proposed for $\mathcal{L}$. As explained in Section 1, we focus on the general setting of possible worlds semantics, in which Kripke semantics has a privileged place. More precisely:

DEFINITION 1. A *possible worlds semantics* for $\mathcal{L}$ is a compositional semantics in which every model $\mathcal{M}$ has a *domain* W of 'worlds', and the semantic value $[\![\varphi]\!]_{\mathcal{M}}$ of each formula φ in $\mathcal{M}$ is a *subset* of W. We can then identify $\mathcal{M}$ with a triple (W, I, V), where $I(\neg)$, $I(\wedge)$, and $I(\Box)$ are appropriate functions on subsets of W, and $V(p) \subseteq W$ for each p. Also, *truth* in a model is defined relative to worlds: φ is *true at* w *in* $\mathcal{M}$ if $w \in [\![\varphi]\!]_{\mathcal{M}}$.

As usual, instead of writing $w \in [\![\varphi]\!]_{\mathcal{M}}$, we often write

$$\mathcal{M}, w \models \varphi.$$

So possible worlds semantics, as defined here, exactly captures one crucial feature of modal semantics: *truth relativity*. This need not involve any further structure among worlds; in particular, no relation of accessibility. But another characteristic feature is what we may call *truth locality*: the truth of φ at w need not depend on the whole model, but only on the part of it that can be 'seen' from w. The idea is most directly implemented in *Kripke semantics*, a possible worlds semantics where models come equipped with an accessibility relation R, and the truth of a formula at w only depends on the part of the model that can be reached from w in successive steps via R.

However, truth locality has also been claimed for other forms of possible worlds semantics, such as topological semantics. In Section 4 we explore ways in which the notion of truth locality can be made precise, and to what extent it is characteristic of Kripke semantics.

2.4. Local interpretations. Fix a domain W and consider interpretations I of $\neg, \wedge, \Box$ as above over subsets of W.

DEFINITION 2. *I respects* a consequence relation $\models$ if

$$\Gamma \models \varphi \text{ implies that for all valuations } V \text{ over } W,$$
$$\bigcap_{\psi \in \Gamma} [\![\psi]\!]_{(W,I,V)} \subseteq [\![\varphi]\!]_{(W,I,V)}.$$

If this holds when $\models = \models_{\mathsf{L}}$ for some modal logic L, we also say that *I respects* L.

https://doi.org/10.1017/S1755020321000083 Published online by Cambridge University Press

From now on we restrict attention to interpretations that respect some modal logic.[2] Then we have the following categoricity result:

THEOREM 3 ([2]). *If I respects some modal logic, then $I(\neg)$ and $I(\wedge)$ are the standard functions, i.e., $I(\neg)$ is complement and $I(\wedge)$ is intersection.*[3]

As a result, an interpretation I for $\mathcal{L}$ (over W) only needs to supply the function interpreting $\Box$, so we can identify I with $I(\Box)$. Let us make this official:

DEFINITION 4 (Local interpretations). A *local interpretation* is a pair (W, F), where $F : \mathcal{P}(W) \to \mathcal{P}(W)$.

If $\mathcal{M} = (W, F, V)$, the truth definition (1) now becomes:

$$\begin{aligned} [\![p]\!]_{\mathcal{M}} &= V(p), \\ [\![\neg\varphi]\!]_{\mathcal{M}} &= W - [\![\varphi]\!]_{\mathcal{M}}, \\ [\![\varphi \wedge \psi]\!]_{\mathcal{M}} &= [\![\varphi]\!]_{\mathcal{M}} \cap [\![\psi]\!]_{\mathcal{M}}, \\ [\![\Box\varphi]\!]_{\mathcal{M}} &= F([\![\varphi]\!]_{\mathcal{M}}). \end{aligned} \tag{2}$$

Also, with the understanding that $\bigcap_{\psi\in\emptyset}[\![\psi]\!]_{\mathcal{M}} = W$, we have:

COROLLARY 5. *(W, F) respects a modal logic L iff $\models_{\mathsf{L}} \varphi$ implies that for all valuations V over W, $[\![\varphi]\!]_{(W,F,V)} = W$.*

2.5. Neighborhood frames. Local interpretations as defined above are in fact familiar objects: they are *neighborhood frames*. Usually, a neighborhood frame is given as (W, N), where $N : W \to \mathcal{P}(\mathcal{P}(W))$. But (W, N) can equally well be presented as (W, m_N) where, for $X \subseteq W$, $m_N(X) = \{w \in W : X \in N(w)\}$. Essentially, these are two ways of describing the same object. In the present context, it is natural to start from the local interpretations (W, F), and define the corresponding families by

$$N_F(w) = \{X \subseteq W : w \in F(X)\}.$$

Sometimes the indexed families $N(w)$ of sets are easier to work with; we shall use both formats, whichever is most convenient for the purpose at hand.

The truth definition in neighborhood semantics is just as (2), or, if you will, (2) with the last clause expressed as follows:[4]

$$[\![\Box\varphi]\!]_{\mathcal{M}} = \{w \in W : [\![\varphi]\!]_{\mathcal{M}} \in N_F(w)\}. \tag{3}$$

Neighborhood semantics for modal logic was suggested in [15, 17], studied in [4, 18], and has recently been revived with a number of applications; see [16] for an extensive introduction. The main motivation for preferring neighborhood frames to relational Kripke frames has come from applications (such as deontic and epistemic logic) taken

[2] Which is to say that they respect the smallest modal logic in the sense of Section 2.1, i.e., the set of $\mathcal{L}$-tautologies.

[3] Actually, this holds for any intensional logic based on classical propositional logic.

[4] Sometimes an alternative *monotone* semantics is used, where (3) is replaced by

$$[\![\Box\varphi]\!]_{\mathcal{M}} = \{w \in W : \exists X \in N_F(w)\ X \subseteq [\![\varphi]\!]_{\mathcal{M}}\}. \tag{m}$$

See, for example, [23]. This validates the rule $\varphi \to \psi / \Box\varphi \to \Box\psi$, whereas the 'strict' semantics only validates the weaker $\varphi \leftrightarrow \psi / \Box\varphi \leftrightarrow \Box\psi$. We shall not pursue the monotone neighborhood semantics in this paper.

https://doi.org/10.1017/S1755020321000083 Published online by Cambridge University Press

to require non-normal modal logics (logics in which not all theorems of K are provable). Also, neighborhood frames have been seen as interesting mathematical structures in themselves, for example, in connection with algebraic semantics (Section 2.8 below), and with topological interpretations of modal logic (see Section 4).

Interestingly, as we have just seen, there is a different and more direct reason for turning to neighborhood frames: the simple idea of using the possible worlds format as a compositional formal semantics for the basic modal language leads directly, via Theorem 3, to these structures as interpretations of $\Box$. Thus, the following is not just a slogan but actually a result (recall Definition 1):[5]

$$\textit{possible worlds semantics} \;=\; \textit{neighborhood semantics}.$$

This means that we can help ourselves to known facts and terminology from neighborhood semantics in what follows. For example, Corollary 5 simply says that (W, F) respects L iff L is sound for (W, F) (also written $(W, F) \models \mathsf{L}$).

DEFINITION 6. A neighborhood frame is *normal* if it respects K.

The following fact is well-known.[6]

PROPOSITION 7. *(W, F) is normal iff each $N_F(w)$ is a filter.*

2.6. Kripkean interpretations. Kripke semantics can be seen as a special case of neighborhood semantics. Recall that a *Kripke frame* is a pair (W, R) where $R \subseteq W^2$, and that, with $\mathcal{M} = (W, R, V)$, the truth clause of boxed formulas in Kripke semantics is

$$\mathcal{M}, w \models \Box\varphi \text{ iff for all } v \text{ such that } wRv, \mathcal{M}, v \models \Box\phi,$$

or, equivalently,

$$[\![\Box\varphi]\!]_{\mathcal{M}} = \{w \in W : R(w) \subseteq [\![\varphi]\!]_{\mathcal{M}}\},$$

where $R(w)$ is the set of R-successors of w.

DEFINITION 8. (W, F) is *Kripkean* iff there is a relation $R \subseteq W^2$ such that F is *standard with respect to* R, in the sense that, $\forall X \subseteq W$, $F(X) = \{w \in W : R(w) \subseteq X\}$.

We shall also need the following notion.

DEFINITION 9. For any local interpretation (W, F), Acc_F, the *potential accessibility relation* of (W, F), is defined by: $w \; Acc_F \; v$ iff $v \in \bigcap N_F(w)$.

The next fact is well-known (except for the new terminology).

FACT 10. *The following are equivalent*:

(a) *(W, F) is Kripkean.*

---

[5] That neighborhood semantics is the most general form of possible worlds semantics is a folklore fact; a precise statement is given in [9] (they use 'extensional' instead of 'compositional'). The novelty here, reflected in Definition 1, is that for this conclusion to follow we need not even presuppose, as is usually done, that the Boolean connectives have their standard meaning.

[6] Unless otherwise stated, proofs of results in neighborhood semantics that we describe as 'well-known' can be found in [16], or in [4].

https://doi.org/10.1017/S1755020321000083 Published online by Cambridge University Press

(b) $N_F(w)$ *is a principal filter for every* w.
(c) F *is standard with respect to* Acc_F.

Thus, if (W, F) is Kripkean, there is a *unique* accessibility relation, namely Acc_F, with respect to which F encodes the standard truth clause for boxed formulas. Indeed, it is practically immediate that the following holds.

FACT 11. *If* (W, F) *is Kripkean, then* (W, F) *and* (W, Acc_F) *are modally equivalent, in the sense that for all valuations* V *over* W *and all formulas* φ, $[\![\varphi]\!]_{(W,F,V)} = [\![\varphi]\!]_{(W,Acc_F,V)}$.

This is the precise sense in which Kripke semantics is a special kind of possible worlds/neighborhood semantics. Particular local interpretations of $\Box$ can sometimes be seen as giving particular meanings to "necessary." For example, the function F_{uni}, defined by

$$F_{\text{uni}}(X) = \begin{cases} W & \text{if } X = W \\ \emptyset & \text{otherwise} \end{cases}$$

embodies logical or metaphysical necessity as truth in *all* possible worlds. (The index "uni" is because $Acc_{F_{\text{uni}}}$ is the universal relation on W.)

However, as explained in Section 1, we are not thinking of Carnap's problem for the meaning of $\Box$ as finding constraints that fix a unique interpretation, not even on a given universe. And the Kripkean local interpretation (W, F_{uni}) is strikingly atypical, since the accessibility relation does not restrain the set of worlds accessible from any world. Rather, we are after constraints that force interpretations of $\Box$ in possible worlds semantics to behave just as in Kripke semantics. At this point, we are able to state a precise version of this question:

> **Carnap's question (Local version)**
> To what extent does respecting modal consequence relations or logics, perhaps in conjunction with invariance constraints, force a local interpretation to be Kripkean?

Since K is sound and complete for the class of all Kripke frames, respecting K will obviously be a minimal requirement. Indeed, for *finite* frames, it gives a complete answer.

THEOREM 12. *The finite Kripkean local interpretations are exactly the normal ones.*

Proof. All Kripkean local interpretations respect K, by Fact 11. Conversely, if (W, F) respects K, then each $N_F(w)$ is a filter (Proposition 7), hence a principal filter when W is finite. This means that $\bigcap N_F(w) \in N_F(w)$, from which it easily follows that F is standard with respect to Acc_F, and hence Kripkean (Fact 10). □

So in the finite case, K forces the interpretations of $\neg$, $\wedge$, and $\Box$ to be the standard ones. However, this is very far from true for infinite frames. We have the following familiar negative result.

FACT 13. *The class of Kripkean local interpretations is not modally definable, in the sense that there is no modal logic* L *such that the local interpretations respecting* L *are exactly the Kripkean local interpretations.*

Proof. This can be proved in various ways; here is one. It is well-known that K is sound and complete with respect to both the class $\mathcal{K}$ of all normal neighborhood

https://doi.org/10.1017/S1755020321000083 Published online by Cambridge University Press

frames and the class $\mathcal{K}'$ of all neighborhood frames (W, F) such that each $N_F(w)$ is a principal filter. By Fact 10, $\mathcal{K}'$ is the class of Kripkean local interpretations. Also, $\mathcal{K}'$ is a proper subset of $\mathcal{K}$. Since $\mathcal{K}$ *is* modally definable (by K), it follows that $\mathcal{K}'$ cannot be modally definable. □

Thus, any answer to Carnap's question for □ in general will involve extra semantic constraints.

2.7. Permutation invariance. Could permutation invariance clinch the answer to the local version of Carnap's question, just as it did for the question about the meaning of ∀ in first-order logic? No: we show in this section that in the modal setting, it is much too restrictive.

A permutation π of W lifts in the usual way to higher-type objects over W, in particular to binary relations on W, and to functions from $\mathcal{P}(W)$ to $\mathcal{P}(W)$. If F is such a function, the function $\pi(F)$ is defined, for $X \subseteq W$, by

$$\pi(F)(X) = \pi(F(\pi^{-1}(X))).^{7}$$

DEFINITION 14. (W, F) is *permutation invariant* (PERM) if, for every permutation π of W, $\pi(F) = F$.

For example, (W, F_{uni}) is permutation invariant. Here are the functions in three more permutation invariant local interpretations:

$$F_{\text{id}}(X) = X \qquad (Acc_{F_{\text{id}}} = id_W),$$

$$F_{\text{emp}}(X) = W \qquad (Acc_{F_{\text{emp}}} = \emptyset),$$

$$F_{\text{diff}}(X) = \begin{cases} W & \text{if } X = W \\ \{w\} & \text{if } X = W - \{w\} \\ \emptyset & \text{otherwise} \end{cases} \qquad (w\, Acc_{F_{\text{diff}}}\, w' \Leftrightarrow w \neq w').$$

There isn't much in the literature on the consequences of invariance constraints on modal operators, but two exceptions are [21] and [13]. MacFarlane [13] lists a number of permutation invariant functions from $\mathcal{P}(W)$ to $\mathcal{P}(W)$, among them F_{id} and F_{uni}, as well as the following:

$$F_1(X) = \begin{cases} W & \text{if } X \neq \emptyset, \\ \emptyset & \text{if } X = \emptyset, \end{cases}$$

$$F_2(X) = \begin{cases} W & \text{if } |X| = 5, \\ \emptyset & \text{otherwise.} \end{cases}$$

Note that F_1 and F_2 are not Kripkean: permutation invariance doesn't enforce Kripkeanity.[8] Still, one may ask exactly which functions from $\mathcal{P}(W)$ to $\mathcal{P}(W)$ are permutation invariant. This is answered by a result in [21].[9] van Benthem shows that

[7] As usual, $\pi(Y) = \{\pi(x) : x \in Y\}$, and $\pi^{-1}(X) = \{x : \pi(x) \in X\}$.

[8] Though F_1 is the dual of F_{uni}: $F_1(X) = W - F_{\text{uni}}(W - X)$, so it is a possibility operator. MacFarlane argues that permutation invariance as defined here is a necessary condition for the *logicality* of operators from sets of worlds to sets of worlds, and particular that one should consider invariance under *all* permutations, not only under those which respect a given accessibility relation on W. On the latter point, see Section 5 below.

[9] van Benthem states his result for functions from sets of individuals rather than sets of possible worlds, but the argument is the same.

https://doi.org/10.1017/S1755020321000083 Published online by Cambridge University Press

if F is a function from $\mathcal{P}(W)^n$ to $\mathcal{P}(W)$, then F is permutation invariant if and only if, for all $X_1, \ldots, X_n \subseteq W$, $F(X_1, \ldots, X_n)$ is a Boolean combination of $X_1, \ldots, X_n$. This implies in particular the following.

THEOREM 15 (van Benthem). *A function $F : \mathcal{P}(W) \to \mathcal{P}(W)$ is permutation invariant iff for each $X \subseteq W$, $F(X)$ is one of $\emptyset$, W, X, and $W - X$.*

This shows how restrictive permutation invariance in itself is, as a constraint on operators from sets of worlds to sets of worlds. Very few of the permutation invariant operators are 'reasonable' interpretations of $\Box$, and in particular, very few are Kripkean: we now show exactly which ones. First, a lemma, whose verification is left to the reader.

LEMMA 16. *If (W, F) is permutation invariant and π is a permutation of W, then, for all $w \in W$, we have $\pi(\bigcap N_F(w)) = \bigcap N_F(\pi(w))$.*

THEOREM 17. *The permutation invariant Kripkean local interpretations over W are exactly (W, F_{uni}), (W, F_{id}), (W, F_{emp}), and (W, F_{diff}).*

Proof. First, it is easily checked that these four local interpretations of $\Box$ are permutation invariant and Kripkean. For the converse, let (W, F) be a permutation invariant Kripkean local interpretation. We claim:

$$Acc_F \text{ is permutation invariant.}$$

In other words, for every permutation π of W,

$$w \, Acc_F \, v \;\Leftrightarrow\; \pi(w) \, Acc_F \, \pi(v).$$

To see this, note that we have $v \in \bigcap N_F(w) \Leftrightarrow \pi(v) \in \pi(\bigcap N_F(w)) = \bigcap N_F(\pi(w))$, by Lemma 16. This proves the claim. But there are exactly four permutation invariant binary relations on any universe, namely, the universal relation, the empty relation, the identity relation, and the difference relation. Since we assumed that (W, F) is Kripkean, i.e., that F is standard with respect to Acc_F, the desired conclusion follows. □

These four local interpretations correspond to familiar logics. When $\mathcal{K}$ is a class of neighborhood frames (or a class of Kripke frames), define $Log(\mathcal{K})$, the *logic of* $\mathcal{K}$, as follows:

$$Log(\mathcal{K}) = \{\varphi : \forall \mathcal{F} \in \mathcal{K} \;\; \mathcal{F} \models \varphi\}. \tag{4}$$

$Log(\mathcal{K})$ is always a modal logic. Now let $\mathcal{K}_{\text{uni}}$ be the class of all neighborhood frames of the form (W, F_{uni}), and similarly for $\mathcal{K}_{\text{id}}$, $\mathcal{K}_{\text{emp}}$, and $\mathcal{K}_{\text{diff}}$. Then we have:

$$\begin{aligned} &Log(\mathcal{K}_{\text{uni}}) = \mathrm{S5}, \\ &Log(\mathcal{K}_{\text{id}}) = \mathsf{K}(\Box p \leftrightarrow p) = \mathsf{L}_\circ, \\ &Log(\mathcal{K}_{\text{emp}}) = \mathsf{K}(\Box \bot) = \mathsf{L}_\bullet, \\ &Log(\mathcal{K}_{\text{diff}}) = \mathsf{KB}(p \wedge \Box p \to \Box\Box p) = \mathit{Else}. \end{aligned} \tag{5}$$

Here $\mathsf{L}_\circ$ and $\mathsf{L}_\bullet$ are the two *trivial* normal modal logics: by a theorem of [14], every consistent normal modal logic is a sublogic of one of these, and the only proper extension of each is the inconsistent modal logic $\mathcal{L}$.[10] *Else* is the 'logic of elsewhere'

[10] These logics have many names in the literature; the ones chosen here derive from the fact that $\mathsf{L}_\circ$ is also the logic of the Kripke frame consisting of a single reflexive point (often

https://doi.org/10.1017/S1755020321000083 Published online by Cambridge University Press

from [24]: $\Box\varphi$ is true at w in this logic iff φ is true everywhere else. (The axiomatization above is from [19].)

However, the main lesson to draw from the facts stated in this section, we think, is that permutation invariance is not an appropriate requirement for our local version of Carnap's question. We are interested in constraints forcing a local interpretation to be Kripkean, which means that truth for boxed formulas is defined in a particular way from an accessibility relation. Permutation invariance drastically restricts the choice of accessibility relations, but we were looking for interpretations that are parametric in an *arbitrary* accessibility relation.

So why is permutation invariance appropriate for first-order logic? The answer seems to be that there we *permute individuals* in the domain which $\forall$ quantifies over, and permutation invariance (or rather isomorphism invariance) implements the reasonable idea that logical constants should be *topic-neutral*. In Kripke semantics for modal logic, on the other hand, *permuting worlds* can make a difference, when the accessibility relation is not one of the four relations above. 'World neutrality' makes sense for some notions of necessity, such as logical necessity, but not in general.

2.8. The algebraic perspective. The semantic frameworks mentioned so far can be seen as special cases of a more general algebraic semantics. Let a *modal algebra* be an algebra $\mathfrak{A} = (A, \vee, \wedge, {}^c, 0, 1, f)$, where $(A, \vee, \wedge, {}^c, 0, 1)$ is a Boolean algebra and f is a unary operation on A. $\mathfrak{A}$ is *normal* if f satisfies $f(a \wedge b) = f(a) \wedge f(b)$ and $f(1) = 1$.[11] A *valuation* is now a map v from *Prop* to A, which is extended inductively to a map $\overline{v}$ from all formulas, with the obvious clauses for the Boolean operations, and with $\overline{v}(\Box\psi) = f(\overline{v}(\psi))$. Any (normal) neighborhood frame $\mathcal{F} = (W, F)$ yields a (normal) modal algebra $\mathcal{F}^+ = (\mathcal{P}(W), \cup, \cap, {}^c, \emptyset, W, F)$, which is modally equivalent to $\mathcal{F}$ in the sense that for any valuation V on W (which is a valuation in the algebraic sense on $\mathcal{P}(W)$) we have, for all formulas φ, $[\![\varphi]\!]_{(W,F,V)} = \overline{V}(\varphi)$.

Algebras of the form $\mathcal{F}^+$ trivially have the following two properties, which we formulate for an arbitrary modal algebra $\mathfrak{A}$.

($\mathcal{C}$) *completeness*: if $X \subseteq A$, then the *join* $\bigvee X$ is an element of A.
($\mathcal{A}$) *atomicity*: any non-zero element is above an *atom* (a minimal non-zero element).

Now consider the property

($\mathcal{V}$) *complete multiplicativity*: if $X \subseteq A$ and $\bigwedge\{f(a) : a \in X\} \in A$, then
($*$) $f(\bigwedge X) = \bigwedge\{f(a) : a \in X\}$.

If $\mathfrak{A}$ satisfies $\mathcal{C}$, then $\mathcal{V}$ just says that f distributes over arbitrary meets.

FACT 18. *(W, F) is Kripkean iff for all $\mathcal{U} \subseteq \mathcal{P}(W)$, $F(\bigcap \mathcal{U}) = \bigcap\{F(X) : X \in \mathcal{U}\}$, that is, iff $(W, F)^+$ is completely multiplicative.*

Proof. This is well-known, but here is a proof. If (W, F) is Kripkean, we have $F(X) = \{w : R(w) \subseteq X\}$ for some $R \subseteq W^2$. Complete multiplicativity then follows since

denoted $\circ$), and $\mathsf{L}_\bullet$ is the logic of a single irreflexive point (often denoted $\bullet$). They are trivial in the sense that in each one, every $\mathcal{L}$-formula is equivalent to a $\Box$-free formula.

[11] In the literature, "modal algebra" often means *normal* modal algebra.

https://doi.org/10.1017/S1755020321000083 Published online by Cambridge University Press

$R(w) \subseteq \bigcap \mathcal{U}$ holds iff for all $X \in \mathcal{U}$, $R(w) \subseteq X$. Conversely, given complete multiplicativity, we have in particular $F(X \cap Y) = F(X) \cap F(Y)$, that is, $X \cap Y \in N_F(w)$ iff $X \in N_F(w)$ and $Y \in N_F(w)$, which shows that each $N_F(w)$ is a filter. To show that it is a principal filter, it is enough to show that $\bigcap N_F(w) \in N_F(w)$, but this is immediate from complete multiplicativity. So (W, F) is Kripkean by Fact 10. □

Properties $\mathcal{C}$, $\mathcal{A}$, and $\mathcal{V}$ figure in the *duality* between algebraic and model-theoretic semantics for modal logic, which goes back to [11], in which Stone duality was generalized to Boolean algebras with operators and relational frames. Thomason [20] and Goldblatt [8] applied these techniques to modal logic. Adapting their results to arbitrary neighborhood frames, Došen [5] showed that the map $\mathcal{F} \mapsto \mathcal{F}^+$, together with a map $\mathfrak{A} \mapsto \mathfrak{A}_+$ in the other direction, yields two contravariant functors that establish a *dual equivalence* between the category of (normal, Kripkean) neighborhood frames and bounded morphisms and the category of modal (normal, $\mathcal{CAV}$-) $\mathcal{CA}$-algebras with complete homomorphisms.[12]

Thus, Carnap's question can be rephrased in algebraic terms: To what extent does respecting modal logics (perhaps in conjunction with other constraints) force a (normal) $\mathcal{CA}$-algebra to be completely multiplicative?

Complete multiplicativity (or equivalently, complete additivity) is studied in-depth in [10]. The focus there is on modal *completeness*. A logic L is (sound and) *complete* relative to frames/algebras of a certain kind (Kripke/neighborhood frames, (normal) $\mathcal{CA}$-algebras, etc.) if there is a class $\mathcal{K}$ of such frames/algebras such that $\mathsf{L} = Log(\mathcal{K})$. Holliday and Litak show (among many other things) that there are modal logics incomplete with respect to any class of normal modal $\mathcal{V}$-algebras.

Carnap's question, on the other hand, concerns the role of $\mathcal{V}$ for modal *correspondence*. Define

$$Fr_{\mathrm{nbd}}(\mathsf{L}) = \{(W, F) : (W, F) \models \mathsf{L}\}. \tag{6}$$

A class $\mathcal{K}$ of frames *corresponds* to a modal logic L if $\mathcal{K} = Fr_{\mathrm{nbd}}(\mathsf{L})$; similarly for classes of algebras. We have seen (Fact 13) that the class $\mathcal{KR}_{\mathrm{nbd}}$ of all Kripkean neighborhood frames corresponds to no modal logic; in other words, it is modally undefinable. Dually, while among $\mathcal{CA}$-algebras, the class of normal $\mathcal{CA}$-algebras corresponds to the logic K, the class of $\mathcal{CAV}$-algebras is modally undefinable.

(In)completeness and (un)definability are two sides of the same coin, as is seen from the familiar (antitone) *Galois connection*:

$$\mathcal{K} \subseteq Fr_{\mathrm{nbd}}(\mathsf{L}) \;\Leftrightarrow\; \mathsf{L} \subseteq Log(\mathcal{K}) \tag{GC}$$

(similarly for other kinds of frames or algebras). Although the undefinability of $\mathcal{KR}_{\mathrm{nbd}}$ (or the class of $\mathcal{CAV}$-algebras) is an easy and well-known fact, whereas the incompleteness of certain modal logics relative to normal $\mathcal{V}$-algebras established in [10] solves a long-standing open problem, the algebraic formulation focusing on the role of complete multiplicativity provides an interesting perspective on Carnap's question about the meaning of $\Box$.[13]

[12] That is, $(\mathcal{F}^+)_+ \cong \mathcal{F}$ and $(\mathfrak{A}_+)^+ \cong \mathfrak{A}$. A homomorphism $h : \mathfrak{A} \to \mathfrak{B}$ is complete if $h(\bigwedge X) = \bigwedge_{a \in X} h(a)$. See [16] for the definition of bounded morphisms for neighborhood frames.

[13] Thanks to Wes Holliday and Tadeusz Litak for directing our attention to the algebraic perspective and the role of complete multiplicativity.

https://doi.org/10.1017/S1755020321000083 Published online by Cambridge University Press

2.9. Global interpretations. The function F in a local interpretation/neighborhood frame (W, F) literally interprets $\Box$. The accessibility relation R in a Kripke frame (W, R), on the other hand, is a *parameter* in the interpretation of $\Box$. Each choice of R yields a local Kripkean interpretation function, which we will call F_R: for $X \subseteq W$, define

$$F_R(X) = \{w \in W : R(w) \subseteq X\}. \tag{7}$$

The class of frames of the form (W, F_R) for $R \subseteq W^2$, i.e., the class of Kripkean local interpretations, can be seen as *the* standard global/parametric interpretation. But in general possible worlds semantics, as defined here, there are no accessibility relations in the background. So we shall simply say that a (global) interpretation is any class of neighborhood frames.

A bit of terminology may be helpful. Let us name the above map from Kripke frames to neighborhood frames, as well as a map in the opposite direction:

$$\begin{array}{ll} \text{a.} & nbd(W, R) = (W, F_R); \\ \text{b.} & kr(W, F) = (W, Acc_F). \end{array} \tag{8}$$

Then, with $\mathcal{KR}$ as the class of all Kripke frames, we have, since $Acc_{F_R} = R$ and, when (W, F) is Kripkean, $F = F_{Acc_F}$:

FACT 19.

(a) $\mathcal{KR}_{\text{nbd}} = nbd(\mathcal{KR})$;
(b) $kr(nbd(W, R)) = (W, R)$;
(c) $(W, F) \in \mathcal{KR}_{\text{nbd}} \Leftrightarrow nbd(kr(W, F)) = (W, F)$.

DEFINITION 20 (Interpretations). A (global) *interpretation* is a class $\mathcal{K}$ of neighborhood frames. We identify $\{(W, F)\}$ with (W, F). The *standard interpretation* is $\mathcal{KR}_{\text{nbd}}$. $\mathcal{K}$ is *Kripkean* if $\mathcal{K} \subseteq \mathcal{KR}_{\text{nbd}}$. Also, $\mathcal{K}$ *respects* a modal logic L if each element of $\mathcal{K}$ respects L. (All this agrees with our earlier definitions when $\mathcal{K} = \{(W, F)\}$.)

We have:

$$\mathcal{K} \text{ respects } \mathsf{L} \text{ iff } \mathsf{L} \subseteq Log(\mathcal{K}) \text{ iff } \mathcal{K} \subseteq Fr_{\text{nbd}}(\mathsf{L}). \tag{9}$$

The global version of Carnap's question is thus to what extent respecting modal consequence relations, perhaps in conjunction with invariance constraints, force an interpretation to be Kripkean, and in particular when it is forced to be the standard interpretation. We can now restate Theorem 12 and Fact 13 as follows.

THEOREM 21.

(a) *A class $\mathcal{K}$ of finite frames is Kripkean iff $\mathcal{K}$ respects* K.
(b) *There is no modal logic* L *such that* $Fr_{\text{nbd}}(\mathsf{L})$ *is the standard interpretation*.

Clearly, a *meaning* of $\Box$ cannot be identified with a single local interpretation/neighborhood frame—not even over a given universe. This is clear already in case of Kripke semantics, where the accessibility relation is a parameter. Global interpretations as defined above are better candidates for such meanings.

https://doi.org/10.1017/S1755020321000083 Published online by Cambridge University Press

§3. Logics with only Kripkean interpretations. Since the standard interpretation $\mathcal{KR}_{\text{nbd}}$ is undefinable, it is natural to ask which modal logics force their interpretations to be Kripkean. Let us give the same name to these logics.

DEFINITION 22. A modal logic L is *Kripkean* if $(W, F) \models \mathsf{L}$ implies that (W, F) is Kripkean, that is, if $Fr_{\text{nbd}}(\mathsf{L}) \subseteq \mathcal{KR}_{\text{nbd}}$.

Which modal logics are Kripkean? We state a folklore result.[14] Recall the B axiom,

$$\varphi \to \Box\Diamond\varphi, \tag{B}$$

which on Kripke frames corresponds to *symmetry* of the accessibility relation.

PROPOSITION 23. *The logic* KB *and its extensions are Kripkean*.

Proof. Suppose $(W, F) \models \mathsf{KB}$, and define $G(X) = W - F(W - X)$ (so G interprets $\Diamond$). Note that, by normality, F and G are monotone. Clearly, (W, F) is consistent with B iff, for all $X \subseteq W$,

$$X \subseteq F(G(X)), \text{ or, equivalently, } G(F(X)) \subseteq X. \tag{10}$$

Next, observe that since (W, F) is consistent with KB, we have

$$X \subseteq F(Y) \Leftrightarrow G(X) \subseteq Y. \tag{11}$$

To see this, suppose $X \subseteq F(Y)$. By monotonicity and (10), $G(X) \subseteq G(F(Y)) \subseteq Y$. Conversely, if $G(X) \subseteq Y$, then, similarly, $X \subseteq F(G(X)) \subseteq F(Y)$.

Now take $w \in W$; we need to show that the filter $N_F(w)$ is principal, i.e., that $\bigcap N_F(w) \in N_F(w)$ or, in other words, $w \in F(\bigcap N_F(w))$. Put differently, we must show that $\{w\} \subseteq F(\bigcap N_F(w))$, which, by (11), is equivalent to

$$G(\{w\}) \subseteq \bigcap N_F(w).$$

So take $v \in G(\{w\})$, and suppose $X \in N_F(w)$, i.e., $w \in F(X)$. If we can show $v \in X$ we are done. But $\{w\} \subseteq F(X)$, and so, by monotonicity and (10),

$$G(\{w\}) \subseteq G(F(X)) \subseteq X.$$

Thus, $v \in X$. □

So S5, for example, is Kripkean, whereas S4 is very far from being Kripkean. Is KB best possible in some sense? Given a modal logic L, what is its smallest Kripkean extension? The following theorem is from [12].[15]

THEOREM 24 (Litak). S5 *is the smallest Kripkean normal extension of* S4.

We leave this topic here, but refer to [12] for several further observations and suggestions.

[14] After we had proved a weaker version, Wes Holliday pointed out to us that Proposition 23 is a known fact. It corresponds to an easily proved and more general algebraic fact (see [12]); we give an elementary proof here to make the paper more self-contained.

[15] In response to our conjecture (at a seminar in Berkeley in October 2016) that S5 is the smallest Kripkean normal extension of S4.3, Tadeusz Litak (within a week or so) proved Theorem 24, and wrote the note [12].

https://doi.org/10.1017/S1755020321000083 Published online by Cambridge University Press

§4. Locality as subframe invariance. As we said in Section 2.3, truth locality, in the sense that the truth of a formula at w in $\mathcal{M}$ only depends on the part of $\mathcal{M}$ that can be 'reached' from w is a perfectly clear notion within Kripke semantics. van Benthem and Bezhanishvili [22] claim that it holds in topological semantics too: "[T]opological semantics for the basic modal language is still 'local', not in the sense of binary accessibility, but in being restricted to what is true in open neighborhoods of the current point." (p. 222).

Recall that a *topological frame* is a neighborhood frame of the form (W, int_τ), where τ is a topology on W (a set of *open* subsets of W containing $\emptyset, W$ and closed under finite intersections and arbitrary unions), and int_τ is the interior function of τ: $int_\tau(X) = \cup\{Y \subseteq X : Y \in \tau\}$. Topological semantics is simply neighborhood semantics for topological frames.

FACT 25. *A neighborhood frame is topological iff it respects* S4.

Proof. This is well-known. Indeed, if (W, F) respects S4, one can show that $range(F)$ is a topology on W such that $F = int_{range(F)}$. □

Now, what the above quote means is the following (*ibid.*, p. 224).

FACT 26. *If τ is a topology on W, $\mathcal{M}_\tau = (W, int_\tau, V)$, and $A \subseteq W$ is open, then, for all formulas φ, $[\![\varphi]\!]_{\mathcal{M}_{\tau_A}} = [\![\varphi]\!]_{\mathcal{M}_\tau} \cap A$.*

Here $\mathcal{M}_{\tau_A} = (A, int_{\tau_A}, V_A)$ is the restriction of $\mathcal{M}_\tau$ to A: τ_A is the *subtopology* generated by A, i.e., $\tau_A = \{X \cap A : X \in \tau\}$, and $V_A(p) = V(p) \cap A$.

Two obvious questions, then, are: (a) Does this notion of locality extend beyond topological frames and Kripkean frames? (b) Is there some stronger notion of locality that singles out the Kripkean neighborhood frames?

4.1. Invariance under generated subframes. To answer the first question, we define a notion of *generated subframe* for arbitrary neighborhood frames, which specializes to the usual notion for topological frames and for Kripke frames.[16] Starting with the families of sets format: given (W, N) and $A \subseteq W$, the *subframe*

$$(A, N_A)$$

is defined by letting $N_A(w) = \{X \cap A : X \in N(w)\}$, for $w \in A$.[17]

DEFINITION 27. (A, N_A) is a *generated subframe* of (W, N), in symbols, $(A, N_A) \subseteq_g (W, N)$, if:

$$\text{For all } X \subseteq W \text{ and all } w \in A,\ X \in N(w) \text{ iff } X \cap A \in N_A(w). \tag{12}$$

If we use the functional format instead, one readily checks that given (W, F) and $A \subseteq W$, condition (12) becomes:

$$\text{For all } X \subseteq W,\ F(X) \cap A = F(X \cap A) \cap A. \tag{13}$$

---

[16] We haven't found such a notion in the literature. However, Pacuit [16] defines *disjoint unions* and *p-morphisms* for arbitrary neighborhood frames, and it easy to check that they relate to our notion of generated subframe in expected ways. For example, ρ is a p-morphism from (W, F) to (W', F') iff the image of (W, F) under ρ is a generated subframe of (W', F').

[17] We apologize for the notation; earlier we defined N_F as the indexed families of sets corresponding to a functional frame (W, F). But since F is a function from subsets of W to subsets of W, and A is a subset of W, using 'N_A' as above shouldn't cause confusion.

https://doi.org/10.1017/S1755020321000083 Published online by Cambridge University Press

Define the function F_A from $\mathcal{P}(A)$ to $\mathcal{P}(A)$, for $X \subseteq A$, by

$$F_A(X) = F(X) \cap A.$$

Then, when (12), or (13), holds, (A, F_A) is the functional version of (A, N_A): for $X \subseteq A$ and $w \in A$,

$$w \in F_A(X) \text{ iff } X \in N_A(w).$$

We omit the straightforward verification of the next fact. In (i), the generated subframe relation $(W', R') \subseteq_g (W, R)$ between Kripke frames is the usual one.

FACT 28.

(i) *If* (W, F) *is Kripkean, then* $(A, F_A) \subseteq_g (W, F)$ *iff* $(A, Acc_{F_A}) \subseteq_g (W, Acc_F)$.
(ii) *If* (W, int_τ) *is a topological frame, then* $(A, (int_\tau)_A) \subseteq_g (W, int_\tau)$ *iff* A *is open*.

Next, we have:

PROPOSITION 29. *If* $(A, F_A) \subseteq_g (W, F)$, *then, for any valuation* V *on* W *and any formula* φ, $[\![\varphi]\!]_{(A,F_A,V_A)} = [\![\varphi]\!]_{(W,F,V)} \cap A$.

Proof. Easy induction on φ. Let us check the case $\varphi = \Box\psi$. We have, with $\mathcal{M} = (W, F, V)$ and $\mathcal{M}_A = (A, F_A, V_A)$:

$$\begin{aligned}
[\![\Box\psi]\!]_{\mathcal{M}} \cap A &= F([\![\psi]\!]_{\mathcal{M}}) \cap A \\
&= F([\![\psi]\!]_{\mathcal{M}} \cap A) \cap A \quad \text{(by (13))} \\
&= F_A([\![\psi]\!]_{\mathcal{M}} \cap A) \\
&= F_A([\![\psi]\!]_{\mathcal{M}_A}) \quad \text{(ind. hyp.)} \\
&= [\![\Box\psi]\!]_{\mathcal{M}_A}.
\end{aligned}$$

□

Thus, we see that invariance under generated subframes in Kripke semantics and in topological semantics, in particular Fact 26, are special cases of Proposition 29. In other words, locality in this sense is *not* specific to these special forms of possible worlds semantics, but *holds in general*. If we want to use truth locality to single out the standard interpretation, we must find a stronger notion.

4.2. Rooted subframes. In Kripke semantics, but not, for example, in topological semantics, there is always a *smallest* generated subset containing a given point w. Is this a unique feature of Kripkean frames? We show in this subsection that the answer is Yes, among topological frames, but not in general.

DEFINITION 30. If (W, F) is a neighborhood frame and $w \in W$, define the *rooted* subframe $(W, F)[w] = (W', F_{W'})$ of (W, F), where

$$W' = \bigcap \{X \subseteq W : w \in X \text{ and } (X, F_X) \subseteq_g (W, F)\}.$$

FACT 31. *If* $(W, F)[w] \subseteq_g (W, F)$, *then* $(W, F)[w]$ *is the smallest generated subframe of* (W, F) *containing* w.

Proof. We need to check that if $(A, F_A) \subseteq_g (W, F)$ and $w \in A$, then we have $(W, F)[w] \subseteq_g (A, F_A)$. The verification of this is straightforward. □

It follows that when (W, F) is Kripkean, so F is standard with respect to a relation R, then $(W, F)[w] = nbd((W, R)[w])$, where $(W, R)[w]$ is the usual Kripke subframe

https://doi.org/10.1017/S1755020321000083 Published online by Cambridge University Press

of (W, R) generated by w. But in general, *rooted subframes need not be generated subframes*, even when (W, F) is normal. One may check, for example, that this fails at 0 in a frame based on the natural numbers where the family $N(n)$ is the set of sets containing n for $n \neq 0$, and $N(0)$ is the set of co-finite sets containing 0.

We can use this criterion as one notion of locality.

DEFINITION 32. (W, F) is *strongly local* if for each $w \in W$, $(W, F)[w] \subseteq_g (W, F)$.

Thus, Kripkean frames are strongly local. Here is a partial converse.

THEOREM 33. *Strongly local topological frames are Kripkean*.

Proof. Suppose (W, F) is strongly local, where $F = int_\tau$ for some topology τ. By Fact 28(ii), for $X \subseteq W$,

$$(X, F_X) \subseteq_g (W, F) \text{ iff } X \text{ is open.} \tag{a}$$

It follows that every set in $N_F(w)$ includes an open set in $N_F(w)$ (namely, its interior). Now take any $w \in W$, and let $A^w = \bigcap\{X \subseteq W : w \in X \text{ and } X \text{ is open}\}$. We have:

$$A^w \subseteq \bigcap N_F(w). \tag{b}$$

To see this, take $v \in A^w$ and let X be any set in $N_F(w)$; we must show $v \in X$. Let Y be an open subset Y of X in $N_F(w)$. Then $w \in F(Y) = int_\tau(Y) \subseteq Y$, so $w \in Y$ and therefore, by the definition of A^w, $v \in Y \subseteq X$, and (b) is proved. Next, we claim that for $X \subseteq W$,

$$X \subseteq W \text{ is open iff } X \subseteq F(X). \tag{c}$$

To prove (c), note that by (a), X is open iff for all $Y \subseteq W$, $F(Y) \cap X = F(X \cap Y) \cap X$. With $Y = W$ we get, since $F(W) = int_\tau(W) = W$, that $X = F(X) \cap X$, that is, $X \subseteq F(X)$. For the other direction, recall that $int_\tau(X \cap Y) = int_\tau(X) \cap int_\tau(Y)$. So if $X \subseteq F(X)$, then, for any $Y \subseteq W$, $F(X \cap Y) \cap X = F(X) \cap F(Y) \cap X = F(Y) \cap X$, so X is open. This proves (c).[18] Finally,

$$A^w \in N_F(w). \tag{d}$$

This is because $w \in A^w$ by definition, and A^w is open by (a) and the assumption of strong locality, so we have $A^w \subseteq F(A^w)$ by (c), and thus $w \in F(A^w)$, i.e., $A^w \in N_F(w)$. So (d) holds. But then, since $N_F(w)$ is a filter, it follows by (b) that $\bigcap N_F(w) \in N_F(w)$, which means that $N_F(w)$ is principal. Since w was arbitrary, this implies as before that (W, F) is Kripkean. □

Say that a global interpretation $\mathcal{K}$ is *strongly local* if each $(W, F) \in \mathcal{K}$ is strongly local. The following corollary is immediate.

COROLLARY 35. *If $\mathcal{K}$ is a strongly local class of topological frames, then $\mathcal{K}$ is Kripkean*.

$(W, F)[w]$ is a reasonable notion of point-generated subframe, *provided* (W, F) is strongly local.[19] But the next example shows that strong locality is not in general sufficient to enforce the standard interpretation, even for normal interpretations.

---

[18] As can be seen from the proof, the following more general claim holds:

FACT 34. *If (W, F) is normal, then $(X, F_X) \subseteq_g (W, F)$ iff $X \subseteq F(X)$.*

[19] Another indication of this is the following fact (see note 16), whose proof is straightforward:

https://doi.org/10.1017/S1755020321000083 Published online by Cambridge University Press

EXAMPLE 36. Let $W = \mathbb{N}$, and define $(\mathbb{N}, F)$ via the families $N_F(n)$ as follows:

$$N_F(n) = \{X \subseteq \mathbb{N} : n+1 \in X\}, \text{ for } n > 0,$$
$$N_F(0) = \{X \subseteq \mathbb{N} : 1 \in X \text{ and } X \text{ is co-finite}\}.$$

Each $N_F(n)$ is a filter, so $(\mathbb{N}, F)$ respects K, but $N_F(0)$ is non-principal, so $(\mathbb{N}, F)$ is not Kripkean. The potential accessibility relation is the successor relation: $Acc_F = \{(n, n+1) : n \in \mathbb{N}\}$. Using the fact observed in note 18, one checks that the domain of each $(\mathbb{N}, F)[n]$ is $\{k : k \geq n\}$. In particular, $(\mathbb{N}, F)[0] = (\mathbb{N}, F)$. It is now easy to verify that $(\mathbb{N}, F)$ is strongly local.

§5. Locality as bisimulation invariance. Presumably the most characteristic expression of locality in Kripke semantics is bisimulation invariance. In this section we lift the Kripkean version of bisimulation to non-Kripkean frames, and discuss the conditions under which requiring a certain kind of invariance under this concept forces interpretations to be Kripkean.[20]

5.1. Filter frames. In order to do so, it will prove handy to take yet another perspective on our general semantics for the modal operator. Invariance under Kripkean bisimulation shall require that all that matters to the action of (W, F) be its Kripkean component, the potential accessibility relation hidden in F. The idea can be made more vivid by actually decomposing F into two parts, that accessibility relation on the one hand and a filtering component on the other, corresponding to the special twist that neighborhood semantics adds to Kripke semantics.

Consider interpreting $\Box$ over a set of worlds W by means of a pair (FF, R) where for every w, $FF(w)$ is a free filter over W, and R is an accessibility relation in the usual sense. Recall that a filter is *free* if the intersection of all its members is empty. Such a frame (W, FF, R) interprets $\Box$ by means of the following semantic clause, for $\mathcal{M} = (W, FF, R, V)$:

$$\mathcal{M}, w \models \Box\varphi \text{ iff } R(w) \subseteq [\![\varphi]\!]_{\mathcal{M}} \text{ and } [\![\varphi]\!]_{\mathcal{M}} \in FF(w). \tag{14}$$

The clauses for atomic formulas, negations, and conjunctions are as usual. We shall call such frames *filter interpretations*:[21] formulas of the form $\Box\varphi$ deemed true at a world w are those such that φ is true at worlds reachable from w, just as for standard Kripkean interpretations, and φ satisfies the condition imposed by a local filter; this is the neighborhood twist. On this view, the filter part of the frame has filtering as its only business. Because the information about the accessibility relation is encoded separately,

(i) If (W, F) is strongly local, then it is a p-morphic image of the disjoint union of its rooted subframes $(W, F)[w]$ for $w \in W$.

[20] Requiring invariance with respect to transformations which respect an accessibility relation is in keeping with the conclusions we reached in Section 2.7: one should not require that worlds are indistinguishable. In contrast with [13], we enforce this idea through bisimulations rather than isomorphisms, since the former but not the latter reflects the local perspective of modal evaluation.

[21] Došen [5] uses 'filter frame' for normal neighborhood frames, i.e., in which each $N(w)$ is a filter. The terminology we propose here is different, although closely related, as Fact 37 below shows.

https://doi.org/10.1017/S1755020321000083 Published online by Cambridge University Press

local filters are now required to be free, that is, not to encapsulate a requirement to include any fixed subset of W.

There is a natural correspondence between filter frames (W, FF, R) and the usual normal neighborhood frames (W, N). To get from a filter interpretation (W, FF, R) to its associated neighborhood frame $(W, N) = (W, FF, R)^*$, one simply needs to reconstruct the neighborhood $N(w)$ at every point, by setting, for each $w \in W$, $N(w) = FF(w) + R(w)$, where, for $\mathcal{A} \subseteq \mathcal{P}(W)$ and $B \subseteq W$, $\mathcal{A} + B$ is $\{X \cup B : X \in \mathcal{A}\}$. In the other direction, to get from a neighborhood frame (W, N) to a filter interpretation $(W, FF, R) = (W, N)^o$, one needs, at each $w \in W$, to divide $N(w)$ into two components, an underlying free filter and worlds accessible at w. The latter are simply defined by setting $R(w) = \bigcap N(w)$, so that R is nothing but Acc_F. To get the former, we define $FF(w)$ as $N(w)$ 'set free', by letting $FF(w) = N(w) - \bigcap N(w)$, where $N(w) - \bigcap N(w) = \{A \subseteq W :$ there is $B \in N(w)$ such that $B - \bigcap N(w) \subseteq A\}$.[22] Shifting in this way from neighborhood semantics to filter models and vice versa preserves truth of modal formulas: it is simply a matter of perspective:

FACT 37.

(i) *For any filter frame* (W, FF, R), $(W, FF, R)^*$ *is a modally equivalent normal neighborhood frame.*
(ii) *For any normal neighborhood frame* (W, N), $(W, N)^o$ *is a modally equivalent filter frame and* $((W, N)^o)^* = (W, N)$.

Proof. For (i), we need to prove (a) that $(W, FF, R)^*$ is a normal frame, and (b) that, for a given valuation V and world w, it makes the same formulas true as (W, FF, R).

(i.a) We check that $FF(w) + R(w)$ is a filter. First, it is closed under finite intersections. Let $X \cup R(w)$ and $Y \cup R(w)$ be two sets in $FF(w) + R(w)$ with $X, Y \in FF(w)$. Because $FF(w)$ is a filter, $X \cap Y$ is in $FF(w)$, hence $(X \cap Y) \cup R(w)$ is in $FF(w) + R(w)$. Second, it is upward closed. Let $X \cup R(w)$ be a set in $FF(w) + R(w)$ and $X \cup R(w) \subseteq Y$. Y may be written as $(Y - (R(w) - X)) \cup R(w)$. Again, because $FF(w)$ is a filter, $(Y - (R(w) - X)) \supseteq X$ guarantees that $Y - (R(w) - X)$ is in $FF(w)$, hence $(Y - (R(w) - X)) \cup R(w)$ is in $FF(w) + R(w)$.

(i.b) Given the semantic clause for filter interpretations, it is enough to show that for all $A \subseteq W$, $R(w) \subseteq A$ and $A \in FF(w)$ iff $A \in FF(w) + R(w)$. The direction from left to right is immediate by definition of $FF(w) + R(w)$. From right to left, if $A \in FF(w) + R(w)$, $A = X \cup R(w)$ for some $X \in FF(w)$, and since $FF(w)$ is upward closed, $X \cup R(w)$ is also in $FF(w)$.

For (ii), we must show (a) that $(W, N)^o$ is a filter frame, (b) that, for a given valuation and world, it makes the same formulas true as (W, N), and (c) that transforming it back into a neighborhood frame yields (W, N) itself.

(ii.a) We check that $N(w) - \bigcap N(w)$ is a free filter. First, it is a filter. It is sufficient to show that $N(w) - \bigcap N(w)$ is closed under finite intersections. Take $A, B \in N(w) - \bigcap N(w)$. So there are $A', B' \in N(w)$ such that $A' - \bigcap N(w) \subseteq A$ and $B' - \bigcap N(w) \subseteq B$. By normality, $A' \cap B' \in N(w)$, and $(A' \cap B') - \bigcap N(w) \subseteq A \cap B$, so $A \cap B \in N(w) - \bigcap N(w)$. Second, $N(w) - \bigcap N(w)$ is free. If not, there would be a $w' \in W$ such that $w' \in \bigcap FF(w)$. Thus, $w' \in \bigcap (N(w) - \bigcap N(w))$, and it follows that

[22] Note that what we get by subtracting $\bigcap N(w)$ is a *filter basis* and not a filter, which is why we need to build in upward closure in the definition of $FF(w)$.

https://doi.org/10.1017/S1755020321000083 Published online by Cambridge University Press

$w' \in \bigcap N(w)$. But then, since $W \in N(w)$ and we have $w' \notin W - \bigcap N(w) \in FF(w)$, we get that $w' \notin \bigcap FF(w)$, a contradiction.

(ii.b) Given the semantic clause for filter models, we need to show that $A \in N(w)$ iff $\bigcap N(w) \subseteq A$ and $A \in N(w) - \bigcap N(w)$. The direction from left to right is immediate: if $A \in N(w)$, then $\bigcap N(w) \subseteq A$ and $A - \bigcap N(w) \subseteq A$, so $A \in N(w) - \bigcap N(w)$. The direction from right to left is simple as well. Since $A \in N(w) - \bigcap N(w)$, A is a superset of some set of the form $A' - \bigcap N(w)$ with $A' \in N(w)$. But since $\bigcap N(w) \subseteq A$, A is also a superset of A', hence A is in $N(w)$ by upward closure again.

(ii.c) We need to show that for any $X \subseteq W$, $X \in (N(w) - \bigcap N(w)) + \bigcap N(w)$ iff $X \in N(w)$. By definition, we have $X \in (N(w) - \bigcap N(w)) + \bigcap N(w)$ iff

$$\exists Y, Z \subseteq W[Z \in N(w) \ \& \ Z - \bigcap N(w) \subseteq Y \ \& \ X = Y \cup \bigcap N(w)]. \qquad (*)$$

From left to right, $(*)$ entails that $Z \subseteq X$, so $X \in N(w)$ by closure under supersets. From right to left we can take $Y = Z = X$, since $X \in N(w)$ implies $\bigcap N(w) \subseteq X$. □

What do Kripkean frames amount to in the world of filter frames? (W, FF, R) will semantically behave like a Kripke frame when there is no filtering, that is when the second conjunct $[\![\varphi]\!]_{\mathcal{M}} \in FF(w)$ in the interpretation of $\Box$ by clause (14) is entailed by the first conjunct. More precisely, say that a filter frame (W, FF, R) is *Kripkean* when its neighborhood match $(W, FF, R)^*$ is. Then we have the following.

LEMMA 38. *(W, FF, R) is Kripkean iff for all $w \in W$ and all $X \subseteq W$, if $R(w) \subseteq X$, then $X \in FF(w)$.*

Proof. From right to left, since $X \in FF(w)$ for all $X \supseteq R(w)$, $FF(w) + R(w)$ is the principal filter generated by $R(w)$, therefore, by Fact 10, $(W, FF, R)^*$ is Kripkean. From left to right, since $(W, FF, R)^*$ is Kripkean, $FF(w) + R(w)$ is the principal filter generated by $R(w)$. So if $R(w) \subseteq X$, we have $X \in FF(w) + R(w)$, and thus $X = Y \cup R(w)$ for some $Y \in FF(w)$, which entails that $X \in FF(w)$ by closure under supersets. □

Special cases of Kripkean filter frames are those of the form $(W, c_{\mathcal{P}(W)}, R)$ where $c_{\mathcal{P}(W)} : W \to \mathcal{P}(\mathcal{P}(W))$ is the constant function defined by $c_{\mathcal{P}(W)}(w) = \mathcal{P}(W)$. One might have expected (W, FF, R) to be Kripkean only if $FF = c_{\mathcal{P}(W)}$, but this is not mandatory, since the filtering only matters for propositions true at reachable worlds. Say that two frames (W, FF, R) and (W, FF', R) are *R-equivalent* if, whenever $w \in W$ and $R(w) \subseteq X \subseteq W$, we have $X \in FF(w)$ if and only if $X \in FF'(w)$.

FACT 39. *If (W, FF, R) and (W, FF', R) are R-equivalent, they are modally equivalent.*

Proof. By R-equivalence, we have $R(w) \subseteq X \ \& \ X \in FF(w)$ iff $R(w) \subseteq X \ \& \ X \in FF'(w)$, and the result follows. □

Given a logic L, we can look at all the filter frames which make it valid, and set $Fr_{\mathrm{ff}}(\mathsf{L}) = \{(W, FF, R) : (W, FF, R) \models \mathsf{L}\}$ in keeping with the previous notation for $Fr_{\mathrm{nbd}}(\mathsf{L})$. One gets correspondence results between axioms of modal logic and conditions on FF and R, in the spirit of traditional correspondence results for Kripke frames which relate axioms and conditions on R. Here are a few examples:

FACT 40. *The following correspondence results hold*:

(i) $Fr_{\mathrm{ff}}(\mathsf{KT})$ *is the class of frames (W, FF, R) such that R is reflexive.*
(ii) $Fr_{\mathrm{ff}}(\mathsf{K4})$ *is the class of frames (W, FF, R) such that R is transitive.*

https://doi.org/10.1017/S1755020321000083 Published online by Cambridge University Press

(iii) $Fr_{\mathrm{ff}}(\mathsf{KD})$ *is the class of frames* (W, FF, R) *such that for all* $w \in W$, *if* $FF(w) = \mathcal{P}(W)$, *there is a* v *such that* wRv.

Proof. (i) It must be shown that $(W, FF, R) \models \Box p \rightarrow p$ if and only if R is reflexive. From right to left, assume that $(W, FF, R, V), w \models \Box p$. This says that $V(p) \in FF(w)$ and $R(w) \subseteq V(p)$. Since R is reflexive, we also have wRw and $w \in V(p)$, which guarantees $(W, FF, R, V), w \models p$. From left to right, assume that $(W, FF, R, V), w \models \Box p$ but $(W, FF, R, V), w \not\models p$. We have that $R(w) \subseteq V(p) \in FF(w)$ and $w \notin V(p)$, hence $w \notin R(w)$, establishing that R is not reflexive.

We omit here the proof for (ii), which follows the classical proof for Kripkean models just like the proof for (i).

(iii) We need to show that $(W, FF, R) \models \Box p \rightarrow \Diamond p$ if and only if for all $w \in W$, if $FF(w) = \mathcal{P}(W)$, there is a v such that wRv. From right to left, assume that $(W, FF, R, V), w \models \Box p$. This implies that $V(p) \in FF(w)$ and $R(w) \subseteq V(p)$. If there is an X which is not in $FF(w)$, $W - V(p)$ cannot be in $FF(w)$ since $V(p)$ is, hence $(W, FF, R, V), w \models \Diamond p$. And if not, we have v with wRv, and since $R(w) \subseteq V(p)$, this also ensures that $(W, FF, R, V), w \models \Diamond p$.

We prove the left to right direction by contraposition. Let w be a world in W such that that $FF(w) = \mathcal{P}(W)$ and $R(w) = \emptyset$. Take any $X \subseteq W$ and let V be such that $V(p) = X$. Then $(W, FF, R, V), w \models \Box p$, but since $FF(w) = \mathcal{P}(W)$, $W - V(P)$ is also in $FF(w)$, and by hypothesis, no world is reachable from w, so $(W, FF, R, V), w \not\models \Diamond p$. □

Thus, some correspondence results for filter models are simple rewritings of classical correspondences, as witnessed by the T and 4 axioms. But sometimes they involve a congenial mix of conditions on FF and R, as illustrated by D. It would be interesting to find other axioms with the same behavior as D, and also to characterize the class of modal axioms for which correspondence smoothly transfers as it does for T and 4.

5.2. Bisimulation invariance. A Kripkean bisimulation shall simply be a bisimulation in the usual sense between the underlying Kripkean structures of two models. In the context of filter models (W, FF, R, V), where this underlying Kripkean structure is made explicit as R, the definition is straightforward:

DEFINITION 41. A *k-bisimulation* between two filter models (W, FF, R, V) and (W', FF', R', V') is a non-empty binary relation $E \subseteq W \times W'$ such that whenever wEw', we have:

(i) w and w' satisfy the same proposition symbols;
(ii) if wRv, then there exists a v' in W' such that vEv' and $w'R'v'$;
(iii) if $w'R'v'$, then there exists a v in W such that vEv' and wRv.

The locality of modal logic with respect to Kripke models is often framed in terms of invariance under bisimulation: bisimilar worlds satisfy the same modal formulas, so that only the local features of the structure of the Kripke models, as captured by the zigzag clauses (ii) and (iii) in the definition of a bisimulation, matter to modal satisfaction. How much of Kripkeanity does invariance under bisimulation give us in the present context? In general, k-bisimilarity between filter frames does not guarantee modal equivalence, since the filtering part of those frames may differ widely. Is there a sense in which Kripkean frames are the k-bisimulation invariant part of our general frames?

https://doi.org/10.1017/S1755020321000083 Published online by Cambridge University Press

First, invariance under k-bisimulation is a property of classes of frames. Just as in the usual case, it amounts to requiring that bisimilar models based on frames in the class are modally equivalent:

DEFINITION 42. A class $\mathcal{J}$ of filter frames is *invariant under k -bisimulation* iff for any (W, FF, R), (W', FF', R') in $\mathcal{J}$, and for any valuations V on W and V' on W', if E is a k-bisimulation between (W, FF, R, V) and (W', FF', R', V') and wEw', then w and w' satisfy the same modal formulas.

For a given logic L, our target is the subclass of $Fr_{\mathrm{ff}}(\mathsf{L})$ consisting only of Kripkean filter frames, which we label $Fr_{\mathrm{kff}}(\mathsf{L})$. $Fr_{\mathrm{kff}}(\mathsf{L})$ is invariant under k-bisimulation, but in general, there will be more than one maximal invariant subclass of $Fr_{\mathrm{ff}}(\mathsf{L})$. Taking K for L, one may simply pick a handful of non k-bisimilar non-Kripkean frames, and grow those frames into a maximal k-bisimulation invariant class which will not be comparable with $Fr_{\mathrm{kff}}(\mathsf{K})$. One needs to assume that the Kripkean frames are also there, or invariance cannot do the job. We shall phrase this as a closure condition on classes of frames.

DEFINITION 43. A class $\mathcal{J}$ of filter frames is *closed under k-extension* iff for any (W, FF, R) in $\mathcal{J}$, $(W, c_{\mathcal{P}(W)}, R)$ is also in $\mathcal{J}$.

We are now able to characterize the Kripkean interpretations of a normal modal logic:

THEOREM 44. *For any normal modal logic* L, $Fr_{\mathrm{kff}}(\mathsf{L})$ *is the greatest subclass of* $Fr_{\mathrm{ff}}(\mathsf{L})$ *which is closed under k-extension and invariant under k-bisimulation*.

Proof. $Fr_{\mathrm{kff}}(\mathsf{L})$ is closed under k-extension and invariant under k-bisimulation, so we need to show that any subclass of $Fr_{\mathrm{ff}}(\mathsf{L})$ which is closed under k-extension and invariant under k-bisimulation is a subclass of $Fr_{\mathrm{kff}}(\mathsf{L})$. Let $\mathcal{J}$ be such a class and (W, FF, R) a frame in $\mathcal{J}$. We must show that (W, FF, R) is Kripkean. Take any $w \in W$ and any $X \subseteq W$ such that $R(w) \subseteq X$. By Lemma 38, it is sufficient to prove that $X \in FF(w)$. Now, since $\mathcal{J}$ is closed under k-extension, $(W, c_{\mathcal{P}(W)}, R)$ is also in $\mathcal{J}$. Let V be a valuation on W such that $V(p) = X$. Clearly, the identity relation on W is a k-bisimulation between $(W, c_{\mathcal{P}(W)}, R, V)$ and (W, FF, R, V). Since $R(w) \subseteq X$, we have that $(W, c_{\mathcal{P}(W)}, R), w \models \Box p$. By k-bisimulation invariance, $(W, FF, R), w \models \Box p$ as well, which ensures that $X \in FF(w)$. □

Kripkean logics as discussed in Section 3 can be readily characterized on the same basis:

COROLLARY 45. *A normal modal logic* L *is Kripkean iff* $Fr_{\mathrm{ff}}(\mathsf{L})$ *is invariant under k-extension and k-bisimulation*.

Proof. The direction from left to right is immediate (use Fact 37 (i)). From right to left, if $Fr_{\mathrm{ff}}(\mathsf{L})$ is invariant under k-extension and k-bisimulation, it is the greatest subclass of $Fr_{\mathrm{ff}}(\mathsf{L})$ to be so, hence by Theorem 44, it is $Fr_{\mathrm{kff}}(\mathsf{L})$. □

If $Fr_{\mathrm{ff}}(\mathsf{L})$ were always invariant under k-extension, one could infer from Theorem 44 that a modal logic is Kripkean iff it simply is invariant under k-bisimulation. But as a case in point, $Fr_{\mathrm{ff}}(\mathsf{KD})$ is not closed under k-extension. By Fact 40, D is valid on non-serial frames on condition that worlds without successors come with proper filters, so that closure under k-extension fails.

https://doi.org/10.1017/S1755020321000083 Published online by Cambridge University Press

When does it hold? We end by formulating this question, which as far as we know is open, in terms of ordinary neighborhood frames rather than filter frames:

QUESTION: For which normal modal logics L is it the case that if (W, F) validates L, so does the corresponding Kripke frame (W, Acc_F)?

The logics K, KT, K4 (hence S4), and K5 have this property, as does any normal logic extending KB, by Proposition 23. But it fails, for example, for KD and K45D.

§6. Conclusion. Reverse engineering Kripke semantics is no simple task. For the well-known reason that there are several general semantics which validate K, one cannot hope to directly get Kripke semantics from the basic modal axioms (except for finite frames). Starting there, two routes are open. One may acknowledge this impossibility for K and look above K for modal logics that make it possible to recover Kripke semantics from their axioms. We did not explore this route, except for a few relevant results stated for the occasion. Alternatively, one may look into supplementary semantic constraints that would single out Kripkean models of K. This is the strategy we had successfully followed for first-order logic, and our primary goal has been to adapt it to the modal case.

The distinctive feature of modal logic, as a fragment of first-order logic, is the fact that modal evaluation is local; everything is taking place at a world and at worlds reachable from that world. Accordingly, we have been looking at semantic constraints that are classically taken to express the locality of modal logic. The first such constraint, that we labeled as 'strong locality', consists in the possibility to restrict a frame to its rooted subframes. However, strong locality characterizes Kripkean frames among topological frames, but not among all normal neighborhood frames. Thus, reverse engineering Kripke semantics would still require going above K to, in this case, S4.

Invariance under k-bisimulation, as a property of classes of frames, paves the way for more general results. Interestingly however, it does not suffice on its own, and Theorem 44 uniquely characterizes Kripkean frames in terms of invariance under k-bisimulation and a closure condition that requires Kripkean frames to be there when their non-standard counterparts are. Phrasing this last result proved convenient in the setting of filter frames, a slight change in perspective on neighborhood frames where the accessibility relation is kept distinct from the filtering performed on top of it by normal neighborhood frames. A systematic study of filter frames might provide new perspectives on the model theory of neighborhood semantics as developed by Pacuit [16]: we hope that the correspondence results stated as Fact 40 will be an incentive for further research along these lines.

Acknowledgements We thank Johan van Benthem, Paul Egré, Sebastian Enqvist, Valentin Goranko, Wesley Holliday, and Tadeusz Litak for helpful comments during the writing of this paper. The insightful remarks by an anonymous referee helped us clarify some formulations. The second author's work on the paper was supported by a grant from the Swedish Research Council, no. 2016-02458.

BIBLIOGRAPHY

[1] Bonnay, D., & Speitel, S. (2021). The ways of logicality: Invariance and categoricity. In Sagi, G. and Woods, J., editors. *The Semantic Conception of Logic*. Cambridge: Cambridge University Press, 2021, pp. 41–60.

https://doi.org/10.1017/S1755020321000083 Published online by Cambridge University Press

[2] Bonnay, D., & Westerståhl, D. (2016). Compositionality solves Carnap's problem. *Erkenntnis*, **81**(4), 721–739.

[3] Carnap, R. (1943). *Formalization of Logic*. Studies in Semantics, Vol. 2. Cambridge: Cambridge University Press.

[4] Chellas, B. (1980). *Modal Logic. An Introduction.* Cambridge: Cambridge University Press. Online publication 2012.

[5] Došen, K. (1989). Duality between modal algebras and neighbourhood frames. *Studia Logica,* **48**(2), 219–234.

[6] Engström, F. (2014). Implicitly definable generalized quantifiers. In Kaså, M., editor. *Idées Fixes: A Festschrift Dedicated to Christian Bennet on the Occasion of his 60th Birthday*. Gothenburg: University of Gothenburg, pp. 65–71.

[7] Feferman, S. (2015). Which quantifiers are logical? A combined semantical and inferential criterion. In Zora, A., editor. *Quantifiers, Quantifiers, and Quantifiers: Themes in Logic, Metaphysics, and Language*. New York: Springer, pp. 19–31.

[8] Goldblatt, R. (1974). Metamathematics of Modal Logic. Ph.D. Thesis, Victoria University of Wellington.

[9] Hansson, B., & Gärdenfors, P. (1973). A guide to intensional semantics. In Kanger, S., editor. *Modality, Morality, and Other Problems of Sense and Nonsense; Essays Dedicated to Sören Halldén*. Lund: CWK Gleerups Bokförlag, pp. 151–167.

[10] Holliday, W. H., & Litak, T. (2019) Complete additivity and modal incompleteness. *Review of Symbolic Logic,* **12**(3), 487–535.

[11] Jónsson, B., & Tarski, A. (1951). Boolean algebras with operators. Part I. *American Journal of Mathematics,* **73**(4), 891–939.

[12] Litak, T. (2018). *On a problem of Westerståhl* (manuscript).

[13] MacFarlane, J. (2000). What Does It Mean to Say That Logic Is Formal? Ph.D. Thesis, University of Pittsburgh.

[14] Makinson, D. (1971). Some embedding theorems for modal logic. *Notre Dame Journal of Formal Logic*, **XII**, 252–254.

[15] Montague, R. (1968). Pragmatics. In Klibansky, R., editor. *Contemporary Philosophy: A Survey*. Florence: La Nuova Italia Editrice, pp. 102–122. Reprinted as Chapter 3 in Montague, R. (1974). *Formal Philosophy*. New Haven, CT: Yale University Press.

[16] Pacuit, E. (2017). *Neighborhood Semantics for Modal Logic*. New York: Springer.

[17] Scott, D. (1970). Advice on modal logic. In Lambert, K., editor. *Philosophical Problems in Logic*. Dordrecht: D. Reidel, pp. 143–173.

[18] Segerberg, K. (1971). *An Essay in Classical Modal Logic*. Number 13 in Filosofiska Studier. Uppsala: Department of Philosophy, Uppsala University.

[19] _____, (1980). A note on the logic of elsewhere. *Theoria*, **46**, 183–187.

[20] Thomason, S. K. (1975). Categories of frames for modal logic. *The Journal of Symbolic Logic*, **40**, 439–442.

[21] van Benthem, J. (1989). Logical constants across varying types. *Notre Dame Journal of Formal Logic*, **30**, 315–342.

[22] van Benthem, J., & Bezhanishvili, G. (2007). Modal logics of space. In Aiello, M., Pratt-Hartmann, I., and van Benthem, J., editors. *Handbook of Spatial Logics*, Dordrecht: Springer, pp. 217–298.

[23] van Benthem, J., Bezhanishvili, N., Enqvist, S., & Ju, J. (2017). Instantial neighbourhood logic. *Review of Symbolic Logic*, **10**(1), 116–144.

[24] von Wright, G. H. (1979). A modal logic of place. In Sosa, E., editor. *The Philosophy of Nicholas Rescher*. Dordrecht: D. Reidel, pp. 65–73.

[25] Zucker, J. I. (1978). The adequacy problem for classical logic. *Journal of Philosophical Logic*, **7**, 517–535.

UNIVERSITÉ PARIS NANTERRE
200 AVENUE DE LA RÉPUBLIQUE
92000 NANTERRE, FRANCE
E-mail: denis.bonnay@gmail.com

STOCKHOLM UNIVERSITY
UNIVERSITETSVÄGEN 10 D
106 91 STOCKHOLM, SWEDEN
and
TSINGHUA UNIVERSITY
30 SHUANGQING ROAD
HAIDIAN DISTRICT, BEIJING, CHINA
E-mail: dag.westerstahl@philosophy.su.se

https://doi.org/10.1017/S1755020321000083 Published online by Cambridge University Press