

Christopher Steinsvold

## Some Formal Semantics for Epistemic Modesty

**Abstract.** Given the frequency of human error, it seems rational to believe that some of our own rational beliefs are false. This is the axiom of epistemic modesty. Unfortunately, using standard propositional quantification, and the usual relational semantics, this axiom is semantically inconsistent with a common logic for rational belief, namely **KD45**. Here we explore two alternative semantics for **KD45** and the axiom of epistemic modesty. The first uses the usual relational semantics and bisimulation quantifiers. The second uses a topological semantics and standard propositional quantification. We show the two different semantics validate many of the same formulas, though we do not know whether they validate exactly the same formulas. Along the way we address various philosophical concerns.

**Keywords:** belief; topology; bisimulation quantifiers; derived set; modesty; humility; KD45; formal epistemology; derivative

### 1. Introduction

Using symbols for propositional quantification and interpreting the box of modal logic as rational belief, one can express a type of epistemic modesty using the following sentence:

$$\Box(\exists p)(\Box p \wedge \neg p) \quad \mathbf{A}^{\text{EM}}$$

This is the *Axiom of Epistemic Modesty*. In words: the agent rationally believes that at least one of her rational beliefs is false. Given the frequency of human error, this seems like a reasonable axiom. By adopting it as an axiom, we need not assume that in every possible world we can imagine,  $\mathbf{A}^{\text{EM}}$  is justifiable. Rather, we can assume we are limiting ourselves to those worlds where  $\mathbf{A}^{\text{EM}}$  is justifiable.

This work is a philosophical and technical exploration into two different formal semantics where  $\mathbf{A}^{\text{EM}}$  is valid, and which also validate a common axiom system for rational belief, namely **KD45**. We focus on the single agent case.

The first formal semantics we focus on interprets propositional quantifiers as bisimulation quantifiers. Our understanding of bisimulation quantifiers is indebted to the work of French (2005, 2006a,b). The second semantics is topological and our approach follows (Steinsvold, 2003, 2007, 2008).

Consider the following axioms of **KD45**:

<i>Closure</i>	$\Box(p \rightarrow q) \rightarrow (\Box p \rightarrow \Box q)$	K
<i>Consistency</i>	$\Box p \rightarrow \Diamond p$	D
<i>Positive Introspection</i>	$\Box p \rightarrow \Box \Box p$	4
<i>Negative Introspection</i>	$\Diamond p \rightarrow \Box \Diamond p$	5

To introduce our basic motivation, consider the following axiom,

$$\Box(\Box p \rightarrow p) \qquad \mathbf{Q}$$

which is a theorem of **KD45**. The validity of **Q** corresponds to:

- $(\forall x)(\forall y)(xRy \Rightarrow yRy)$  secondary reflexivity

Now consider the *standard* interpretation of the propositional quantifier (where quantifiers range over all subsets of the set of possible worlds). Under the standard interpretation,  $(\exists p)\varphi$  is true at a world  $w$ , in a model  $M$ , if there is a  $p$ -variant of  $M$ ,  $M'$ , and  $\varphi$  is true at  $w$  in  $M'$ .

The problem which motivates us can be put as follows. Under the standard interpretation of the propositional quantifier and the usual interpretation of the box, the validity of both **Q** and **D** together in a frame is *semantically* inconsistent with  $\mathbf{A}^{\text{EM}}$ .

To see the truth of this, consider the following argument. Let  $F = \langle W, R \rangle$  be any frame in which both **Q** and **D** are valid. Let  $M = \langle W, R, V \rangle$  be any model on this frame, and let  $w$  be any world in  $M$ . By the validity of axiom **D**, for some  $z$  we have  $wRz$ . By the validity of **Q**, we have  $zRz$ . Now if  $\mathbf{A}^{\text{EM}}$  is true at  $w$ , then  $(\exists p)(\Box p \wedge \neg p)$  is true at  $z$ . But this is impossible because we have  $zRz$ .

Thus, our basic motivation is to investigate a formal semantics where **Q**, **D**, and  $\mathbf{A}^{\text{EM}}$  can all be valid together, because the most popular semantics simply will not allow it. In slightly stronger terms, since **KD45** is a

common axiom system for rational belief (and  $\mathbf{Q}$  is a theorem of  $\mathbf{KD45}$ ), we are motivated to investigate any semantics which validates  $\mathbf{KD45}$  and  $\mathbf{A}^{\text{EM}}$ , and we focus on two semantics satisfying this motivation.

The axioms of  $\mathbf{KD45}$  are idealized, and we see no difference in principle between the use of idealized agents in formal epistemology and the use of idealized lines and planes in geometry. Nonetheless, the more down to earth our agents are the better. We are modeling idealized versions of ourselves. With this in mind the appeal of  $\mathbf{A}^{\text{EM}}$  is straightforward. Human agents have been impressively wrong throughout history, even the history of science is littered with false, but rational, beliefs. Meditating on our own case, we should be able to recall many cases, trivial or non-trivial, where we had a rational belief which we later discovered to be false. Considering all this, it seems plain that there is enough evidence to inductively conclude: one of our current rational beliefs is false (though, of course, we do not know which one).<sup>1</sup> Thus,  $\mathbf{A}^{\text{EM}}$  is worthy of exploration as an axiom for rational belief. In the face of all the mistakes humans have made, it seems simply immodest to deny our fallibility by denying  $\mathbf{A}^{\text{EM}}$ .<sup>2</sup>

The first semantics we look at interprets propositional quantifiers as bisimulation quantifiers. Bisimulation quantifiers are often studied in relation to uniform interpolation and were introduced in (Ghilardi and Zawadowski, 1995) and (Visser, 1996), and have roots in the work of Pitts (1992). They have also been studied in relation to the  $\mu$ -calculus by D'agostino and Lenzi (2005). In (French, 2006a) it was shown that the multi-agent version of  $\mathbf{KD45}$  with bisimulation quantifiers is decidable (as well as multi-agent  $\mathbf{K}$ ,  $\mathbf{K4}$ ,  $\mathbf{S4}$ ,  $\mathbf{S5}$ , and  $\mathbf{GL}$ ). In contrast, the monomodal versions of many modal logics with normal propositional quantification are often not so well-behaved (see Fine, 1970). Standard propositional quantification for  $\mathbf{S5}$  with two agents is known to be intertranslatable with full second order logic (first shown by Antonelli and Thomason (2002) and later given a simpler proof in (Kuhn, 2004)), though the single agent case is decidable (see Bull, 1969; Fine, 1970; Kaplan, 1970).

The second semantics is topological, and we interpret the diamond as the derived set, whereas the propositional quantifiers are interpreted

---

<sup>1</sup> Those interested in the myriad ways in which we can be wrong should consult the popular book by Schulz (2010).

<sup>2</sup> Despite the previous argument, Evnine (2001) has produced an interesting argument *against* the rationality of  $\mathbf{A}^{\text{EM}}$ .

in the standard way. As far as we can tell, both semantics validate the same formulas, though we have no proof of this. Also, we've found no valid formula which is epistemically counter-intuitive, though without a completeness proof for a reasonable set of axioms, we cannot be sure there are none.

In keeping with the theme of modesty, we will make clear the limitations of this paper. First and foremost, this paper is a preliminary investigation, and while we do present a number of answers to various questions, there are still more questions (some technical and some philosophical), which we do not have answers to. Second, we are not taking the most general approach. We are only focusing on semantics where the axioms of **KD45** are valid, as opposed to weaker systems. While this is definitely a limitation, we also hope it will bring some simplicity to a topic which seems technically, and philosophically, not so simple.

This paper is organized as follows. In Section 2 we present the basics of bisimulation quantifiers and prove various theorems. In Section 3 we present a certain philosophical account of modesty and explore it. In Section 4 we present a topological semantics, prove various theorems and address certain philosophical concerns, as well as make comparisons and address concerns with a *rival* topological approach to belief, as presented in (Baltag et al., 2018).

## 2. Semantics for bisimulation quantifiers

### 2.1. Preliminary definitions

A *frame*  $F = \langle W, R \rangle$  is a pair, where  $W$  is non-empty and  $R \subseteq W \times W$ . The members of  $W$  are *points* or *worlds*. A valuation,  $V$ , is a function from propositional variables to  $2^W$  (the power set of  $W$ ). A *model*  $M = \langle W, R, V \rangle$  is a frame with a valuation. Where  $M = \langle W, R, V \rangle$ , we often write  $w \in M$  to mean  $w \in W$ .

The set of propositional variables is  $\text{PROP} = \{p_1, p_2, \dots\}$ , and we often use  $p$  and  $q$  as arbitrary members of  $\text{PROP}$ . Our formulas are defined with:

$$\alpha := p \mid \perp \mid (\alpha_1 \rightarrow \alpha_2) \mid \Box\alpha \mid (\exists p)\alpha$$

The other connectives are defined as usual. A particular occurrence of  $p$  in  $\Phi$  is *free* in  $\Phi$  if it is not within the scope of  $(\exists p)$  or  $(\forall p)$  in  $\Phi$ , it is *bound* otherwise. A formula is *closed* if it has no free variables.

The following definition of a  $\Theta$ -bisimulation from (French, 2006a), is a refinement of the classic notion of a bisimulation.

DEFINITION 2.1. Given two models  $M = \langle W, R, V \rangle$ ,  $M' = \langle W', R', V' \rangle$  and  $\Theta \subseteq \text{PROP}$ , a  $\Theta$ -bisimulation between  $M$  and  $M'$  is a non-empty relation  $\sim_\Theta \subseteq W \times W'$  such that for all  $w \in W$ ,  $w' \in W'$ , if  $w \sim_\Theta w'$ , then:

1. for any  $q \notin \Theta$ :  $w \in V(q)$  iff  $w' \in V'(q)$ ;
2. if  $wRz$  then there is some  $z'$  such that  $z \sim_\Theta z'$  and  $w'Rz'$ ;
3. if  $w'Rz'$  then there is some  $z$ ,  $z \sim_\Theta z'$  and  $wRz$ .

If there is a  $\Theta$ -bisimulation between  $M$  and  $M'$  we write  $M \sim_\Theta M'$  and say that  $M$  and  $M'$  are  $\Theta$ -bisimilar.

If  $\Theta$  is empty and  $M \sim_\Theta M'$ , then we have the (usual) definition of a bisimulation—in this case we say  $M$  and  $M'$  are bisimilar and write  $M \sim M'$ . We write  $M \sim_p M'$  instead of  $M \sim_{\{p\}} M'$ . Given two frames  $F$  and  $F'$  which satisfy conditions 2 and 3 of Definition 2.1, we say that  $F$  and  $F'$  are bisimilar and write  $F \sim F'$ . Note that  $M \sim_{\text{PROP}} M'$  is equivalent to saying the underlying frames of  $M$  and  $M'$  are bisimilar.

Just like bisimulations,  $\Theta$ -bisimulations are equivalence relations. The identity relation is a  $\Theta$ -bisimulation (reflexivity), the composition of two  $\Theta$ -bisimulations is a  $\Theta$ -bisimulation (transitivity), and the inverse of a  $\Theta$ -bisimulation is a  $\Theta$ -bisimulation (symmetry).

We introduce the following notation to save space.

Notation 2.1. Define:  $M(w) \sim_\Theta M'(w')$ , to mean:  $w \in M$  and  $w' \in M'$ ;  $M \sim_\Theta M'$  and  $w \sim_\Theta w'$ . We will use this notation in combination with some of the other conventions/definitions already mentioned. For instance, we write  $M(w) \sim_p M'(w')$  instead of  $M(w) \sim_{\{p\}} M'(w')$ . We write  $F(w) \sim F'(w')$  to mean there is a bisimulation between  $F$  and  $F'$  which connects  $w$  and  $w'$ . And so on.

Let  $\mathcal{C}^{\text{stE}}$  represent the class of serial, transitive and Euclidean frames, i.e., frames having the following properties:

- $(\forall x)(\exists y) xRy$  seriality
- $(\forall x)(\forall y)(\forall z)(xRy \ \& \ yRz \Rightarrow xRz)$  transitivity
- $(\forall x)(\forall y)(\forall z)(xRy \ \& \ xRz \Rightarrow yRz)$  Euclidean

We write  $M \in \mathcal{C}^{\text{stE}}$  to mean  $M$  is a model based on a frame in  $\mathcal{C}^{\text{stE}}$ . Definition of truth at a world in a model is as follows:

- $M, w \models p$  iff  $w \in V(p)$ ,
- $M, w \models \perp$  iff  $0 = 1$ ,
- $M, w \models \Phi \rightarrow \Psi$  iff if  $M, w \models \Phi$  then  $M, w \models \Psi$ ,
- $M, w \models \Box\Phi$  iff  $(\forall x)(wRx \Rightarrow M, x \models \Phi)$ ,
- $M, w \models (\exists p)\Phi$  iff  $(\exists M' \in \mathcal{C}^{\text{stE}})(M(w) \sim_p M'(w') \ \& \ M', w' \models \Phi)$ .

Note the use of Notation 2.1 in the clause for  $(\exists p)\Phi$ , as well as the fact that it is defined relative to  $\mathcal{C}^{\text{stE}}$  (and not just any class of frames). We take the definition of the truth clause for  $(\exists p)\Phi$  from (French, 2006a).

We sometimes say  $w$  forces  $\Phi$  to mean:  $\Phi$  is true at world  $w$ .  $\Phi$  is valid in a model iff  $\Phi$  is true at every point in the model.  $\Phi$  is valid in a frame iff  $\Phi$  is valid in every model based on the frame. We write  $\mathcal{C}^{\text{stE}} \models \Phi$  to mean  $\Phi$  is valid in every member of  $\mathcal{C}^{\text{stE}}$ . The validity in  $\mathcal{C}^{\text{stE}}$  of axioms D, 4 and 5 correspond, respectively, to seriality, transitivity, and the Euclidean property.

Note that, because we are assuming  $R$  is transitive, we can give the following simplified definition of a generated submodel (see, e.g., Goldblatt, 1992, p. 10.)

DEFINITION 2.2. Given  $F = \langle W, R \rangle$ , for any  $w \in W$  we put  $W^w := \{w\} \cup \{y \mid wRy\}$  and  $R^w := R \cap (W^w \times W^w)$ . Furthermore, for any valuation  $V$  and any  $q \in \text{PROP}$ , let  $V^w(q) := V(q) \cap W^w$ . Then  $F^w := \langle W^w, R^w \rangle$  is the subframe of  $F$  generated by  $w$  and  $M^w := \langle W^w, R^w, V^w \rangle$  is the submodel of  $M = \langle W, R, V \rangle$  generated by  $w$ .

In the next section we prove various results to be used for later philosophical discussion, and also for later comparison with the second semantics we will explore.

## 2.2. Some results for bisimulation quantifiers

Since we now have bisimulation quantifiers in our object language, one valuable result we need to show is that bisimulations still preserve the truth of our formulas, i.e. if  $M, M' \in \mathcal{C}^{\text{stE}}$  and  $M(w) \sim M'(w')$ , then

$$M, w \models \Phi \quad \text{iff} \quad M', w' \models \Phi.$$

This is Corollary 2.6. Tim French has established this for  $\mathcal{C}^{\text{stE}}$  (and various other classes), and we reproduce the result here, as it is crucial for further steps. Philosophically speaking, the most important results of this section are Theorem 2.11 and Corollary 2.10; these two results will help guide our discussion of modesty in the next section.

We take the following definition from (French, 2006a).

DEFINITION 2.3. Let  $\mathcal{C}$  be a class of frames. Call  $\mathcal{C}$  *amalgamative* if,

- for all  $M, M' \in \mathcal{C}$ , if  $M(w) \sim_{\Theta \cup \Gamma} M'(w')$ , then there is  $M^* \in \mathcal{C}$  such that  $M(w) \sim_{\Theta} M^*(w^*)$  and  $M'(w') \sim_{\Gamma} M^*(w^*)$ .

We need to show that  $\mathcal{C}^{\text{stE}}$  is amalgamative. In (French, 2006b), an elegant construction is used to show that the class of all frames is amalgamative, and it can be used for other classes as well. We use this construction to show  $\mathcal{C}^{\text{stE}}$  is amalgamative, and the proof (which we include for convenience) is essentially the same as French's.

LEMMA 2.4 (French, 2006b).  $\mathcal{C}^{\text{stE}}$  is amalgamative.

PROOF. Let  $M = \langle W, R, V \rangle$  and  $M' = \langle W', R', V' \rangle$ . Assume that  $M, M' \in \mathcal{C}^{\text{stE}}$  and  $M(w) \sim_{\Theta \cup \Gamma} M'(w')$ . Then the trick is to turn the  $\Theta \cup \Gamma$ -bisimulation itself into a model  $M^* \in \mathcal{C}^{\text{stE}}$ , where  $M(w) \sim_{\Theta} M^*(w^*)$  and  $M'(w') \sim_{\Gamma} M^*(w^*)$ . Let  $M^* = \langle W^*, R^*, V^* \rangle$ , where:

- $R^* := \sim_{\Theta \cup \Gamma} := \{ \langle a, a' \rangle \mid a \sim_{\Theta \cup \Gamma} a' \}$ ,
- $\langle a, a' \rangle R^* \langle b, b' \rangle$  iff  $aRb$  and  $a'R'b'$ ,
- $\langle a, a' \rangle \in V^*(p)$  iff either both  $a \in V(p)$  and  $p \notin \Theta$ , or both  $a' \in V'(p)$  and  $p \notin \Gamma$ .

Finally, let  $w^* = \langle w, w' \rangle$ .

Define the relation  $\sim_{\Theta}$  between  $M$  and  $M^*$  with  $a \sim_{\Theta} \langle a, a' \rangle$ . Similarly, define the relation  $\sim_{\Gamma}$  between  $M'$  and  $M^*$  with  $a' \sim_{\Gamma} \langle a, a' \rangle$ . We leave it to the reader to show  $\sim_{\Theta}$  is indeed a  $\Theta$ -bisimulation between  $M$  and  $M^*$  (and similarly that  $\sim_{\Gamma}$  is indeed a  $\Gamma$ -bisimulation between  $M'$  and  $M^*$ ). Since  $w \sim_{\Theta} w^*$  and  $w' \sim_{\Gamma} w^*$  (because  $w^* = \langle w, w' \rangle$ ), we now have:

$$M(w) \sim_{\Theta} M^*(w^*) \text{ and } M'(w') \sim_{\Gamma} M^*(w^*).$$

We still need to show  $M^* \in \mathcal{C}^{\text{stE}}$ . To see  $M^*$  is serial assume  $\langle a, a' \rangle \in M^*$ . Since  $R$  is serial,  $aRb$  for some  $b$ . Since  $a \sim_{\Theta \cup \Gamma} a'$ , there is some  $b', a'R'b'$  and  $b \sim_{\Theta \cup \Gamma} b'$ . Thus,  $\langle b, b' \rangle \in M^*$ , and thus, by the definition of  $R^*$ ,  $\langle a, a' \rangle R^* \langle b, b' \rangle$ .

Assume  $\langle a, a' \rangle R^* \langle b, b' \rangle$  and  $\langle b, b' \rangle R^* \langle c, c' \rangle$ . By the definition of  $R^*$ ,  $aRb$  and  $bRc$  and  $a'R'b'$  and  $b'R'c'$ . Since both  $R$  and  $R'$  are transitive,  $aRc$  and  $a'R'c'$ , and thus  $R^*$  is transitive as well. Similarly, that  $R^*$  is Euclidean follows from the fact that  $R$  and  $R'$  are both Euclidean.

Thus,  $M^* \in \mathcal{C}^{\text{stE}}$ . ⊣

The following is a narrow version of the more general Lemma 2.31 in (French, 2006a). The more general Lemma applies to any amalgamative class, whereas here we apply it only to  $\mathcal{C}^{\text{stE}}$ . The proof is essentially the same (and we include it for convenience).

LEMMA 2.5 (French, 2006a). *Assume  $M, M' \in \mathcal{C}^{\text{stE}}$ ,  $M(w) \sim_{\Theta} M'(w')$ , and that for all  $p \in \Theta$ ,  $p$  does not occur free in  $\Phi$ . Then:*

$$M, w \models \Phi \text{ iff } M', w' \models \Phi.$$

PROOF. The proof is by induction on the complexity of formulas. We show for the case of the existential quantifier as the other cases are straightforward. Assume the hypothesis and  $M', w' \models (\exists p)\Psi$ . Thus, for some  $M^* \in \mathcal{C}^{\text{stE}}$  we have  $M'(w') \sim_p M^*(w^*)$  and  $M^*, w^* \models \Psi$ . By assumption  $M(w) \sim_{\Theta} M'(w')$ . Linking  $\sim_{\Theta}$  and  $\sim_p$  together yields:  $M(w) \sim_{\Theta \cup \{p\}} M^*(w^*)$ .

By Lemma 2.4, there is  $M^{\nabla} \in \mathcal{C}^{\text{stE}}$  such that:  $M(w) \sim_p M^{\nabla}(w^{\nabla})$  and  $M^*(w^*) \sim_{\Theta \setminus \{p\}} M^{\nabla}(w^{\nabla})$ . By assumption, no free variable of  $(\exists p)\Psi$  occurs in  $\Theta$ . Though  $p$  may be free in  $\Psi$ , it cannot be in  $\Theta \setminus \{p\}$  (we've subtracted it out). Thus by induction hypothesis (hereafter IH)  $M^{\nabla}, w^{\nabla} \models \Psi$ . And since  $M(w) \sim_p M^{\nabla}(w^{\nabla})$ ,  $M, w \models (\exists p)\Psi$ . The other direction is similar.  $\dashv$

One of the more fundamental results in (French, 2006a) is that the non-modal axioms for propositional quantification are sound in any class of frames which is amalgamative (and thus they are sound in  $\mathcal{C}^{\text{stE}}$ ). This result is non-trivial as there are some classes of frames where vacuous quantification can fail (see French, 2006a)!

From Lemma 2.5, for  $\Theta := \emptyset$  we obtain:

COROLLARY 2.6. *Given  $M, M' \in \mathcal{C}^{\text{stE}}$  and  $M(w) \sim M'(w')$ ,*

$$M, w \models \Psi \text{ iff } M', w' \models \Psi.$$

Since submodels are bisimilar to the models they are generated from (using the identity relation), we have the next result:

COROLLARY 2.7. *If  $M^w$  is the submodel of  $M$  generated by  $w$ , for all  $x \in W^w$ :  $M, x \models \Psi$  iff  $M^w, x \models \Psi$ .*

The next theorem shows that any two points in any two frames in  $\mathcal{C}^{\text{stE}}$  are bisimilar.



LEMMA 2.8. *Let  $F = \langle W, R \rangle$  and  $F' = \langle W', R' \rangle$  be any two frames in  $\mathcal{C}^{\text{stE}}$ , and let  $w \in W$  and  $w' \in W'$ , then*

$$F(w) \sim F'(w').$$

PROOF. Assume the hypothesis. Since  $R$  and  $R'$  are serial,  $W \times W'$  is a bisimulation between  $F$  and  $F'$  which links  $w$  and  $w'$ .

For  $\langle w, w' \rangle \in W \times W'$ , and if  $w'R'z'$ , then let  $z$  be some world which  $w$  bears  $R$  to (there must be one by seriality), clearly  $\langle z, z' \rangle \in W \times W'$ . The other direction is similar.  $\dashv$

The following shows that if a closed formula is true at any point in any model of  $\mathcal{C}^{\text{stE}}$ , then the formula is valid in  $\mathcal{C}^{\text{stE}}$ .

LEMMA 2.9. *For any closed formula  $\Phi$ :*

$$\text{if } \mathcal{C}^{\text{stE}} \not\models \neg\Phi, \text{ then } \mathcal{C}^{\text{stE}} \models \Phi.$$

PROOF. Assume  $\Phi$  is a formula with no free variables. If  $\mathcal{C}^{\text{stE}} \not\models \neg\Phi$  then for some  $w$  in some  $M \in \mathcal{C}^{\text{stE}}$ ,  $M, w \models \Phi$ . Let  $w'$  be a point in any model  $M'$  in  $\mathcal{C}^{\text{stE}}$ .

By Lemma 2.8 there is a bisimulation between the frames of  $M$  and  $M'$  which also connects  $w$  and  $w'$ . Thus (recall that PROP is the entire set of propositional variables), we have:  $M(w) \sim_{\text{PROP}} M'(w')$ . Since there is no free occurrence of any variable in  $\Phi$ , we use Lemma 2.5 (let  $\Theta = \text{PROP}$ ) to conclude  $M', w' \models \Phi$ .

Since  $w'$  was an arbitrary point in an arbitrary model,  $\mathcal{C}^{\text{stE}} \models \Phi$ .  $\dashv$

It is not clear at this point, but the previous Lemma, together with the following corollary, reflects a certain philosophical account of modesty (discussed in the next section).

COROLLARY 2.10. *For any closed  $\Phi$  formula:*

$$\mathcal{C}^{\text{stE}} \models \Box\Phi \text{ iff } \mathcal{C}^{\text{stE}} \models \Phi$$

PROOF. The right to left direction follows by Necessitation. For the other direction assume  $\mathcal{C}^{\text{stE}} \models \Box\Phi$ . By axiom D, this implies that  $\Phi$  is true at some point in some model, but then by Lemma 2.9,  $\Phi$  is valid in  $\mathcal{C}^{\text{stE}}$ .  $\dashv$

To be sure, in the following Theorem, and throughout this entire article, ‘**KD45**’ refers to normal **KD45** (without any quantifiers).

THEOREM 2.11. *If  $\Phi$  is a formula with no quantifiers and  $p_i, \dots, p_j$  is a list of the propositional variables in  $\Phi$ , then*

$$\mathbf{KD45} \not\vdash \Phi \text{ iff } \mathcal{C}^{\text{stE}} \models (\exists p_i) \dots (\exists p_j) \neg \Phi.$$

PROOF. Assume  $\Phi$  is a formula with no quantifiers and  $p_i, \dots, p_j$  are all the propositional variables in  $\Phi$ . Since  $\mathbf{KD45}$  is complete with respect to  $\mathcal{C}^{\text{stE}}$ , if  $\mathbf{KD45} \not\vdash \Phi$ , then for some  $M \in \mathcal{C}^{\text{stE}}$  and some  $w \in M$ ,  $M, w \models \neg \Phi$ . The identity relation on  $M$  is a  $p$ -bisimulation from  $M$  to  $M$ , for any  $p$ , but in particular for  $p_i, \dots, p_j$ . Thus, by the definition of truth in a model,  $M, w \models (\exists p_i) \dots (\exists p_j) \neg \Phi$ . By Lemma 2.9,  $\mathcal{C}^{\text{stE}} \models (\exists p_i) \dots (\exists p_j) \neg \Phi$ .

Conversely, if  $\mathcal{C}^{\text{stE}} \models (\exists p_i) \dots (\exists p_j) \neg \Phi$ , then  $\Phi$  fails in some model in  $\mathcal{C}^{\text{stE}}$ , and since  $\mathbf{KD45}$  is sound for  $\mathcal{C}^{\text{stE}}$ ,  $\mathbf{KD45} \not\vdash \Phi$ .  $\dashv$

### 3. Modesty

Informally, we are interpreting  $\Box \Phi$  with,

*the agent rationally believes  $\Phi$ .*

*Prima facie*, the following fact should seem epistemically curious. Since  $\mathbf{KD45} \not\vdash \Box p \rightarrow p$ , by Theorem 2.11, we obtain:

COROLLARY 3.1.  $\mathcal{C}^{\text{stE}} \models (\exists p)(\Box p \wedge \neg p)$ .

Corollary 3.1 tells us that our rational agent *always* has some false belief, and it is this curious fact we aim to justify. Furthermore, by the validity of Necessitation and Corollary 3.1, we obtain:

COROLLARY 3.2.  $\mathcal{C}^{\text{stE}} \models \Box(\exists p)(\Box p \wedge \neg p)$ .

$\mathbf{A}^{\text{EM}}$ , the axiom of epistemic modesty, is valid in  $\mathcal{C}^{\text{stE}}$ . Now, granting that  $\mathbf{A}^{\text{EM}}$  is a reasonable axiom for rational belief, we now have a solid justification for accepting  $(\exists p)(\Box p \wedge \neg p)$  as well, for consider the following *reductio* style argument.<sup>3</sup>

- |    |   |            |
|----|---|------------|
| 1. | $\Box(\exists p)(\Box p \wedge \neg p)$ | assumption |
| 2. | $\neg(\exists p)(\Box p \wedge \neg p)$ | assumption |

---

<sup>3</sup> MacIntosh (1980) credits Prior in (1971) as an early source of the following derivation. Also, strictly speaking this is an informal derivation, as no formal deductive apparatus had been made clear here. The point goes for the following derivation as well.

- |  |  |
|--|--|
| 3. $(\forall p)(\Box p \rightarrow p)$   | from 2                                     |
| 4. $\Box p \rightarrow p$  | from 3                                     |
| 5. $\Box(\exists p)(\Box p \wedge \neg p) \rightarrow (\exists p)(\Box p \wedge \neg p)$ | $p/(\exists p)(\Box p \wedge \neg p)$ in 4 |
| 6. $(\exists p)(\Box p \wedge \neg p)$   | from 1, 5 and modus ponens                 |
| 7. <i>contradiction</i>  | from 2 and 6                               |

Thus,  $\mathbf{A}^{\text{EM}}$  implies  $(\exists p)(\Box p \wedge \neg p)$ . That is, we obtain:

$$\Box(\exists p)(\Box p \wedge \neg p) \rightarrow (\exists p)(\Box p \wedge \neg p)$$

In English: *if you believe that at least one of your beliefs is false, then at least one of your beliefs is false*. Thus, given the aforementioned argument, we can say: as axioms,  $(\exists p)(\Box p \wedge \neg p)$  and  $\Box(\exists p)(\Box p \wedge \neg p)$  are equivalent. And so Corollary 3.1 is only as philosophically questionable as  $\mathbf{A}^{\text{EM}}$ .

As French observes (2006a), the Barcan formula,  $(\forall p)\Box\Psi \rightarrow \Box(\forall p)\Psi$ , fails in  $\mathcal{C}^{\text{stE}}$ . Not only does Barcan fail, it fails significantly. That is, there is an instance of the Barcan formula where the antecedent is valid in  $\mathcal{C}^{\text{stE}}$  and the negation of the consequent is also valid. For  $(\forall p)\Box(\Box p \rightarrow p)$  is valid in  $\mathcal{C}^{\text{stE}}$ , and  $\Diamond(\exists p)(\Box p \wedge \neg p)$  is valid, by Corollary 3.1 and **D**.

There is a philosophical difference between  $(\forall p)\Box(\Box p \rightarrow p)$ , which is valid in  $\mathcal{C}^{\text{stE}}$ , and  $\Box(\forall p)(\Box p \rightarrow p)$ , whose negation is valid in  $\mathcal{C}^{\text{stE}}$ . The validity of  $(\forall p)\Box(\Box p \rightarrow p)$  tells us that it is rational to believe that any one of our rational beliefs is correct. Philosophically, it seems that the whole point of having sufficient justification is to be able to believe that the belief is correct.  $\Box(\forall p)(\Box p \rightarrow p)$ , on the other hand, tells us it is rational to believe that all our beliefs are correct, which, again, sounds epistemically immodest.

As the derivation above shows, believing  $(\exists p)(\Box p \wedge \neg p)$  is a self-fulfilling prophecy: if you believe some of your beliefs are incorrect, then some of your beliefs are incorrect. Interestingly, there are other principles like this. Assuming the agent's beliefs are consistent (which, by axiom **D**, we are), we can argue for:

$$\Box(\exists p)(p \wedge \neg\Box p) \rightarrow (\exists p)(p \wedge \neg\Box p)$$

Consider the following *reductio* style argument:

- |   |            |
|---|------------|
| 1. $\Box(\exists p)(p \wedge \neg\Box p)$ | assumption |
| 2. $\neg(\exists p)(p \wedge \neg\Box p)$ | assumption |
| 3. $(\forall p)(p \rightarrow \Box p)$    | from 2     |

4.  $p \rightarrow \Box p$  from 3
5.  $(\forall p)(p \rightarrow \Box p) \rightarrow \Box(\forall p)(p \rightarrow \Box p)$   $p/(\forall p)(p \rightarrow \Box p)$  in 4
6.  $\Box(\forall p)(p \rightarrow \Box p)$  from 3, 5 and modus ponens
7.  $\Box\neg(\forall p)(p \rightarrow \Box p)$  from 1
8.  $\Box\perp$  from 6, 7 and regularity
9.  $\neg\Box\perp$  from **D**
10. *contradiction* from 8 and 9

Thus if you believe your beliefs are incomplete, and your beliefs are consistent, then you must be correct about this. Granting that we should wish to have our agents resemble ourselves as much as we can, adopting  $\Box(\exists p)(p \wedge \neg\Box p)$  as an axiom for rational belief seems reasonable, and thus by the aforementioned self-referential argument,  $(\exists p)(p \wedge \neg\Box p)$  is reasonable as well. To be sure, we have:

COROLLARY 3.3.  $\mathcal{C}^{\text{stE}} \models (\exists p)(p \wedge \neg\Box p)$  and  $\mathcal{C}^{\text{stE}} \models \Box(\exists p)(p \wedge \neg\Box p)$ .

PROOF. **KD45**  $\not\models p \rightarrow \Box p$ , so by Theorem 2.11,  $\mathcal{C}^{\text{stE}} \models (\exists p)(p \wedge \neg\Box p)$ . Then, by Necessitation,  $\mathcal{C}^{\text{stE}} \models \Box(\exists p)(p \wedge \neg\Box p)$ .  $\dashv$

Before continuing it will be appropriate to say some informal remarks about the very nature of modesty itself. We've assumed from the start that **A<sup>EM</sup>** reflects some sort of modesty on the part of the agent, and we now attempt to make it clear just how this is a form of modesty.

The general notion of modesty, as a virtue, has received much attention in recent decades, much of it seems to be in reaction to Driver's article (1989). Simply put, Driver's account of modesty involves underestimating one's self. Various authors have disagreed. One of the early responses to Driver's account was given by Flanagan (1990), which we will focus on. Flanagan's account of modesty is a non-overestimation account. Flanagan writes:

According to the nonoverestimation account, the modest person may well have a perfectly accurate sense of her accomplishments and worth but she does not overestimate them. (1990, p. 424)

Richards (1992) has a similar account to Flanagan, though Richards uses the word 'humility' instead of 'modesty.' Like Flanagan, Richards argues against the idea that being modest implies underestimating oneself. He writes:

I have identified humility as having a proper sense of oneself and one's accomplishments. (Richards, 1992, p. 9)

Clearly our object language is too impoverished to completely encapsulate this notion of modesty (or humility), but it can reflect it to an extent. We will call this type of account, represented by Richards and Flannagan, an *accuracy* account of modesty, as the main point seems to be having an accurate awareness of one's qualities, and thus, by implication, not underestimating or overestimating one's qualities.

There are two senses of 'modesty' relevant here. The first sense, which we designate with 'modest<sub>1</sub>' simply means: moderate, limited, or small. When one talks of a 'modest sum of money' one means it in this first sense. The second sense, which we designate with 'modest<sub>2</sub>' is the more philosophically rich notion of modesty, and pertains to the agent's attitude towards her own qualities. For the purpose of current discussion we are assuming that the accuracy account is the proper account of modesty<sub>2</sub>. Thus, to say 'Michael Jordan's ability as a basketball player is modest<sub>1</sub>' is simply false. His ability, or the quality of his play, is excellent, and as such is not modest<sub>1</sub>. But this does not imply Michael Jordan is not modest<sub>2</sub>. As long as he does not overestimate his skill to play basketball, his attitude (towards his skill) may be modest<sub>2</sub>.

In many ways the qualities of our idealized agent are not modest<sub>1</sub>. The agent's knowledge of logical truths is not modest<sub>1</sub> (due to Necessitation), and the quality of the agent's (positive and negative) introspection is not modest<sub>1</sub> either (cf. axioms 4 and 5). Similarly, von Neumann's ability to calculate was not modest<sub>1</sub>. Yet for von Neumann to say 'I am exceptional at calculating' is not necessarily not modest<sub>2</sub>, for it is a correct estimation of his ability to calculate. Alternatively, if one has a certain limitation, then modesty<sub>2</sub> requires that one not overestimate that limitation. With this in mind, consider:

**THEOREM 3.4.** *For all closed  $\Phi$ , the following are all equivalent:*

- (1)  $\mathcal{C}^{\text{stE}} \models \Box\Phi$ ,
- (2)  $\mathcal{C}^{\text{stE}} \models \Phi$ ,
- (3)  $(\exists M \in \mathcal{C}^{\text{stE}})(\exists w \in M) M, w \models \Phi$ ,
- (4)  $(\exists M \in \mathcal{C}^{\text{stE}})(\exists w \in M) M, w \models \Box\Phi$ .

**PROOF.** The equivalence of 1 and 2 comes from Corollary 2.10. The equivalence of 2 and 3 follows from Lemma 2.9. Thus, 3 implies 1. We leave the final case for the reader.  $\dashv$

In so far as various closed  $\Phi$  do reflect some quality of the agent, Theorem 3.4 implies that that the agent will be aware of the quality.

In some cases this may have nothing to do with  $\text{modesty}_2$  at all (e.g., if the closed  $\Phi$  is a mere tautology), but in some cases it clearly does. A prime example of  $\text{immodesty}_2$  would be someone who claims that all of their beliefs are correct, when they in fact have some false belief (a straightforward case of overestimation). Theorem 3.4 ensures this cannot happen. Whatever general qualities that are expressed by a closed  $\Phi$ , Theorem 3.4 ensures that the agent will have accurate awareness of that quality (and thus, not be in a position to overestimate it).

Having some false belief is  $\text{modest}_1$ , because it is a limitation.  $\mathbf{A}^{\text{EM}}$  is  $\text{modest}_2$ , because it is a correct estimate of this limitation. The following reveals another general quality of the agent which is  $\text{modest}_1$ .

Since  $\mathbf{KD45} \not\vdash \diamond p \rightarrow \Box p$ , by Theorem 2.11, we obtain:

**COROLLARY 3.5.**  $\mathcal{C}^{\text{stE}} \models (\exists p)(\diamond p \wedge \diamond \neg p)$

An agent whose rational beliefs satisfy  $(\forall p)(\Box p \vee \Box \neg p)$ , though possible, is far removed from any agent we would normally consider or relate to. Such an agent would have their mind made up about everything. Thus it is a natural limitation on our agent to have the negation of this, as we do with Corollary 3.5, and  $\text{modesty}_2$  should require an awareness of this natural limitation. Appropriately, from Corollary 3.5 and Necessitation we have,

**COROLLARY 3.6.**  $\mathcal{C}^{\text{stE}} \models \Box(\exists p)(\diamond p \wedge \diamond \neg p)$

Using Necessitation and Theorem 2.11 we have:

**COROLLARY 3.7.** *If a quantifier-free formula  $\Phi$  with variables  $p_i, \dots, p_j$  is not a theorem of  $\mathbf{KD45}$ , then  $\mathcal{C}^{\text{stE}} \models \Box(\exists p_i) \dots (\exists p_j) \neg \Phi$ .*

Considering Corollary 3.7, it seems our agent never loses sight of the *types* of propositions which fail. Perhaps Corollary 3.7 may be characterized as a counterpart to logical omniscience. As has long been noted, modal agents always believe all the theorems of the system (by Necessitation). Here, we have something akin to ‘non-theorem omniscience’ (so to speak), and is vaguely reminiscent of the 5 axiom itself.

Looking at Theorem 2.11 and Corollary 3.7 a little closer, what should we say about them? For any non-theorem  $\Phi$  of  $\mathbf{KD45}$  we have that (1)  $(\exists p_i) \dots (\exists p_j) \neg \Phi$  and (2)  $\Box(\exists p_i) \dots (\exists p_j) \neg \Phi$  are valid (where  $p_i, \dots, p_j$  are variables in  $\Phi$ ). To give a general sketch at a possible justification, perhaps the validity of each sentence of sort (1) can be justified by claiming it is a  $\text{modest}_1$  type quality of the agent, being a

limitation. And each sentence of sort (2) can be then justified as being  $\text{modest}_2$ , as the agent has accurate knowledge of this limitation.

This section represents a rather  $\text{modest}_1$  attempt to characterize the agent as being a  $\text{modest}_2$  agent, or even as an agent at all. Perhaps there are valid sentences of  $\mathcal{C}^{\text{stE}}$  which simply make no sense, for any sort of agent. Lacking a completeness proof for a reasonable set of axioms, we cannot rule that out.

## 4. Topological semantics

We now turn towards a different semantics which, as far as we can tell, validates the same formulas as  $\mathcal{C}^{\text{stE}}$ . We return to the standard interpretation of the propositional quantifier (where quantifiers range over sets of possible worlds of the model), but change the usual interpretation of the box. The semantics is topological, but we are not focusing on a spatial interpretation. Rather, we are interpreting the topological semantics in terms of an agent with beliefs. The level of point-set topology involved is basic, and our technical goals are to prove analogues of Corollary 2.10 and Theorem 2.11.

### 4.1. Preliminary definitions

DEFINITION 4.1. For any non-empty set  $X$ , let  $\tau$  be a subset of the power set of  $X$  such that:

1.  $X \in \tau$ ,
2.  $\emptyset \in \tau$ ,
3. If  $\mathcal{F} \subseteq \tau$  and  $\mathcal{F}$  is finite, then  $\bigcap \mathcal{F} \in \tau$ ,
4. If  $\mathcal{F} \subseteq \tau$  then  $\bigcup \mathcal{F} \in \tau$ .

Then  $\tau$  is a *topology* on  $X$  and  $\langle X, \tau \rangle$  is a *topological space*.

Given a topological space  $\langle X, \tau \rangle$ , the members of  $X$  are *points* or *worlds*, and the members of the topology  $\tau$  are *open sets* or *opens*, and we use  $O$  and  $U$  as variables for open sets. A set is *closed* if the complement is open. A *topological model*  $M^\tau = \langle X, \tau, V \rangle$  is a topological space with a valuation.

Truth in a topological model  $M^\tau$  at a point  $w$  is the same as in Section 2 for the propositional variables,  $\perp$ , and the conditional (except that we replace  $\models$  with  $\models^\tau$ ). Given two models on the same topological

space,  $M^{\tau'}$  and  $M^{\tau}$ , call  $M^{\tau'}$  a *p-variant* of  $M^{\tau}$ , if  $M^{\tau'}$  and  $M^{\tau}$  agree on all propositional variables with the possible exception of  $p$ . For the box and existential propositional quantifier,

- $M^{\tau}, w \models^{\tau} \Box\Phi$  iff  $(\exists O)(w \in O \text{ and } (\forall x \in O \setminus \{w\})(M, x \models^{\tau} \Phi))$ ,
- $M^{\tau}, w \models^{\tau} (\exists p)\Phi$  iff there is a  $p$ -variant of  $M^{\tau}$ ,  $M^{\tau'}$ , and  $M^{\tau'}, w \models^{\tau} \Phi$ .

For the box, perhaps the most important aspect to note is that it allows for  $\Box p \wedge \neg p$  to be true at a point. Rational beliefs can be false. A formula  $\Phi$  is *valid in a topological model* iff  $\Phi$  is true at every point in the model. A formula  $\Phi$  is *valid in a topological space* iff  $\Phi$  is valid in every topological model based on the topological space. Where  $\mathcal{C}$  is a class of topological spaces, we write  $\mathcal{C} \models^{\tau} \Psi$  to mean  $\Psi$  is valid in  $\mathcal{C}$ .

The following operator,  $d$ , plays the role of a diamond, dual to the box.

DEFINITION 4.2. Given  $\langle X, \tau \rangle$  and  $A \subseteq X$ , let  $d(A)$  be the set of all points  $w$  such that,

- $(\forall O)(\text{if } w \in O, \text{ then } O \setminus \{w\} \cap A \text{ is non-empty})$ ,

Then  $d(A)$  is the *derived set* of  $A$ .

Again, in modal terms,  $d$  is a diamond. Compare:  $M^{\tau}, w \models^{\tau} \Diamond\Phi$  iff  $(\forall O)(\text{if } w \in O \text{ then } (\exists y)(y \in O \setminus \{w\} \text{ and } M^{\tau}, y \models^{\tau} \Phi))$ . Furthermore,  $d(A)$  is usually written as:  $A'$  and called *the set of limit points of A*. In the spaces we are interested in,  $d(\{x\})$  will be empty, for any point  $x$ .

Notice that **K** is always valid. Moreover, the validity of **D**, **4** and **5** correspond, respectively, to:

- **D**:  $\{x\}$  is never open, for any point  $x$ ,
- **4**:  $d(A)$  is closed, for any set  $A$ ,
- **5**:  $d(A)$  is open, for any set  $A$ .

We now define a topology in which the axioms of **KD45** are valid. Let  $X$  be an infinite set. For any  $A \subseteq X$ , we call  $A$  *co-finite* if  $X \setminus A$  is finite. Let  $\tau^{\text{co}}$  be  $\{O \in 2^X \mid O \text{ is co-finite}\} \cup \{\emptyset\}$ . Then  $\tau^{\text{co}}$  is the *co-finite topology* on  $X$ . It is straightforward to show that  $\tau^{\text{co}}$  is indeed a topology. Firstly,  $\emptyset$  and  $X$  belong to  $\tau^{\text{co}}$ , since  $X$  is co-finite. Secondly, the union of any collection of co-finite sets is also co-finite, and the finite intersection of co-finite sets is also co-finite. It is useful to keep in mind that, for all  $A \in 2^X$ , either  $d(A) = \emptyset$  or  $d(A) = X$ . For if  $A$  is finite, then  $d(A) = \emptyset$ ; and if  $A$  is not finite, then  $d(A) = X$ .



Let  $\mathcal{C}^{\text{co}}$  be the class of all topological spaces of the form  $\langle X, \tau^{\text{co}} \rangle$ , where  $X$  is a *countably* infinite set. We specify that each set of points in  $\mathcal{C}^{\text{co}}$  is countably infinite mainly to simplify the work here. For instance, proving Lemma 4.11 would be more involved if we allowed for any cardinality. We write  $M^\tau \in \mathcal{C}^{\text{co}}$  to mean  $M^\tau$  is based on a topological space in  $\mathcal{C}^{\text{co}}$ . Also, when the context is clear, we will often leave the superscript ‘ $\tau$ ’ off of ‘ $M^\tau$ ’ and just use ‘ $M$ ’ for convenience.

We show: **KD45** is sound and complete with respect to  $\mathcal{C}^{\text{co}}$ .

## 4.2. Topological completeness for KD45

To be sure, the overall purpose of this article is to explore and compare two different semantics for an agent with modest qualities. For the first type of semantics, the bulk of the technical work was shown in Section 2.2. Our main goal, at this point, is to show that the results for  $\mathcal{C}^{\text{stE}}$  in section 2.2 also hold for  $\mathcal{C}^{\text{co}}$ . To do this, we need to show that **KD45**, without propositional quantification, is topologically complete, that is,

$$\mathbf{KD45} \vdash \Psi \text{ iff } \mathcal{C}^{\text{co}} \models^\tau \Psi.$$

We start our proof. Assume  $\mathbf{KD45} \not\vdash \Phi$ . Since **KD45** has the finite model property,  $\Phi$  fails at some world  $w$  in some finite  $M \in \mathcal{C}^{\text{stE}}$ , and so  $\Phi$  fails in the submodel generated by  $w$ , that is,  $M^w, w \not\models \Phi$ , where  $M^w = \langle W^w, R^w, V^w \rangle$ .

Note that  $M^w$  is also finite. Requiring that  $M^w$  be finite is significant because ultimately we will be making infinitely many copies of each world in  $W^w$  (with the possible exception of  $w$  itself). We will then form a topology on this new, infinite set of points, and we want that set to be countably infinite (so that it is a member of  $\mathcal{C}^{\text{co}}$ ). This is not essential for showing completeness, but it is helpful for simplifying our work in the following section.

LEMMA 4.3. *Considering  $M^w$ , we have:*

- (1) *If  $w$  bears  $R^w$  to itself, then  $R^w = W^w \times W^w$ ,*
- (2) *If  $w$  does not bear  $R^w$  to itself, then  $R^w = W^w \times (W^w \setminus \{w\})$ .*

PROOF. For (1), assume  $wR^w w$  and let  $x$  and  $y$  be any worlds in  $W^w$ . By construction,  $w$  bears  $R^w$  to both  $x$  and  $y$ , and so by the Euclidean property,  $xR^w y$ .

For (2), assume  $w \not R^w w$ . No world can relate to  $w$ , because then  $w$  would relate to itself, by the Euclidean property (contradiction). So let  $x$  be any world, and let  $y$  be any world that is not  $w$ . Since  $y$  is not  $w$ ,  $w R^w y$ . If  $x = w$ , then done. If not, then  $w R^w x$ , and so by the Euclidean property,  $x R^w y$ .  $\dashv$

Thus,  $R^w$  is a simple relation on  $W^w$ . Either every world relates to every world, or every world relates to every world except  $w$ . Thus, the only possible non-reflexive world is  $w$ . Also, in our proof of the previous theorem, we only used the Euclidean property, however, we are also using a simplified definition of a submodel, made possible by the fact that our initial model was transitive (see Goldblatt, 1992, Exercise 1.6.1).

In the following definition we introduce the notion of a *quasi-explosion*. We make an infinite number of copies of each reflexive world, and so, applying this construction to  $M^w$ , we are making an infinite number of copies of every world, with the possible exception of  $w$ .

DEFINITION 4.4. Let  $M = \langle W, R, V \rangle$ . For any  $x \in W$ , we put:

$$C(x) := \begin{cases} \{\{x, 0\}\} = \{x_0\} & \text{if } x \not R x, \\ \{x\} \times \mathbb{N} = \{x_0, x_1, x_2, \dots\} & \text{if } x R x, \end{cases}$$

and call  $C(x)$  the *copies* of  $x$ . For all  $x_i, x_j \in C(x)$  call  $x_i, x_j$  *fellow copies*. Let  $W^* = \bigcup \{C(x) \mid x \in W\}$ . For all  $x_i, y_j \in W^*$ , let  $x_i R^* y_j$  if  $x R y$ . For any  $x \in W$ , let  $C(x) \subseteq V^*(p)$  if  $x \in V(p)$ . Call  $M^* = \langle W^*, R^*, V^* \rangle$  the *quasi-explosion* of  $M$ .

Linking the copies back to the originals is a bisimulation (and also a p-morphism) and gives:

LEMMA 4.5. *If  $M^*$  is the quasi-explosion of  $M^w$ ,  $z \in M^w$ , and  $z_j \in C(z)$ ,*

$$M^w, z \models \Psi \text{ iff } M^*, z_j \models \Psi.$$

From the above lemma we have,

COROLLARY 4.6. *If  $M^*$  is the quasi-explosion of  $M^w$ , and  $z_k, z_j$  are fellow copies,*

$$M^*, z_k \models \Psi \text{ iff } M^*, z_j \models \Psi.$$

PROOF. Assume  $M^*, z_k \models \Psi$ . By Lemma 4.5 we have  $M^w, z \models \Psi$ , and applying Lemma 4.5 again we have  $M^*, z_j \models \Psi$ .  $\dashv$

We need,

LEMMA 4.7. *If  $M^*$  is the quasi-explosion of  $M^w$ ,  $M^* \in \mathcal{C}^{\text{stE}}$ .*

PROOF. Assume the hypothesis and assume  $x_i R^* y_j$  and  $y_j R^* z_k$ . Thus  $x R^w y$  and  $y R^w z$ , and because  $R^w$  is transitive,  $x R^w z$ . And so  $x_i R^* z_k$ . Thus  $R^*$  is transitive.

That  $R^*$  is serial and Euclidean is shown similarly.  $\dashv$

If  $\diamond p \rightarrow \Box p$  was valid in the frame of  $M^w$ , it would not necessarily be valid in the frame of  $M^*$ , but it would be valid in the model, by Theorem 4.5. Finally,

THEOREM 4.8. **KD45**  $\vdash \Psi$  iff  $\mathcal{C}^{\text{co}} \models^\tau \Psi$ .

PROOF. The soundness of **KD45** with respect to  $\mathcal{C}^{\text{co}}$  is left to the reader.

Recapitulating our earlier reasoning, we assumed **KD45**  $\not\vdash \Phi$ , and it followed that  $\Phi$  failed in the finite, generated submodel  $M^w \in \mathcal{C}^{\text{stE}}$ , at world  $w$ . Applying Lemma 4.5 we have that  $\Phi$  fails in the quasi-explosion  $M^* = \langle W^*, R^*, V^* \rangle$ , at all the possible copies of  $w$  (note that if  $w$  is not reflexive, then there is just one copy of  $w$ ).

Also, by Lemma 4.7, since  $M^w \in \mathcal{C}^{\text{stE}}$ ,  $M^* \in \mathcal{C}^{\text{stE}}$ .

Here is where seriality comes in.  $W^*$ , of course, is non-empty, but we need to know it is infinite. By the construction in Definition 4.4, we know at least the copies of  $w$  are in  $W^*$ , and if  $w$  is reflexive, then there will be an infinite number of copies in  $W^*$ . But if  $w$  is not reflexive, then there is only one copy of  $w$  (namely  $w_0$ ), but our relation is serial, and so  $w_0$  relates to some world, and by the Euclidean property, that world must relate to itself, and so there will be an infinite number of copies of this second world. Thus,  $W^*$  is infinite.

Let  $\tau^{\text{co}}$  be the co-finite topology on  $W^*$  and let  $M^\tau = \langle W^*, \tau^{\text{co}}, V^* \rangle$ . As mentioned, since we started off with a finite model, and made a countably infinite number of copies of each world (besides possibly  $w$ ),  $W^*$  is countably infinite. Thus  $M^\tau \in \mathcal{C}^{\text{co}}$ .

We need to show:  $M^*, z_j \models \Psi$  iff  $M^\tau, z_j \models^\tau \Psi$ .

The non-modal cases are straightforward. Assume  $M^*, z_j \models \Box \Psi$ . Considering  $M^w$ , if  $w$  bears  $R^w$  to itself, then, by Lemma 4.3, all the worlds in  $W^w$  bear  $R^w$  to each other, and so by the construction of  $M^*$ , all the worlds in  $W^*$  bear  $R^*$  to each other, that is  $R^* = W^* \times W^*$ . And thus all worlds in  $M^*$  force  $\Psi$ . Thus, by the IH, all worlds in  $M^\tau$  force  $\Psi$ . And since  $W^*$  is an open set, and  $z_j$  is in it,  $M^\tau, z_j \models^\tau \Box \Psi$ .

On the other hand, if  $w$  does not bear  $R^w$  to itself, then  $R^w = W^w \times (W^w \setminus \{w\})$ , by Lemma 4.3. And so by the construction of  $M^*$ , we have:  $R^* = W^* \times (W^* \setminus \{w_0\})$ . And to be sure,  $C(w) = \{w_0\}$ . Thus every world in  $M^*$  forces  $\Psi$  with the possible exception of  $w_0$ . Thus, by IH, every world in  $M^\tau$  forces  $\Psi$  with the possible exception of  $w_0$ . Now, since all co-finite sets are open, the following is open:  $(W^* \setminus \{w_0\}) \cup \{z_j\}$ . And  $z_j$  is a member, and so by the definition of truth,  $M^\tau, z_j \models \Box\Psi$ .

Conversely, assume  $M^\tau, z_j \models^\tau \Box\Psi$ . Thus there is some open, co-finite set,  $O$ ,  $z_j \in O$  and every world in  $O \setminus \{z_j\}$  forces  $\Psi$ . By IH, every world in  $O \setminus \{z_j\}$  forces  $\Psi$  in  $M^*$ .

Considering our original world  $w$ , if  $w$  is reflexive then  $w$ , and every other world, has infinitely many copies in  $M^*$ . And since  $O$  is co-finite, then only a finite number of worlds could possibly have  $\Psi$  fail. But then all worlds force  $\Psi$ , by Corollary 4.6. Thus,  $M^*, z_j \models \Box\Psi$ .

On the other hand if  $w$  is not reflexive, then it has only one copy in  $M^*$ , namely  $w_0$ , and  $w_0$  is not reflexive either. Thus  $z_j$  does not relate to  $w_0$ , and all other worlds have infinitely many fellow copies. Since  $O$  is co-finite, all but possibly a finite number of worlds force  $\Psi$  in  $M^*$ , and so by Corollary 4.6 again,  $M^*, z_j \models \Box\Psi$ .

Therefore,  $\Phi$ , our original non-theorem of **KD45**, will fail in  $M^\tau$ .  $\dashv$

Thus, **KD45** is sound and complete for  $\mathcal{C}^{\text{co}}$ . Notice that, this topology won't work for the multi-agent case, as there is only one co-finite topology on a given infinite set—all the agents would be the same. There are, in fact, other topologies which validate the **KD45** axioms, and we mention them later on.

### 4.3. Some results for propositional quantifiers

Here we prove results analogous to the results in Section 2.2. Throughout this section we adopt the following convention: where  $M$  is a model, we will typically write:  $M^p$  to designate some  $p$ -variant of  $M$ . Taking it further, we may write:  $M^{pq}$  to specify a  $q$ -variant of  $M^p$ , which is itself a  $p$ -variant of  $M$ .

In the following we show that if a model and its  $p$ -variant only differ on  $p$  at finitely many points, then the points not in that finite set all force exactly the same formulas in both models.

LEMMA 4.9. *Let  $M \in \mathcal{C}^{\text{co}}$  and let  $F$  be a finite set of points in  $M$ . Let  $M^p$  (with valuation  $V^p$ ) be a  $p$ -variant of  $M$  such that for all  $x \notin F$ ,*

$$x \in V(p) \text{ iff } x \in V^p(p),$$

*then for all  $z \notin F$ ,*

$$M, z \models^\tau \Phi \text{ iff } M^p, z \models^\tau \Phi.$$

PROOF. By induction on the complexity of formulas.

Assume the hypothesis. The cases for the propositional variables,  $\perp$ , and the conditional are straightforward.

If  $x \notin F$ , assume  $M, x \models^\tau \Box\Psi$ . Thus, for some  $O$ ,  $x \in O$  and for all  $y \in O \setminus \{x\}$ ,  $M, y \models^\tau \Psi$ . Since  $x \notin F$ ,  $x \in O \setminus F$  and for all  $y \in (O \setminus F) \setminus \{x\}$ ,  $M, y \models^\tau \Psi$ . Since none of the points in  $F$  are in  $(O \setminus F) \setminus \{x\}$ , we may apply the induction hypothesis and infer: for all  $y \in (O \setminus F) \setminus \{x\}$ ,  $M^p, y \models^\tau \Psi$ . Since  $O \setminus F$  is co-finite, it is open, thus  $M^p, x \models^\tau \Box\Psi$ . The converse is analogous.

If  $x \notin F$ , assume  $M, x \models^\tau (\exists q)\Psi$ . Thus for some  $q$ -variant of  $M$ ,  $M^q, x \models^\tau \Psi$ . Let  $V^q$  be the valuation of  $M^q$ .

Let  $V^{qp}$  be the valuation for a  $p$ -variant of  $M^q$ , such that for all points  $z$ :  $z \in V^{qp}(p)$  if  $z \in V^p(p)$ . Significantly,  $M^{qp}$  is a  $q$ -variant of  $M^p$ . Furthermore, Since  $\Psi$  is true at  $x$  in  $M^q$  and the only possible difference between  $V^q$  and  $V^{qp}$  is the valuation of  $p$  in the finite set  $F$ , by the induction hypothesis,  $M^{qp}, x \models^\tau \Psi$ . By the definition of truth,  $M^p, x \models^\tau (\exists q)\Psi$ . The converse is analogous.  $\dashv$

We will use the following notation for the following lemmas. For any  $\Phi$ , let  $f(\Phi) := \{p \mid p \text{ is free in } \Phi\}$ .

In the next lemma we show that if two points agree on all the free variables of  $\Phi$ , then they agree on  $\Phi$ . We will use the previous lemma to show this.

LEMMA 4.10. *Let  $M \in \mathcal{C}^{\text{co}}$  and let  $x, y \in M$ . Assume for all  $q \in f(\Phi)$ ,  $x \in V(q)$  iff  $y \in V(q)$ . Then  $M, x \models^\tau \Phi$  iff  $M, y \models^\tau \Phi$ .*

PROOF. By induction on the complexity of formulas. Assume the hypothesis. The cases for the propositional variables,  $\perp$ , and the conditional are straightforward.

For the case where  $\Phi = \Box\Psi$ , assume  $M, x \models^\tau \Box\Psi$ . Thus,  $\Psi$  is true at co-finitely many points. Let  $U$  be the set of points where  $\Psi$  is true.  $U$  is open, and so is  $U \cup \{y\}$ . So by definition of truth in a model,  $M, y \models^\tau \Box\Psi$ .

For the case where  $\Phi = (\exists q)\Psi$ , assume  $M, x \models^\tau (\exists q)\Psi$ . Thus there is some  $q$ -variant of  $M$ ,  $M^q$ , and  $\Psi$  is true at  $x$  in  $M^q$ . If  $x$  and  $y$  agree on  $q$  in  $M^q$ , then they agree on all free variables of  $\Psi$ , and so we may use the induction hypothesis to conclude  $M^q, y \models^\tau \Psi$ . And since  $M^q$  is a  $q$ -variant of  $M$ ,  $M, y \models^\tau (\exists q)\Psi$ .

On the other hand, if  $x$  and  $y$  do not agree on  $q$  in  $M^q$ , then consider the  $q$ -variant of  $M^q$ ,  $M^{qq}$ , where  $M^{qq}$  is exactly the same as  $M^q$ , except that we change the valuation of  $q$  at  $y$  to agree with the valuation of  $q$  at  $x$  in  $M^q$ . Here we let  $F = \{y\}$  and apply Lemma 4.9 to conclude that  $M^{qq}, x \models^\tau \Psi$ . And since  $x$  and  $y$ , in  $M^{qq}$ , completely agree on all free variables in  $\Psi$ , we use the induction hypothesis to conclude  $M^{qq}, y \models^\tau \Psi$ . And since  $M^{qq}$  is a  $q$ -variant of  $M$ ,  $M, y \models^\tau (\exists q)\Psi$ .  $\dashv$

LEMMA 4.11. *Let  $M, M' \in \mathcal{C}^{\text{co}}$  and let  $g$  be a 1-1, onto function from  $M$  to  $M'$ . Assume that for all  $p \in f(\Phi)$  and for all points  $z$ ,*

$$z \in V(p) \text{ iff } g(z) \in V'(p),$$

*then, for all  $x$ ,*

$$M, x \models^\tau \Phi \text{ iff } M', g(x) \models^\tau \Phi.$$

PROOF. We assume the hypothesis and proceed by induction on the complexity of formulas, going ahead to the cases for the box and the existential quantifier.

Assume  $M, x \models^\tau \Box\Psi$ . Thus all but finitely many points in  $M$  force  $\Psi$ . By induction hypothesis, all but finitely many points force  $\Psi$  in  $M'$ , thus  $M', g(x) \models^\tau \Box\Psi$ . The converse is similar.

Assume  $M, x \models^\tau (\exists q)\Psi$ . Thus for some  $q$ -variant of  $M$ ,  $M^q$ ,  $x$  forces  $\Psi$  in  $M^q$ . It is possible that  $q$  is free in  $\Psi$ , so we create a  $q$ -variant of  $M'$ ,  $M'^q$ , where for all  $z$ ,  $g(z) \in V'^q(q)$  if  $z \in V^q(q)$ .

Now, each  $x$  in  $M'^q$  agrees with  $g(x)$  in  $M^q$  on all free variables in  $\Psi$ , and so the induction hypothesis applies, and thus  $M'^q, g(x) \models^\tau \Psi$ . And since  $M'^q$  is a  $q$ -variant of  $M'$ ,  $M', g(x) \models^\tau (\exists q)\Psi$ . The converse is similar.  $\dashv$

We can now show a topological analogue to Theorem 2.9:

THEOREM 4.12. *For all closed  $\Phi$ ,*

$$\text{If } \mathcal{C}^{\text{co}} \not\models^\tau \neg\Phi, \text{ then } \mathcal{C}^{\text{co}} \models^\tau \Phi.$$

PROOF. Assume  $\mathcal{C}^{\text{co}} \not\models^\tau \neg\Phi$ . Thus  $\Phi$  is true at a point  $w$  in some  $M \in \mathcal{C}^{\text{co}}$ . Let  $z$  be any point in  $M$ . Since  $\Phi$  is closed,  $f(\Phi)$  is empty,

thus by Lemma 4.10,  $\Phi$  is true at  $z$  in  $M$ , and since  $z$  was arbitrary,  $\Phi$  is true at every point in  $M$ .

Now let  $M^*$  be any model in  $\mathcal{C}^{\text{co}}$ . Since all the topologies in  $\mathcal{C}^{\text{co}}$  are based on a countably infinite set, they all have the same cardinality, and so there is a 1–1, onto function between  $M$  and  $M^*$ . Thus by Lemma 4.11,  $\Phi$  is true at every point in  $M^*$ . And since  $M^*$  is arbitrary,  $\mathcal{C}^{\text{co}} \models^{\tau} \Phi$  ⊣

Thus if a closed formula is true at any point in any model, it is true at every point in every model. Using the theorem above we have the following analogue to Corollary 2.10. The proof is similar to Corollary 2.10, but instead of using Theorem 2.9 we use Theorem 4.12.

**COROLLARY 4.13.** *For all closed  $\Phi$ ,  $\mathcal{C}^{\text{co}} \models^{\tau} \Box\Phi$  iff  $\mathcal{C}^{\text{co}} \models^{\tau} \Phi$ .*

We also have the following analogue to Theorem 2.11.

**THEOREM 4.14.** *If  $\Phi$  is a formula with no quantifiers and  $p_i, \dots, p_j$  is a list of the propositional variables in  $\Phi$ ,*

$$\mathbf{KD45} \not\models \Phi \text{ iff } \mathcal{C}^{\text{co}} \models^{\tau} (\exists p_i) \dots (\exists p_j) \neg \Phi.$$

**PROOF.** Assume  $\Phi$  is a formula with no quantifiers and  $p_i, \dots, p_j$  are all the propositional variables in  $\Phi$ . By Theorem 4.8, **KD45** is complete with respect to  $\mathcal{C}^{\text{co}}$ . Thus if **KD45**  $\not\models \Phi$ , then for some  $M \in \mathcal{C}^{\text{co}}$  and some  $w \in M$ ,  $M, w \models \neg\Phi$ . Trivially,  $M$  is a  $p$ -variant of itself, for any  $p$ , but in particular for  $p_i, \dots, p_j$ . Thus, by the definition of truth in a model,  $M, w \models (\exists p_i) \dots (\exists p_j) \neg\Phi$ . By Lemma 4.12,  $\mathcal{C}^{\text{stE}} \models (\exists p_i) \dots (\exists p_j) \neg\Phi$ .

Conversely, if  $\mathcal{C}^{\text{co}} \models (\exists p_i) \dots (\exists p_j) \neg\Phi$ , then  $\Phi$  fails in some model in  $\mathcal{C}^{\text{co}}$ , and since **KD45** is sound for  $\mathcal{C}^{\text{co}}$ , **KD45**  $\not\models \Phi$ . ⊣

From Theorem 4.14, we have all the corollaries from Corollary 3.1 to Corollary 3.7. And the philosophical interpretation of these theorems, from Section 3 holds over to the topological case as well. All of this leads us to conject the following conjecture:

$$\mathcal{C}^{\text{stE}} \models \Phi \text{ iff } \mathcal{C}^{\text{co}} \models^{\tau} \Phi.$$

#### 4.4. Philosophical interpretation of the topological semantics

In a recent paper by Baltag et al. (2018), where *another* topological approach to belief is presented, some interesting philosophical criticisms

were put forth regarding the topological approach to belief considered here.

[...] *in any topo-model and any state in this model, there is at least one false belief*, that is, the agent always believes the false proposition  $X \setminus \{x\}$  at the actual state  $x$ .

This corresponds to the validity of  $(\exists p)(\Box p \wedge \neg p)$ . The agent always has at least one false, rational belief. Recall that we addressed this curious fact at the start of Section 3, by presenting a syntactic proof of:

$$\Box(\exists p)(\Box p \wedge \neg p) \rightarrow (\exists p)(\Box p \wedge \neg p)$$

Thus  $\mathbf{A}^{\text{EM}}$ , i.e.  $\Box(\exists p)(\Box p \wedge \neg p)$ , implies that the agent has at least one false belief, and so if there is a philosophical problem here, it is with the axiom of epistemic modesty, which does seem plausible.

We now present some basic concepts in topology in order to present this alternative approach from (Baltag et al., 2018). Far from arguing against this alternative approach, our aim is to show it is, in an important case, the same as the approach here.

DEFINITION 4.15. Given  $\langle X, \tau \rangle$  and  $A \subseteq X$ , the *interior* of  $A$  is the set  $\text{Int}(A) = \bigcup \{O \subseteq A \mid O \in \tau\}$ .

Thus  $\text{Int}(A)$  is the union of all open subsets of  $A$ . In 1938, Tsao-Chen (1938) noticed a connection between the closure operator and the diamond of one of Lewis' systems of strict implication, and this precipitated a series of papers (McKinsey, 1941; McKinsey and Tarski, 1944, 1948) exploring the connection more deeply. Because of their work,  $\mathbf{S4}$  is well-known as the logic of the interior operator (see Aiello et al., 2007, for a straightforward completeness proof).

Recall that a set is *closed* when the complement is open. The *closure* of a set  $A$ ,  $\text{Cl}(A)$ , is the intersection of all the closed supersets of  $A$ . And,  $\text{Cl}$  is the dual of  $\text{Int}$ . And so just as the interior operator is the box of  $\mathbf{S4}$ , the closure operator,  $\text{Cl}$ , is the diamond of  $\mathbf{S4}$ . That is,  $\text{Cl}(A) = \neg \text{Int}(\neg A)$ , where for any  $Y \in 2^X$  we put  $\neg Y := X \setminus Y$ . The following is well-known and straightforward to show:  $\text{Cl}(A) = d(A) \cup A$ . And thus, dually, we have:  $\text{Int}(A) = \neg d(\neg A) \cap A$ .

Stalnaker (2006) presents an interesting way to define belief in terms of knowledge.

- $\text{B}\Phi$  is  $\neg \text{K} \neg \text{K}\Phi$

Stalnaker



Keeping in mind that the interior operator acts like a knowledge operator (by obeying **S4**), [Baltag et al. \(2018\)](#) explore Stalnaker's definition in a topological setting. And so their topological version of belief is the closure of the interior. Thus, given a set of possible worlds  $A$ , we have:

- $B(A)$  is  $\text{Cl}(\text{Int}(A))$  Topo-Stalnaker

To be sure, the Topo-Stalnaker approach is presented as a rival approach to the approach discussed here. And here we're interpreting rational belief in the proposition  $A$  as  $-d(-A)$ , where  $d$  is our diamond. However, despite definite differences, we now show an important case where the two approaches coincide.

We show below that in the case where  $d$  obeys the diamond of **KD45**, we have:  $-d(-A) = \text{Cl}(\text{Int}(A))$ . A helpful piece of notation is the boxdot notation,  $\square$  from ([Boolos, 1993](#)):

$$\square\Phi := (\square\Phi \wedge \Phi)$$

which gives us the dual:

$$\diamond\Phi := (\diamond\Phi \vee \Phi)$$

Thus, since  $\text{Cl}(A) = d(A) \cup A$ , and ' $\diamond$ ' represents  $d$ , ' $\square$ ' represents the closure operator. And so ' $\square$ ' represents the interior operator. Putting all this together, *belief that  $p$*  is represented in the Topo-Stalnaker interpretation with:  $\diamond\square p$ . Thus, we want to show:

**THEOREM 4.16.** **KD45**  $\vdash \square p \leftrightarrow \diamond\square p$ .

**PROOF.** 1. **K4**  $\vdash \square p \rightarrow \square\square p$

2. **K4**  $\vdash \square p \rightarrow (\square\square p \wedge \square p)$  from 1

3. **K4**  $\vdash \square p \rightarrow \square(\square p \wedge p)$  from 2, the box distributes over ' $\wedge$ '

4. **K4**  $\vdash \square p \rightarrow \square\square p$  from 3 and the definition of ' $\square$ '

5. **KD4**  $\vdash \square p \rightarrow \diamond\square p$  from 4 and **D**

6. **KD4**  $\vdash \square p \rightarrow (\diamond\square p \vee \square p)$  from 5, weakening the consequent

7. **KD4**  $\vdash \square p \rightarrow \diamond\square p$  from 6 and the definition of ' $\diamond$ '.

Conversely,

1. **K**  $\vdash \square p \rightarrow \square p$

2. **K**  $\vdash \diamond\square p \rightarrow \diamond\square p$  from 1, the diamond is monotonic

3. **K5**  $\vdash \diamond p \rightarrow \square\diamond p$

4. **K5**  $\vdash \diamond\square p \rightarrow \square p$  from 3, contraposition and let  $p$  be  $\neg p$

5. **K5**  $\vdash \diamond\square p \rightarrow \square p$  from 2 and 4

6. **K5**  $\vdash (\diamond \Box p \vee \Box p) \rightarrow \Box p$  from 1 and 5  
 7. **K5**  $\vdash \diamond \Box p \rightarrow \Box p$  from 6 and the definition of ‘ $\diamond$ ’

Thus, we obtain **KD45**  $\vdash \Box p \leftrightarrow \diamond \Box p$ . ⊢

Thus, considering Theorem 4.16, if one is content to use  $d$  and **KD45**, then one should have no objection to the Topo-Stalnaker approach. As Stalnaker’s approach implies **KD45** for belief, this point feels salient.

For the most part, our interpretation of the topological semantics is the same as the relational semantics. We are interpreting points as possible worlds, and sets of worlds as propositions. Thus a proposition  $P$  is true at a world  $w$  if and only if  $w \in P$ . Note that, if we take the infinite intersection of all propositions true at world  $w$ , we get:  $\{w\}$ . Singleton propositions are maximal propositions. A singleton proposition represents a complete picture of a world.

Now, there is one aspect of the topological semantics, related to singleton propositions, which is difficult to interpret, and there is no analogue for it in the relational case. Consider the following topological condition:

$T_1$ : for all points  $x, y \in X$ , we have  $x \notin d(\{y\})$ .

This condition holds of all the topological spaces in  $\mathcal{C}^{\text{co}}$ . In fact, it is equivalent to the condition of every co-finite set being open. The validity of axiom 5 implies this condition, as does the validity of  $\Box(\Box p \rightarrow p)$ .

How should the  $T_1$  condition be interpreted? Again, any singleton proposition  $\{w\}$  represents a *maximal* proposition. In *some* sense, the  $T_1$  condition is saying:  $\{x\}$  is *never* possible, at any world.

In (Steinsvold, 2008) an interpretation inspired by the work of Patrick Grim (1991) is presented. Grim argues, in a variety of ways, that a totality of truths is incoherent. Applying the conclusion here, maximal propositions, i.e.  $\{x\}$ , are incoherent. Thus, we can make sense out of the topological semantics as long as we accept that  $\{x\}$  does not make sense.

Alternatively, perhaps one could argue that the probability of a maximal proposition being true is so infinitesimally small, if not zero, that one is justified in believing it will never be the case. However, such an argument would seem to raise philosophical questions which are not easy to answer.

We conclude this section with various comments.

1. As mentioned, there are other topologies, where the axioms of **KD45** are valid. Let  $X$  be an infinite set, and let  $A$  be an infinite

subset of  $X$ . Let  $\tau^A = \{ U \subseteq X \mid (\exists F)(F \subseteq A \text{ and } F \text{ is finite, and } A \setminus F \subseteq U) \} \cup \{\emptyset\}$ .

Given  $\langle X, \tau^A \rangle$ , any member of  $\tau^A$  (besides  $\emptyset$ ) can be formed by taking  $A$ , removing finitely many points, and then adding any amount of points from  $X$ . Every non-empty member of  $\tau^A$  will contain all but finitely many points of  $A$ . Our completeness proof for single-agent **KD45** is a simplified version of the topological completeness proof for multi-agent **KD45** with common belief, to be found in (Steinsvold, 2007), and it utilizes these types of topologies.

2. In the case axiom 5 is seen as too strong, one might consider **KD4Q**, that is **KD4**+ $\Box(\Box p \rightarrow p)$ . Let  $A$  and  $B$  be two infinite, disjoint sets. Let  $\tau$  be the co-finite topology on  $A$ , let  $\tau'$  be the cofinite topology on  $A \cup B$ , and let  $\tau'' = \tau' \cup \tau$ . The axioms of **KD4Q**, are valid in  $\langle A \cup B, \tau'' \rangle$ , but axiom 5 can fail at any point in  $B$  by letting  $V(p) = B$ .

### 4.5. The Gettier problem

As noted above, we have:

$$\text{Int}(A) = -d(-A) \cap A.$$

Granting that the interior operator represents knowledge, and the dual of  $d$  represents rational belief, we have:

- $K\Phi$  is  $\Box\Phi \wedge \Phi$  the traditional analysis of knowledge

Thus, topologically, knowledge is true, rational belief. Famously, in 1963, Edmund Gettier published a short article (1963) challenging this simple thesis with some interesting counterexamples, and epistemology has never been the same. To be sure, this thesis is typically put in terms of ‘justified belief’ instead of ‘rational belief,’ but for our purposes we can safely ignore this distinction.

To give a quick, oft-used, example. Suppose you are looking out at a field, and there is a tree in the field. Behind the tree, unbeknownst to you, is a sheep. However, in front of you is a fake sheep—something that looks exactly like a normal sheep, but is not a sheep. It does not exactly matter what the fake sheep is, it could be a robot in the form of a sheep, or a dog disguised to look like a sheep, all that matters is that it looks just like a normal sheep, so that you believe it is a sheep, and your belief is justified. Now, consider the following sentence,

S: There is a sheep in the field.

In this example,  $S$  is true, you believe  $S$ , and you are justified in believing  $S$ . Yet, you do not *know* that  $S$ .

Epistemologists agree with Gettier: the traditional analysis of knowledge is incomplete. Trying to solve the Gettier problem is typically thought of in terms of adding an extra-condition to the traditional analysis of knowledge in order to make it complete. However, there has been an impressive and wild diversity of responses to the problem, and epistemologists are far from settled on this question. Some have given up altogether.

The interior operator acts like knowledge, and the derived set, given certain restrictions, acts like rational belief. But if we take both of them as is, we get our own topological version of the Gettier problem. This problem was noted in (Steinsvold, 2007), and the proposed solution was to use clopen sets for knowledge. A set is *clopen* if it is closed and open, i.e., it is open and its complement is open (let  $\text{clop}$  be the set of all clopen sets). Using ‘ $K^c$ ’ to represent knowledge, we have,

- $M, w \models^\tau K^c\Phi$  iff  $(\exists O \in \text{clop})(w \in O \text{ and } (\forall x \in O) M, x \models^\tau \Phi)$ .

Using this, it is straightforward to check that  $K^c\Phi \rightarrow (\Box\Phi \wedge \Phi)$  is valid in all topological spaces, but the converse fails, and thus one avoids the topological Gettier problem. Furthermore, **S4** is sound for  $K^c$ , is it complete? A topological space  $\langle X, \tau \rangle$  is *connected* if  $\text{clop} = \{\emptyset, X\}$ , i.e., the only clopen sets are  $X$  and  $\emptyset$ . It is straightforward to see the **5** axiom for  $K^c$  holds in any connected topological space. Are there any topological spaces where  $d$  obeys **KD45** and **5** fails for  $K^c$ ?

Though using clopen sets may superficially solve this topological Gettier problem, it really is not clear what a *clopen* set represents, epistemically. That is, what does it really have to do with knowledge or the actual Gettier problem? This is not clear.

There is another point worth mentioning in relation to this topological Gettier problem, and it is directly related to a well-known attempt to solve the problem. As noted in (Steinsvold, 2007) and also (Baltag et al., 2018), open sets do not represent just any type of justification, they represent *true* justifications. That is, suppose I believe  $P$ , and I believe  $P$  on the basis of  $Q$ ,  $Q$  being my justification. In everyday life, it is possible that  $Q$  is false. Whereas, considering the definition of truth in a model, if  $\Box p$  is true at  $w$ , then there is an open set  $O$  and  $w \in O$  (etc.). Point being: a justification, i.e. an open set, has to be *true* at a world for an agent to use it.

Call a belief *well-founded* if it is justified by a truth. Here,  $\Box p$  represents a well-founded belief. Thus we can view the topological semantics not just as a semantics for rational belief, but for well-founded, rational beliefs.

Going back to our Gettier example mentioned above: you believe that S, that is, you believe there is a sheep in the field. But on what basis is S believed? Is it well-founded? It is not. You believe S because you believe: ‘*that* thing is a sheep.’ Thus, you believe there is *some* sheep in the field because you believe that particular thing you are looking at is a sheep. But *that* is false, the thing you are looking at is not a sheep. It is a fake sheep.

The diagnosis just given to our Gettier example is representative of a major, early response to the Gettier problem, typically labeled the ‘No False Lemmas’ condition, introduced by Armstrong in (Armstrong, 1974). Since knowledge implies justified belief, various philosophers had argued that if you really do know something, then the relevant belief must be well-founded to block Gettier-type examples. While this certainly does block many Gettier examples, the No False Lemmas condition was argued to be, in some cases, too weak, and in other cases, too strong.

Thus, while we certainly have not solved this problem, it is interesting that the topological semantics has at least one major response to the Gettier problem *built-in*. And, to be thorough, the epistemologist Michael Levin in (Levin, 2006) has argued that many of the supposed arguments against the No False Lemma condition do not work. While, as Levin says, it is ‘textbook orthodoxy’ that the No False Lemmas condition is unsatisfactory, Levin’s work shows that this is a premature judgement.<sup>4</sup> Perhaps the No False Lemmas condition simply fell out of philosophical fashion before it could be given its best defense.

Typically, when one interprets a topology epistemically, the open sets are treated as evidence. And the fact that topologies are closed under *finite* intersection represents a familiar form of reasoning. To quote (Baltag et al., 2018),

[...] closure under finite intersection capture’s an agent’s ability to put finitely many pieces into a single piece.

But what about arbitrary intersections? Topologies closed under arbitrary intersection are *Alexandroff* topologies. And there is a one-to-one

---

<sup>4</sup> To be thorough, Levin believes the well-known fake-barn case is still a problem for the No False Lemmas condition.

correspondence between Alexandroff topological models and transitive and reflexive relational models (for a proof see [Aiello et al., 2007](#)). Thus, any Alexandroff topology is essentially a frame for **S4** and any frame for **S4** is essentially an Alexandroff topology.

Thus, it seems safe to say, if we want to represent an agent in a way that is different than the usual relational semantics, we should not be using Alexandroff topologies for that representation. With that in mind, there are no Alexandroff topologies for **KD45** when using  $d$  as the diamond (and we leave the proof for the reader).

And to make a point which reflects the general theme of this paper: since a topology can represent an agent which has finite intersection, but also *lacks* infinite intersections, this represents a significant limitation on our agent, and thus reflects a modest<sub>1</sub> quality of the agent.

## 5. Conclusion

As far as we can tell, both semantics explored in this paper validate the same formulas, so in that regard they are equal. And of course we'd be grateful to anyone who could prove, or disprove, the conjecture that both semantics validate the exact same formulas. However, the topological semantics, while tantalizingly epistemic, raises difficult interpretive questions. Thus, if one wants to represent a modest agent, the relational semantics with bisimulation quantifiers seems to be the more straightforward approach.

**Acknowledgements.** Thanks to Tim French for reading an earlier version of this work, and to Kit Fine and Rohit Parikh for discussion early on which lead to this paper. And thanks to an anonymous referee for a careful review and many helpful comments.

## References

- Aiello, M., I. Pratt-Hartmann, and J. van Benthem (eds.), 2007, *Handbook of Spatial Logics*, Springer, Dordrecht. DOI: [10.1007/978-1-4020-5587-4](https://doi.org/10.1007/978-1-4020-5587-4)
- Antonelli, G. A., and R. H. Thomason, 2002, "Representability in second-order propositional poly-modal logic", *The Journal of Symbolic Logic* 67: 1039–1054. DOI: [10.2178/jsl/1190150147](https://doi.org/10.2178/jsl/1190150147)

- Armstrong, D., 1974, *Belief, Truth, and Knowledge*, Cambridge University Press, Cambridge, Mass. DOI: [10.1017/CB09780511570827](https://doi.org/10.1017/CB09780511570827)
- Baltag, A., N. Bezhanishvili, A. Özgün, and S. Smets, 2019, “A topological approach to full belief”, *Journal of Philosophical Logic* 48 (2): 205–244. DOI: [10.1007/s10992-018-9463-4](https://doi.org/10.1007/s10992-018-9463-4)
- Boolos, G., 1993, *The Logic Of Provability*, Cambridge University Press, Cambridge, Mass. DOI: [10.1017/CB09780511625183](https://doi.org/10.1017/CB09780511625183)
- Bull, R. A., 1969, “On modal logics with propositional quantifiers”, *The Journal of Symbolic Logic* 34: 257–263. DOI: [10.2307/2271102](https://doi.org/10.2307/2271102)
- D’agostino, G., and G. Lenzi, 2005, “An axiomatization of bisimulation quantifiers via the  $\mu$ -calculus”, *Theoretical Computer Science* 338: 64–95. DOI: [10.1016/j.tcs.2004.10.040](https://doi.org/10.1016/j.tcs.2004.10.040)
- Driver, J., 1989, “The virtues of ignorance”, *The Journal of Philosophy* 86: 373–384. DOI: [10.2307/2027146](https://doi.org/10.2307/2027146)
- Evnine, S. J., 2001, Learning from one’s mistakes: epistemic modesty and the nature of belief”, *Pacific Philosophical Quarterly* 82: 157–177. DOI: [10.1111/1468-0114.00123](https://doi.org/10.1111/1468-0114.00123)
- Fine, K., 1970, “Propositional quantifiers in modal logic”, *Theoria*, 36: 336–346. DOI: [10.1111/j.1755-2567.1970.tb00432.x](https://doi.org/10.1111/j.1755-2567.1970.tb00432.x)
- Flanagan, O., 1990, “Virtue and ignorance”, *Journal of Philosophy*, 87 (8): 420–428. DOI: [10.2307/2026736](https://doi.org/10.2307/2026736)
- French, T., 2005, “Bisimulation quantified logics: undecidability”, pages 396–407 in S. Sarukkai and S. Sen (eds.), *FSTTCS 2005: Foundations of Software Technology and Theoretical Computer Science*, Lecture Notes in Computer Science, vol. 3821, Springer, Berlin, Heidelberg. DOI: [10.1007/11590156\\_32](https://doi.org/10.1007/11590156_32)
- French, T., 2006a, “Bisimulation quantifiers for modal logic”, PhD thesis, University of Western Australia.
- French, T. 2006b, “Bisimulation quantified modal logics: decidability”, *Advances in Modal Logic* 6: 147–166.
- Gettier, E., 1963, “Is justified true belief knowledge?”, *Analysis* 23 (6): 121–123. DOI: [10.1093/analys/23.6.121](https://doi.org/10.1093/analys/23.6.121)
- Ghilardi, S., and M. Zawadowski, 1995, “Undefinability of propositional quantifiers in the modal system S4”, *Studia Logica* 55: 259–271. DOI: [10.1007/BF01061237](https://doi.org/10.1007/BF01061237)

- Goldblatt, R., 1992, *Logics of Time and Computation*, 2nd Edition, CSLI Lecture Notes no. 7.
- Grim, P., 1991, *The Incomplete Universe*, MIT Press, Cambridge, Mass.
- Kaplan, D., 1970, "S5 with quantifiable propositional variables", *The Journal of Symbolic Logic* 35 (2): 355.
- Kuhn, S., 2004, "A simple embedding of T into double S5", *Notre Dame Journal of Formal Logic* 45 (1): 13–18. DOI: [10.1305/ndjfl/1094155276](https://doi.org/10.1305/ndjfl/1094155276)
- Levin, M., 2006, "Gettier cases without false lemmas?", *Erkenntnis* 65 (3): 381–392. DOI: [10.1007/s10670-005-5470-2](https://doi.org/10.1007/s10670-005-5470-2)
- MacIntosh, J. J., 1980, "An extension of a proof of Prior's or when thinking makes it so", *Analysis* 40 (2): 86–89. DOI: [10.2307/3327417](https://doi.org/10.2307/3327417)
- McKinsey, J. C. C., 1941, "A solution of the decision problem for the Lewis systems S2 and S4, with an application to topology", *The Journal of Symbolic Logic* 6 (4): 117–134. DOI: [10.2307/2267105](https://doi.org/10.2307/2267105)
- McKinsey, J. C. C., and A. Tarski, 1944, "The algebra of topology", *Annals of Mathematics* 45: 141–191. DOI: [10.2307/1969080](https://doi.org/10.2307/1969080)
- McKinsey, J. C. C., and A. Tarski, 1948, "Some theorems about the sentential calculi of Lewis and Heyting", *Journal Of Symbolic Logic* 13: 1–15. DOI: [10.2307/2268135](https://doi.org/10.2307/2268135)
- Pitts, A. M., 1992, "On an interpretation of second order quantification in first order intuitionistic propositional logic", *The Journal of Symbolic Logic* 57: 33–52. DOI: [10.2307/2275175](https://doi.org/10.2307/2275175)
- Prior, A. N., 1971, *Objects of Thought*, Clarendon, Oxford, DOI: [10.1093/acprof:oso/9780198243540.001.0001](https://doi.org/10.1093/acprof:oso/9780198243540.001.0001)
- Richards, N., 1992, *Humility*, Temple University Press, Philadelphia.
- Schulz, K., 2010, *Being Wrong: Adventures in the Margin of Error*, Ecco, New York.
- Stalnaker, R., 2006, "On logics of knowledge and belief", *Philosophical Studies* 128 (1): 169–199. DOI: [10.1007/s11098-005-4062-y](https://doi.org/10.1007/s11098-005-4062-y)
- Steinsvold, C., 2003, "Towards a topology of knowledge and belief", Talk given at the Workshop on Reasoning about Space, during the NASSLI conference, Bloomington, Indiana.
- Steinsvold, C., 2007, "Topological models of belief logics", PhD thesis, CUNY GSUC, New York, NY.



- Steinsvold, C., 2008, “A grim semantics for logics of belief”, *The Journal of Philosophical Logic* 37 (1): 45–56. DOI: [10.1007/s10992-007-9055-1](https://doi.org/10.1007/s10992-007-9055-1)
- Tsao-Chen, T., 1938, “Algebraic postulates and a geometric interpretation for the Lewis calculus of strict implication”, *Bulletin of the American Mathematical Society* 44: 737–744. DOI: [10.1090/S0002-9904-1938-06860-7](https://doi.org/10.1090/S0002-9904-1938-06860-7)
- Visser, A., 1996, “Uniform interpolation and layered bisimulation”, pages 139–164 in P. Hájek (ed.), *Gödel’96, Logical Foundations of Mathematics, Computer Science and Physics – Kurt Gödel’s Legacy*, Springer, Berlin. DOI: [10.1017/9781316716939.010](https://doi.org/10.1017/9781316716939.010)

CHRISTOPHER STEINSVOLD  
Department of Philosophy  
Brooklyn College, City University of New York  
New York, USA  
[steinsvold1@verizon.net](mailto:steinsvold1@verizon.net)