

# Evidentialist Logic

Matthew P. Wampler-Doty

## Abstract

In this paper we present a novel epistemic logic, which we shall call *Evidentialist Logic*. The logic is found to be non-compact, and sound and complete for a Hilbert style proof system. The logic is further shown to be a conservative extension of basic modal logic; subsequently the decision problem is observed to be PSPACE hard.

## 1 Introduction

In Hintikka (1962), Hintikka originally imagined Epistemic logic as system for reasoning about the knowledge of a single agent in a static environment. His original analysis was to consider what an agent would *eventually* conclude, given enough time to reach every derivable logical conclusion. While his original work has been widely celebrated, it lacked an element fundamental to 20th century epistemology - the notion of *justification*.

In Artemov and Nogina (2005), an answer was given to this problem with the development of *justification logic*, which took inspiration from the *logic of proofs* (Artemov, 1994). The fundamental innovation behind this approach is to introduce new *syntax*, namely formulae of the form  $t : \phi$ , which are intended to mean “ $t$  is evidence for  $\phi$ .”

In this paper we offer an approach in the same spirit, however the developments we make are essentially *semantic* rather than syntactic. We consider beliefs to be formed of a deductive process, but consider them to be conclusions drawn from a body of *premises*, which the agent has acquired through their senses and education forming a web of belief. Throughout this essay, we shall consider these premises as evidence, and use these terms interchangeably. Our chosen reading of  $\Box_X \phi$  is now “The agent named  $X$  has an argument for  $\phi$ , based on their current premises and background information.” We also offer 2 novel modalities:

- $\Diamond_X \phi$  – which we read “for some accessible restriction of the agent  $X$ ’s current premises,  $\phi$  holds” where  $\phi$  is some formula representing an epistemic statement
- $\Diamond_X \phi$  – which we read “for some accessible extension of the agent  $X$ ’s premises,  $\phi$  holds”

In addition to this two modalities, we shall also employ the following new syntax:

- $\bigcirc$  – which we read “agent  $X$ ’s current premises are sound”

To further motivate the intuition behind these semantics, we imagine that agents can choose to ignore premises in order to compose stronger deductions, in the spirit of René Descartes’s Meditation’s II (Descartes and Tweyman, 1993):

I will, nevertheless, make an effort, and try anew the same path on which I had entered yesterday, that is, proceed by casting aside all that admits of the slightest doubt, not less than if I had discovered it to be absolutely false; and I will continue always in this track until I shall find something that is certain, or at least, if I can do nothing more, until I shall know with certainty that there is nothing certain.

We shall also consider the *dual* of this behavior, where the agent entertains premises beyond what they are currently considering. In both of these operations, we will assume that the state of affairs outside of the agent’s psychology remains fixed, and only the premises undergo reconsideration.

We have named the logic behind these concerns *Evidentialist Logic*, after the philosophical view that one has knowledge of a proposition  $\phi$  if it follows from appropriate premises. Furthermore, we have chosen to consciously lampoon ourselves by abbreviating *Evidentialist Logic* as EVIL throughout this essay.

## 2 Grammar & Semantics

Let  $\Phi$  be an infinite class of sentence letters, and let  $\mathcal{A}$  be a class of agents. The grammar of EVIL is given by the following Backus-Naur form context free grammar:

$$\phi ::= p \mid \phi \rightarrow \psi \mid \perp \mid \Box_X \phi \mid \Box_X \phi \mid \Box_X \phi \mid \bigcirc_X$$

... where  $X \in \mathcal{A}$  and  $p \in \Phi$ . We shall refer to this language as alternately  $\mathcal{L}(\Phi, \mathcal{A})$  or  $\mathcal{L}(\Phi)$ . We follow the usual abbreviations for other connectives, diamonds, and logical signs.

Next, we shall define recursively the Tarski truth predicate ( $\models$ ). For an arbitrary  $\mathfrak{M} \subseteq (\wp\Phi) \times (\wp\mathcal{L}(\Phi))^{\mathcal{A}}$  and  $(a, A) \in \mathfrak{M}$  we write:<sup>1</sup>

- 1  $\mathfrak{M}, (a, A) \models p \iff p \in a$
- 2  $\mathfrak{M}, (a, A) \models \phi \vee \psi \iff$  either  $\mathfrak{M}, (a, A) \models \phi$  or  $\mathfrak{M}, (a, A) \models \psi$
- 3  $\mathfrak{M}, (a, A) \models \neg\phi \iff \mathfrak{M}, (a, A) \not\models \phi$
- 4  $\mathfrak{M}, (a, A) \models \Box_X \phi \iff$  for all  $(b, B) \in \mathfrak{M}$  where  $\mathfrak{M}, (b, B) \models A_X$ , we have  $\mathfrak{M}, (b, B) \models \phi$
- 5  $\mathfrak{M}, (a, A) \models \Box_X \phi \iff$  for all  $(b, B) \in \mathfrak{M}$  where  $B_X \subseteq A_X$  and  $a = b$ , then  $\mathfrak{M}, (b, B) \models \phi$
- 6  $\mathfrak{M}, (a, A) \models \Box_X \phi \iff$  for all  $(b, B) \in \mathfrak{M}$  where  $A_X \subseteq B_X$  and  $a = b$ , then  $\mathfrak{M}, (b, B) \models \phi$
- 7  $\mathfrak{M}, (a, A) \models \bigcirc_X \iff \mathfrak{M}, (a, A) \models A_X$

---

<sup>1</sup>In the following, for all  $A \in (\wp\mathcal{L}(\Phi))^{\mathcal{A}}$  we shall write  $A(X)$  as  $A_X$

Considered as a recursive function,  $(\models)$  is not *total*; there are inputs for which it does not terminate. For instance, if one lets  $\mathfrak{M} := \{(\emptyset, S)\}$ , where  $S := \{\Box \perp\}$  and  $\perp := p \wedge \neg p$  for some  $p \in \Phi$ , then  $\mathfrak{M}, (\emptyset, S) \models \Box \perp$  is undecidable.

We shall focus on semantics known to have decidable truth conditions. Let  $\mathcal{L}(\Phi)|_{\text{Prop}}$  be the restriction of  $\mathcal{L}(\Phi)$  to propositional formula. For clarification, we may observe that the BNF grammar of  $\mathcal{L}(\Phi)|_{\text{Prop}}$  is:

$$\phi_{\text{Prop}} ::= p \in \Phi \mid \phi_{\text{Prop}} \vee \psi_{\text{Prop}} \mid \neg \phi_{\text{Prop}}$$

Next let  $\wp_\omega X$  denote the finite subsets of  $X$ , and  $X \subseteq_\omega Y$  express that  $X$  is a finite subset of  $Y$ . One may then show by induction that if  $\mathfrak{M} \subseteq_\omega (\wp_\omega \Phi) \times (\wp_\omega \mathcal{L}(\Phi)|_{\text{Prop}})^{\mathcal{A}}$ , then  $\mathfrak{M}, (a, A) \models \phi$  is decidable in finite time for all  $\phi \in \mathcal{L}(\Phi)$  and for all  $(a, A) \in \mathfrak{M}$ , under the proviso that all of the functions  $(\wp_\omega \mathcal{L}(\Phi)|_{\text{Prop}})^{\mathcal{A}}$  we are considering are also decidable. Throughout the rest of this discussion, we shall adhere to this restriction.

### 3 Failure of Compactness

Before proceeding to the axiomatic theory of EVIL, we shall demonstrate that the logic is not compact by giving an example of an infinite set of formula for which every finite subset is satisfiable while the entirety is not. In the subsequent discussion we shall restrict ourselves to models with one agent and cease to employ subscripts.

**Lemma 3.0.1.** *Let  $\tau : \Phi \rightarrow \mathcal{L}$  be defined as follows:*

$$\begin{aligned} \tau(p) &:= p \wedge \Diamond \top \wedge \\ &\quad \Box (\neg p \wedge \Diamond \top \wedge \\ &\quad \Box (p \wedge \Diamond p)) \end{aligned}$$

*Every finite subset of  $\tau[\Phi]$  is satisfiable, but not the entirety, which is infinite.*  $\dashv$

*Proof.* That  $\tau[\Phi]$  is infinite is immediate, as  $\Phi$  was stipulated to be infinite.

So let  $S \subseteq \tau[\Phi]$  be finite. We shall find a model that satisfies  $S$ . First observe that there is a finite  $\Psi \subset \Phi$  such that  $S = \tau[\Psi]$ . Let  $\{q, r, s\} \subseteq \Phi \setminus \Psi$  be three distinct letters; we know that three exist for  $\Phi \setminus \Psi$  is infinite. Let  $\mathfrak{M}$  be as in Fig. 1, where

$$\begin{aligned} a &:= \Psi \cup \{s\} \\ b &:= \{q\} \\ c &:= \Psi \cup \{r\} \end{aligned}$$

One may check that  $\mathfrak{M}, (a, \{q\}) \models \tau(p)$  for all  $p \in \Psi$ .

On the other hand, suppose there was some model  $\mathfrak{N}$  such that  $\mathfrak{N}, (a, A) \models \tau(p)$  for all  $p \in \Psi$ . This implies that  $\mathfrak{N}, (a, A) \models p$  for each  $p \in \Phi$ . Moreover, let  $(b, B)$  be such that  $\mathfrak{N}, (b, B) \models A$  (we know one exists since by hypothesis  $\mathfrak{N}, (a, A) \models \Diamond p$ ). By the semantics of  $\tau(p)$ , it is evident

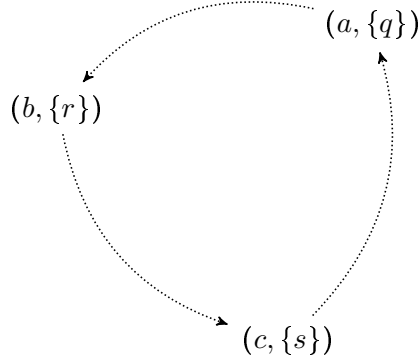


Figure 1: Model  $\mathfrak{M}$

that  $\mathfrak{N}, (b, B) \models \neg p$  and that there's a  $(c, C)$  such that  $\mathfrak{N}, (c, C) \models B$  and  $\mathfrak{N}, (c, C) \models p \wedge \Diamond p$ . But then from this we may infer that  $a$  and  $c$  both contain exactly the same sentence letters, which means that  $\mathfrak{N}, (a, A) \models B$  as well. However it cannot be that  $\mathfrak{N}, (a, A) \models \Diamond p$  since we have that  $\mathfrak{N}, (a, A) \models \Box \neg p$  by  $\tau(p)$ . Thus it is impossible that  $\mathfrak{N}, (a, A) \models \tau[\Phi]$ .  $\nmid$  QED

## 4 Axiomatics

### 4.1 Axioms and Rules

We now turn to providing an axiomatization for this system. In these  $\phi, \psi$ , and  $\chi$  are taken to be any formulae in  $\mathcal{L}(\Phi)$  while  $p$  taken as any letter in  $\Phi$ . With each axiom we provide an informal philosophical reading – see Table 1 for details.

**Lemma 4.1.1.** *We have the following provable equivalences:*

$$\begin{array}{lll}
 \vdash \Box_X \phi \leftrightarrow \Box_X \Box_X \phi & \vdash \Box_X \phi \leftrightarrow \Box_X \Box_Y \phi & \vdash \Box_X \phi \leftrightarrow \Box_X \Box_Y \phi \\
 \vdash \Box_X \phi \leftrightarrow \Box_X \Box_X \phi & \vdash \Box_X \phi \leftrightarrow \Box_X \Box_X \phi & \vdash \Box_X \leftrightarrow \Box_X \Box_X
 \end{array}$$

+

### 4.2 Soundness & Completeness

In the subsequent discussion, let the Tarski predicate ( $\models$ ) denote the usual Kripke semantics for the language  $\mathcal{L}(\Phi)$  for structures  $\langle W, V, P_X, R_{\Box_X}, R_{\Box_X}, R_{\Box_X} \rangle$ . The only unconventional thing we shall do is to declare  $P_X \subseteq W$  and  $\mathfrak{M}, w \models \Box_X$  if and only if  $w \in P_X$ . For more information on basic modal semantics see Blackburn et al. (2001). In the subsequent discussion we shall employ the Fraktur typeface (such as  $\mathfrak{M}$ ) to denote models in EVIL semantics and the blackboard typeface (such as  $\mathbb{M}$ ) to denote models in Kripke semantics.

From the definitions so far the following can be seen to hold:

(1)	$\vdash \phi \rightarrow \psi \rightarrow \phi$	
(2)	$\vdash (\phi \rightarrow \psi \rightarrow \chi) \rightarrow (\phi \rightarrow \psi) \rightarrow \phi \rightarrow \chi$	<i>Axioms for basic propositional logic</i>
(3)	$\vdash (\neg\phi \rightarrow \neg\psi) \rightarrow \psi \rightarrow \phi$	
(4)	$\vdash \boxplus_X \phi \rightarrow \phi$	<i>If <math>\phi</math> holds under any further evidence <math>X</math> considers, then <math>\phi</math> holds simpliciter, since considering no additional evidence is trivially considering further evidence</i>
(5)	$\vdash \boxplus_X \phi \rightarrow \boxplus_X \boxplus_X \phi$	<i>If <math>\phi</math> holds under any further evidence <math>X</math> considers, then <math>\phi</math> holds whenever <math>X</math> considers even further evidence beyond that</i>
(6)	$\vdash p \rightarrow \boxplus_X p$	
(7)	$\vdash p \rightarrow \boxplus_X p$	<i>Changing one's mind does not bear on matters of fact</i>
(8)	$\vdash \Diamond_X \phi \rightarrow \boxplus_X \Diamond_X \phi$	<i>The more evidence <math>X</math> discards, the freer her imagination can run</i>
(9)	$\vdash \Box_X \phi \rightarrow \Box_X \boxplus_X \phi$	
(10)	$\vdash \Box_X \phi \rightarrow \Box_X \boxplus_X \phi$	<i>If <math>X</math> believes a proposition, she believes it regardless of what anyone else thinks</i>
(11)	$\vdash \bigcirc_X \rightarrow \Box_X \phi \rightarrow \phi$	<i>If <math>X</math>'s premises are sound, then her logical conclusion are correct</i>
(12)	$\vdash \bigcirc_X \rightarrow \boxplus_X \bigcirc_X$	<i>If <math>X</math>'s premises are sound then any subset of will be sound as well</i>
(13)	$\vdash \phi \rightarrow \boxplus_X \Diamond_X \phi$	
(14)	$\vdash \phi \rightarrow \boxplus_X \Diamond_X \phi$	<i>Embracing evidence is the inverse of discarding evidence</i>
(15)	$\vdash \Box_X (\phi \rightarrow \psi) \rightarrow \Box_X \phi \rightarrow \Box_X \psi$	
(16)	$\vdash \boxplus_X (\phi \rightarrow \psi) \rightarrow \boxplus_X \phi \rightarrow \boxplus_X \psi$	<i>Variations on axiom K</i>
(17)	$\vdash \boxplus_X (\phi \rightarrow \psi) \rightarrow \boxplus_X \phi \rightarrow \boxplus_X \psi$	
(I)	$\frac{\vdash \phi \rightarrow \psi \quad \vdash \phi}{\vdash \psi}$	<i>Modus Ponens</i>
(II)	$\frac{\vdash \phi}{\vdash \Box_X \phi}$	
(III)	$\frac{\vdash \Box_X \phi}{\vdash \phi}$	<i>Variations on necessitation</i>
(IV)	$\frac{\vdash \phi}{\vdash \boxplus_X \phi}$	

Table 1: A Hilbert style axiom system for EVIL

**Lemma 4.2.1** (Soundness). *If  $\vdash \phi$  then for any model  $\mathfrak{M}$  and any  $(a, A) \in \mathfrak{M}$  we have that  $\mathfrak{M}, (a, A) \models \phi$*  →

The proof of the converse, that is *completeness*, proceeds by a three stage construction:

- The first step is to construct a Kripke model  $\mathfrak{T}^\phi$  consisting of finite maximally consistent sets of formulae related to  $\phi$  where  $\mathfrak{T}^\phi, w \not\models \phi$  for some world  $w \in W^{\mathfrak{T}^\phi}$ . This model will be shown to make true nine properties.
- The second step is to construct a model  $\mathfrak{I}^{\mathfrak{T}^\phi}$  which is bisimilar to  $\mathfrak{T}^\phi$ . This model also makes true these nine properties as well as an additional tenth property.
- The final third step is to construct an EVIL model  $\mathfrak{X}_\phi^{\mathfrak{I}^{\mathfrak{T}^\phi}}$ . We shall then show that for each  $w \in W^{\mathfrak{I}^{\mathfrak{T}^\phi}}$  there is a corresponding  $(a, A) \in \mathfrak{X}_\phi^{\mathfrak{I}^{\mathfrak{T}^\phi}}$  such that  $\mathfrak{I}^{\mathfrak{T}^\phi}, w \models \psi$  if and only if  $\mathfrak{X}_\phi^{\mathfrak{I}^{\mathfrak{T}^\phi}}, (a, A) \models \psi$  for all subformulae  $\psi$  of  $\phi$ .

These three steps together suffice to prove completeness. We shall now proceed to demonstrate these constructions.

#### 4.2.1 Subformula Model Construction

In this section we provide definitions and lemmas related to the subformula construction  $\mathfrak{T}^\phi$ . We consciously imitate Boolos (1993) in our approach, as well as the “Fischer-Ladner Closure” used in the completeness theorem of PDL (Blackburn et al., 2001).

**Definition 4.2.2.**

$$\sim \phi := \begin{cases} \psi & \text{if } \phi = \neg \psi \\ \neg \phi & \text{o/w} \end{cases} \quad \boxtimes_X \phi := \begin{cases} \phi & \text{if } \phi = \boxtimes_X \psi \\ \boxtimes_X \phi & \text{o/w} \end{cases} \quad \boxtimes_X \phi := \begin{cases} \phi & \text{if } \phi = \boxplus_X \psi \\ \boxplus_X \phi & \text{o/w} \end{cases}$$

→

**Lemma 4.2.3.** *By Lemma 4.1.1 we have*

$$\vdash \sim \phi \leftrightarrow \neg \phi \quad \vdash \boxtimes_X \phi \leftrightarrow \boxtimes_X \phi \quad \vdash \boxtimes_X \phi \leftrightarrow \boxplus_X \phi$$

... moreover ...

$$\boxtimes_X \phi = \boxtimes_X \boxtimes_X \phi \quad \boxtimes_X \phi = \boxtimes_X \boxplus_X \phi$$

→

**Definition 4.2.4.** *Let  $\delta(\phi) \subseteq \mathcal{A}$  be the set of agents that occur in  $\phi^2$*  →

**Definition 4.2.5.** *Define  $\Sigma(\Delta, \phi)$  using primitive recursion as follows:*

---

<sup>2</sup>In natural language, we read  $\delta(\phi)$  as “the dudes mentioned by  $\phi$ .”

$$\begin{aligned}
\Sigma(\Delta, p) &:= \{p, \neg p, \perp, \neg \perp\} \cup \bigcup \{\{\Box_X p, \neg \Box_X p, \Box_X p, \neg \Box_X p\} \mid X \in \Delta\} \\
\Sigma(\Delta, \perp) &:= \{\perp, \neg \perp\} \\
\Sigma(\Delta, \mathcal{O}_X) &:= \{\mathcal{O}_X, \neg \mathcal{O}_X, \Box_X \mathcal{O}_X, \neg \Box_X \mathcal{O}_X, \perp, \neg \perp\} \\
\Sigma(\Delta, \phi \rightarrow \psi) &:= \{\phi \rightarrow \psi, \neg(\phi \rightarrow \psi)\} \cup \Sigma(\Delta, \phi) \cup \Sigma(\Delta, \psi) \\
\Sigma(\Delta, \Box_X \phi) &:= \{\Box_X \phi, \neg \Box_X \phi, \Box_X \Box_X \phi, \neg \Box_X \Box_X \phi\} \\
&\quad \cup \bigcup \{\{\Box_X \Box_Y \phi, \neg \Box_X \Box_Y \phi, \Box_X \Box_Y \phi, \neg \Box_X \Box_Y \phi, \Box_Y \phi, \neg \Box_Y \phi, \Box_Y \phi, \neg \Box_Y \phi\} \mid Y \in \Delta\} \\
&\quad \cup \Sigma(\Delta, \phi) \\
\Sigma(\Delta, \Box_X \phi) &:= \{\Box_X \phi, \neg \Box_X \phi\} \cup \Sigma(\Delta, \phi) \\
\Sigma(\Delta, \Box_X \phi) &:= \{\Box_X \phi, \neg \Box_X \phi\} \cup \Sigma(\Delta, \phi)
\end{aligned}$$

→

**Lemma 4.2.6.**  $\Sigma(\delta(\phi), \phi)$  is finite. Moreover, we have the following:

- If  $\psi \in \Sigma(\delta(\phi), \phi)$  then  $\sim \psi \in \Sigma(\delta(\phi), \phi)$
- If  $\psi \in \Sigma(\delta(\phi), \phi)$  and  $\chi$  is a subformula of  $\psi$ , then  $\chi \in \Sigma(\delta(\phi), \phi)$
- If  $\Box_X \phi \in \Sigma(\delta(\phi), \phi)$  then  $\Box_X \phi \in \Sigma(\delta(\phi), \phi)$
- If  $\Box_X \phi \in \Sigma(\delta(\phi), \phi)$  then  $\Box_X \phi \in \Sigma(\delta(\phi), \phi)$

→

**Definition 4.2.7.** Let  $At(\Psi)$  denote the maximally consistent subsets of  $\Psi$

→

**Lemma 4.2.8** (Lindenbaum Lemma). If  $\Gamma \not\vdash \phi$  and  $\Gamma \subseteq \Sigma(\delta(\phi), \phi)$ , then there is a  $\Gamma' \in At(\Sigma(\delta(\phi), \phi))$  such that  $\Gamma \subseteq \Gamma'$  and  $\Gamma' \not\vdash \phi$

→

**Definition 4.2.9.** Define  $\mathbf{t}^\phi := \langle W^{\mathbf{t}^\phi}, V^{\mathbf{t}^\phi}, P_X^{\mathbf{t}^\phi}, R_{\Box_X}^{\mathbf{t}^\phi}, R_{\Box_X}^{\mathbf{t}^\phi}, R_{\Box_X}^{\mathbf{t}^\phi} \rangle$  where:

$$\begin{aligned}
W^{\mathbf{t}^\phi} &:= At(\Sigma(\delta(\phi), \phi)) \\
V^{\mathbf{t}^\phi}(p) &:= \{w \in W^{\mathbf{t}^\phi} \mid p \in w\} \\
P_X^{\mathbf{t}^\phi} &:= \{w \in W^{\mathbf{t}^\phi} \mid \mathcal{O}_X \in w\} \cup \{w \in W^{\mathbf{t}^\phi} \mid X \notin \delta(A)\} \\
R_{\Box_X}^{\mathbf{t}^\phi} &:= \{(w, v) \in W^{\mathbf{t}^\phi} \times W^{\mathbf{t}^\phi} \mid \{\psi \mid \Box_X \psi \in w\} \subseteq v\} \\
R_{\Box_X}^{\mathbf{t}^\phi} &:= \{(w, v) \in W^{\mathbf{t}^\phi} \times W^{\mathbf{t}^\phi} \mid \bigcup \{\{\psi, \Box_X \psi\} \mid \Box_X \psi \in w\} \subseteq v \wedge \bigcup \{\{\psi, \Box_X \psi\} \mid \Box_X \psi \in v\} \subseteq w\} \\
R_{\Box_X}^{\mathbf{t}^\phi} &:= \{(v, w) \in W^{\mathbf{t}^\phi} \times W^{\mathbf{t}^\phi} \mid \bigcup \{\{\psi, \Box_X \psi\} \mid \Box_X \psi \in w\} \subseteq v \wedge \bigcup \{\{\psi, \Box_X \psi\} \mid \Box_X \psi \in v\} \subseteq w\}
\end{aligned}$$

→

**Lemma 4.2.10** (Truth Lemma). For any subformula  $\psi \in \Sigma(\delta(\phi), \phi)$  and any  $w \in W^{\mathbf{t}^\phi}$ , we have that  $\mathbf{t}^\phi, w \Vdash \psi$  if and only if  $\psi \in w$

*Proof.* The proof proceeds by induction on  $\psi$ . Most of the steps are routine, with the exception of the right to left directions for the boxes.

We shall demonstrate the right to left direction for  $\boxtimes_X$ . Assume that  $\boxtimes_X \psi \notin w$ , then  $w \not\vdash \boxtimes_X \psi$ . By Lemma 4.2.3 this is true if and only if  $w \not\vdash \boxtimes_X \psi$ . Now abbreviate:

$$\begin{aligned} A &:= \bigcup \{ \{ \chi, \boxtimes_X \chi \} \mid \boxtimes_X \chi \in w \} \\ B &:= \{ \sim \boxtimes_X \chi \mid \boxtimes_X \chi \in \Sigma(\delta(\phi), \phi) \wedge \sim \chi \in w \} \end{aligned}$$

Now suppose towards a contradiction that  $\{ \sim \psi \} \cup A \cup B \vdash \perp$ . Then  $A \cup B \vdash \psi$ , and furthermore by Lemma 4.2.3 and rule (III) from the axioms we have that  $\boxtimes_X A \cup \boxtimes_X B \vdash \boxtimes_X \psi$ .<sup>3</sup> But then let

$$\begin{aligned} A' &:= \{ \boxtimes_X \chi \mid \boxtimes_X \chi \in w \} \\ B' &:= \{ \sim \chi \mid \sim \chi \in w \} \end{aligned}$$

Since  $\boxtimes_X \boxtimes_X \chi = \boxtimes_X \chi$  by Lemma 4.2.3, we have  $A' = \boxtimes_X A$ . Moreover, by Lemma 4.2.3, axiom 13, and classical logic we can see that

$$\vdash \sim \chi \rightarrow \boxtimes_X \sim \boxtimes_X \psi$$

Thus for every  $\beta \in \boxtimes_X B$  we have that  $B' \vdash \beta$ . Hence by  $n$  applications of the Cut rule we can arrive at

$$A' \cup B' \vdash \boxtimes_X \chi$$

However, evidently  $A' \cup B' \subseteq w$ , hence  $w \vdash \boxtimes_X \psi$ , which contradicts what has been stipulated.  $\nexists$

Hence it must be that  $\{ \sim \psi \} \cup A \cup B \not\vdash \perp$ . In addition, from the fact that  $w \subseteq \Sigma(\delta(\phi), \phi)$  with Lemma 4.2.6 and the hypothesis we have that  $\{ \sim \psi \} \cup A \cup B \subseteq \Sigma(\delta(\phi), \phi)$ . Hence by the Lindenbaum Lemma we have that there is some  $v \in At(\Sigma(\delta(\phi), \phi))$  such that  $\{ \sim \psi \} \cup A \cup B \subseteq v$ . By the inductive hypothesis we have that  $\mathfrak{t}^\phi, v \not\vdash \psi$ .

To complete the argument, we have to show that  $w R_{\boxtimes_X}^{\mathfrak{t}^\phi} v$ . Since  $A \subseteq v$  we just need to check that  $\bigcup \{ \{ \psi, \boxtimes_X \psi \} \mid \boxtimes_X \psi \in v \} \subseteq w$ . Suppose that  $\boxtimes_X \psi \in v$  but  $\psi \notin w$ . Since  $w$  is maximally consistent we have then that  $\psi \notin w$ . Thus  $\sim \boxtimes_X \psi \in v$ , which contradicts that  $v$  is consistent.  $\nexists$  Now suppose that  $\boxtimes_X \psi \in v$  but  $\boxtimes_X \psi \notin w$ , hence  $\sim \boxtimes_X \psi \in w$  and thus  $\sim \boxtimes_X \boxtimes_X \psi \in v$ . However we know from Lemma 4.2.3 that  $\boxtimes_X \boxtimes_X \psi = \boxtimes_X \psi$ , which once again implies that  $v$  is inconsistent.  $\nexists$  QED

**Lemma 4.2.11** ( $\mathfrak{t}^\phi$  is Partly EVIL).  $\mathfrak{t}^\phi$  makes true the following properties:

- (1)  $R_{\boxtimes_X}^{\mathfrak{t}^\phi} \subseteq W^{\mathfrak{t}^\phi} \times W^{\mathfrak{t}^\phi}$
- (2)  $W^{\mathfrak{t}^\phi}$  is finite
- (3) For all  $w \in W^{\mathfrak{t}^\phi}$  we have  $w R_{\boxtimes_X}^{\mathfrak{t}^\phi} w$
- (4) If  $w R_{\boxtimes_X}^{\mathfrak{t}^\phi} v$  and  $v R_{\boxtimes_X}^{\mathfrak{t}^\phi} z$  then  $w R_{\boxtimes_X}^{\mathfrak{t}^\phi} z$
- (5)  $R_{\boxtimes_X}^{\mathfrak{t}^\phi} = (R_{\boxtimes_X}^{\mathfrak{t}^\phi})^{-1}$
- (6) If  $w R_{\boxtimes_X}^{\mathfrak{t}^\phi} v$  then  $w \in V^{\mathfrak{t}^\phi}(p)$  if and only if  $v \in V^{\mathfrak{t}^\phi}(p)$
- (7) If  $w R_{\boxtimes_X}^{\mathfrak{t}^\phi} v$  and  $v R_{\boxtimes_X}^{\mathfrak{t}^\phi} u$  then  $w R_{\boxtimes_X}^{\mathfrak{t}^\phi} u$
- (8) If  $w R_{\boxtimes_X}^{\mathfrak{t}^\phi} v$  then  $u R_{\boxtimes_X}^{\mathfrak{t}^\phi} w$  if and only if  $u R_{\boxtimes_X}^{\mathfrak{t}^\phi} v$

---

<sup>3</sup>Here  $\boxtimes_X S$  is shorthand for  $\{ \boxtimes_X \chi \mid \chi \in S \}$ .



(9) If  $w \in P_X^{\dagger^\phi}$  then  $wR_{\Box_X}^{\dagger^\phi}w$

... for all  $\{X, Y\} \subseteq \mathcal{A}$ . Any model with the same modal similarity type as  $\dagger^\phi$  that makes the above true is said to be **partly EViL**  $\dashv$

Unfortunately, while  $\dagger^\phi$  is nearly what is necessary to derive completeness for our semantics, it is not perfect. Another stage of the construction is necessary.

## 4.2.2 Bisimulation

We first introduce a Backus-Naur form grammar for the **Either** type constructor, which may be viewed as a coproduct in category theory (in the category of Sets)<sup>4</sup>:

$$\text{Either } a \text{ } b ::= a_l \mid b_r$$

**Definition 4.2.12.** Let  $\mathbb{M}$  be a Kripke model, then define  $\dagger^{\mathbb{M}}$  as a model

$$\langle W^{\dagger^{\mathbb{M}}}, V^{\dagger^{\mathbb{M}}}, P_X^{\dagger^{\mathbb{M}}}, R_{\Box_X}^{\dagger^{\mathbb{M}}}, R_{\Box_X}^{\dagger^{\mathbb{M}}}, R_{\Box_X}^{\dagger^{\mathbb{M}}} \rangle$$

... where ...

$$\begin{aligned} W^{\dagger^{\mathbb{M}}} &:= \bigcup \{ \{w_l, w_r\} \mid w \in W^{\mathbb{M}} \} \\ V^{\dagger^{\mathbb{M}}}(p) &:= \bigcup \{ \{w_l, w_r\} \mid w \in V^{\mathbb{M}}(p) \} \\ P_X^{\dagger^{\mathbb{M}}} &:= \bigcup \{ \{w_l, w_r\} \mid w \in P_X^{\mathbb{M}} \} \\ R_{\Box_X}^{\dagger^{\mathbb{M}}} &:= \bigcup \{ \{ (w_l, v_r), (w_r, v_l) \} \mid wR_{\Box_X}^{\mathbb{M}}v \wedge w \notin P_X^{\mathbb{M}} \} \cup \bigcup \{ \{w_l, w_r\} \times \{v_l, v_r\} \mid wR_{\Box_X}^{\mathbb{M}}v \wedge w \in P_X^{\mathbb{M}} \} \\ R_{\Box_X}^{\dagger^{\mathbb{M}}} &:= \bigcup \{ \{ (w_l, v_l), (w_r, v_r) \} \mid wR_{\Box_X}^{\mathbb{M}}v \} \\ R_{\Box_X}^{\dagger^{\mathbb{M}}} &:= \bigcup \{ \{ (w_l, v_l), (w_r, v_r) \} \mid wR_{\Box_X}^{\mathbb{M}}v \} \end{aligned}$$

$\dashv$

**Lemma 4.2.13.** For any Kripke model  $\mathbb{M} = \langle W, V, P_X, R_{\Box_X}, R_{\Box_X}, R_{\Box_X} \rangle$ , we have the following bisimulation  $Z$  between  $\mathbb{M}$  and  $\dagger^{\mathbb{M}}$ :

$$wZw_l \quad \& \quad wZw_r$$

**Lemma 4.2.14.** If  $\mathbb{M}$  is partly EViL then  $\dagger^{\mathbb{M}}$  is partly EViL as well. It also makes true another, novel property:

(10) If  $wR_{\Box_X}^{\dagger^{\mathbb{M}}}w$  then  $w \in P_X^{\dagger^{\mathbb{M}}}$

Any partly EViL Kripke model that makes true this tenth property is said to be **completely EViL**  $\dashv$

---

<sup>4</sup>Either is taken from the functional programming language **Haskell**

### 4.2.3 Translation

In the subsequent discussion, it will be useful to exploit certain properties of *partly* EVIL models. To this end we introduce the concept of a *column*.

**Definition 4.2.15.** *Let  $\mathbb{M}$  be a partly EVIL Kripke structure. We shall make the following definition:*

$$\ulcorner w \urcorner_{\mathbb{M}} := \{v \mid w(R_{\Box X}^{\mathbb{M}} \cup R_{\Box X}^{\mathbb{M}})^* v\}$$

... where  $R^*$  is the reflexive transitive closure of  $R$

→

**Lemma 4.2.16** (Column Lemma). *The following hold if  $\mathbb{M}$  is partly EVIL:*

- (1) For all  $w$  we have  $w \in \ulcorner w \urcorner_{\mathbb{M}}$
- (2) If  $w \in \ulcorner v \urcorner_{\mathbb{M}}$  then  $\ulcorner w \urcorner_{\mathbb{M}} = \ulcorner v \urcorner_{\mathbb{M}}$
- (3) If  $wR_{\Box X}^{\mathbb{M}} v$  then for all  $u \in \ulcorner v \urcorner_{\mathbb{M}}$  we have  $wR_{\Box X}^{\mathbb{M}} u$
- (4) If  $w \in \ulcorner v \urcorner_{\mathbb{M}}$  then  $w \in V^{\mathbb{M}}(p)$  if and only if  $v \in V^{\mathbb{M}}(p)$  for all  $p \in \Phi$

→

**Definition 4.2.17.** *Let  $L(\phi) := \{p \in \Phi \mid p \text{ is a subformula of } \phi\}$*

*Let  $\mathbb{W}^{\mathbb{M}} := \bigcup \{\{\{w\}, \ulcorner w \urcorner_{\mathbb{M}}\} \mid w \in W^{\mathbb{M}}\}$*

*Let  $\rho_{\phi}^{\mathbb{M}} : \mathbb{W}^{\mathbb{M}} \rightarrow \Phi \setminus L(\phi)$  be an injection*

*Let  $\mathbb{E}\mathbb{I}\mathbb{S}_{\phi}^{\mathbb{M}} : W^{\mathbb{M}} \rightarrow \wp\Phi \times \wp(\mathcal{L}|_{\text{Prop}}(\Phi))$  be defined such that:*

$$\mathbb{E}\mathbb{I}\mathbb{S}_{\phi}^{\mathbb{M}}(w) := (\{p \in L(\phi) \mid \mathbb{M}, w \Vdash p\} \cup \{\rho_{\phi}^{\mathbb{M}}(\ulcorner w \urcorner_{\mathbb{M}})\}, \lambda X. \{\neg\rho_{\phi}^{\mathbb{M}}(\ulcorner v \urcorner_{\mathbb{M}}) \mid \neg wR_{\Box X}^{\mathbb{M}} v\} \cup \{\perp \rightarrow \rho_{\phi}^{\mathbb{M}}(\{v\}) \mid wR_{\Box X}^{\mathbb{M}} v\})$$

*Let  $\mathbb{A}_{\phi}^{\mathbb{M}} := \mathbb{E}\mathbb{I}\mathbb{S}_{\phi}^{\mathbb{M}}[W^{\mathbb{M}}]$*

→

**Lemma 4.2.18.** *Let  $\mathbb{M}$  be a completely EVIL Kripke structure. Then for any subformula  $\psi$  of  $\phi$  and any  $w \in W^{\mathbb{M}}$ , we have  $\mathbb{M}, w \Vdash \psi$  if and only if  $\mathbb{A}_{\phi}^{\mathbb{M}}, \mathbb{E}\mathbb{I}\mathbb{S}_{\phi}^{\mathbb{M}}(w) \models \psi$*

→

*Proof.* Apply induction. The only challenging cases involve the boxes, so we shall illustrate  $\Box_X \psi$ .

Assume that  $\mathbb{M}, w \not\Vdash \Box_X \psi$ , then there's some  $v \in W^{\mathbb{M}}$  such that  $wR_{\Box X}^{\mathbb{M}} v$  and  $\mathbb{M}, v \not\Vdash \psi$ . Let  $(a, A) := \mathbb{E}\mathbb{I}\mathbb{S}_{\phi}^{\mathbb{M}}(w)$  and  $(b, B) := \mathbb{E}\mathbb{I}\mathbb{S}_{\phi}^{\mathbb{M}}(v)$ . By the inductive hypothesis it suffices to show that  $\mathbb{A}_{\phi}^{\mathbb{M}}, (b, B) \models A_X$ . But the only things in  $A_X$  are tautologies or formulae of the form  $\neg\rho_{\phi}^{\mathbb{M}}(\ulcorner u \urcorner_{\mathbb{M}})$  where  $\neg wR_{\Box X}^{\mathbb{M}} u$ . But then Lemma 4.2.16 it can't be that  $\neg\rho_{\phi}^{\mathbb{M}}(\ulcorner v \urcorner_{\mathbb{M}}) \in A_X$ , and this suffices.

Now assume that  $\mathbb{A}_{\phi}^{\mathbb{M}}, (a, A) \not\models \Box_X \psi$  where  $(a, A) = \mathbb{E}\mathbb{I}\mathbb{S}_{\phi}^{\mathbb{M}}(w)$ , so there must be some  $v \in W^{\mathbb{M}}$  such that  $\mathbb{A}_{\phi}^{\mathbb{M}}, (b, B) \not\models \psi$  where  $(b, B) = \mathbb{E}\mathbb{I}\mathbb{S}_{\phi}^{\mathbb{M}}(v)$  and  $\mathbb{A}_{\phi}^{\mathbb{M}}, (b, B) \models A_X$ . By the inductive hypothesis it suffices to show that  $wR_{\Box X}^{\mathbb{M}} v$ , but this must be the case for otherwise  $\neg\rho_{\phi}^{\mathbb{M}}(\ulcorner v \urcorner_{\mathbb{M}}) \in A_X$  and then it couldn't be that  $\mathbb{A}_{\phi}^{\mathbb{M}}, (b, B) \models A_X$  since  $\rho_{\phi}^{\mathbb{M}}(\ulcorner v \urcorner_{\mathbb{M}}) \in B$ .

The inductive steps for the other boxes follow by similar reasoning.

QED

#### 4.2.4 Completeness

**Theorem 4.2.19.** *If  $\not\models \phi$  then there is some model  $\mathfrak{M}$  and some  $(a, A) \in \mathfrak{M}$  such that  $\mathfrak{M}, (a, A) \not\models \phi$*   $\dashv$

*Proof.* Assume  $\not\models \phi$ , then by Lemmas 4.2.10 and 4.2.11 we have some partly EViL Kripke structure and world such that  $\mathbb{A}, a \not\models \phi$ . By Lemma 4.2.13 we have that there is a completely EViL Kripke structure  $\mathbb{B}$  such that  $\mathbb{A} \leftrightarrow \mathbb{B}$ , thus there is some world  $b$  such that  $\mathbb{B}, b \not\models \phi$ . Finally by Lemma 4.2.18 we have that there's a model  $\mathfrak{C}$  in EViL semantics and a pair  $(c, C) \in \mathfrak{C}$  such that  $\mathfrak{C}, (c, C) \not\models \phi$ . QED

## 5 Complexity

In this section, we shall establish a lower bound on the computational complexity of EViL. We shall first prove the following lemma:

**Lemma 5.0.20.** *EViL with a single agent is a conservative extension of basic modal logic with just axiom  $K$ . That is, if  $\not\models_K \phi$  then  $\not\models_{\text{EViL}} \phi$*

*Similarly, EViL with  $m > n$  agents is a conservative extension of EViL with  $n$  agents*  $\dashv$

*Proof.* Assume that  $\not\models_K \phi$ , then we know from modal logic that there's a finite Kripke Structure  $\mathbb{M} := \langle W, V, R \rangle$  such and a world  $w \in W$  such that  $\mathbb{M}, w \not\models \phi$ . Now extend  $\mathbb{M}$  to  $\mathbb{M}' := \langle W, V, P, R_\square, R_\Box, R_\Box \rangle$  where

- $P := \{(v, v) \mid vRv\}$
- $R_\Box := R_\Box := \{(w, w) \mid w \in W\}$

... this model is trivially completely EViL. Moreover we know that  $\mathbb{M}$  is an elementary submodel of  $\mathbb{M}'$ , so  $\mathbb{M}', w \not\models \phi$ . Hence by the Lemma 4.2.18 we have a model  $\mathfrak{M}$  and  $(a, A) \in \mathfrak{M}$  such that  $\mathfrak{M}, (a, A) \not\models \phi$ ; so by soundness for EViL we have the desired result.

Similarly, if we  $\not\models_{\text{EViL}_A} \phi$  then by completeness can find a witnessing  $\mathfrak{M}$  and  $(a, A) \in \mathfrak{M}$  such that  $\mathfrak{M}, (a, A) \not\models \phi$ . But then we can embed  $\mathfrak{M}$  into  $\mathfrak{M}'$  for agents  $\mathcal{B} \supseteq \mathcal{A}$  where  $\mathfrak{M}' := \{(a, A') \mid (a, A) \in \mathfrak{M}\}$  and

$$A'_X := \begin{cases} A_X & X \in \mathcal{A} \\ \emptyset & X \notin \mathcal{A} \end{cases}$$

QED

**Lemma 5.0.21.** *EViL is PSPACE hard*  $\dashv$

*Proof.* This follows trivially from the fact that EViL is a conservative extension of basic modal logic, and the decision problem for basic modal logic is PSPACE complete. QED

## References

- S. Artemov. Logic of proofs. *Annals of Pure and Applied Logic*, 67(1):29–60, 1994.
- S. Artemov and E. Nogina. Introducing justification into epistemic logic, 2005.
- P. Blackburn, M. De Rijke, and Y. Venema. *Modal logic*. Cambridge Univ Pr, 2001. URL <http://logic.swu.edu.cn/upload/sharedocs/reference/P.BlackburnModalLogic.pdf>.
- G. Boolos. *The Logic of Provability*. Cambridge: Cambridge University Press, 1993.
- R. Descartes and S. Tweyman. *Meditations on first philosophy in focus*. Routledge, 1993.
- J. Hintikka. *Knowledge and belief: an introduction to the logic of the two notions*. Cornell University Press, 1962.