Evidentialist Logic

MSc Thesis (*Afstudeerscriptie*)

written by

Matthew P. Wampler-Doty
(born May 23rd, 1984 in Boston, Massachusetts, USA)

under the supervision of **Prof.dr J. F. A. K. van Benthem**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:	**Members of the Thesis Committee:**
August 24, 2010	Prof. Jan van Eijck
	Dr. Alessandra Palmigianio
	Prof. Frank Veltman

Institute for Logic, Language and Computation

Contents

Chapter 1

Philosophy

1.1 Forward

The purpose of this thesis is to present a formal framework which tries to present a novel modal logic for reasoning about knowledge. Subsequently, we shall conform to the following structure:

§1 First, we shall elaborate the on our philosophical intuition behind our epistemic logic, and provide a sketch of how the system will ultimately be formulated.

§2 Next, we give formal details of the system we will develop. Single agent semantics for concrete models is developed and an elimination theorem is derived. The semantics are then extended to a multi-agent setting, and finally Kripke semantics as an abstraction on our previous concrete semantics.

§3 Here we present several axiom systems for the abstract and concrete semantics. We show via our investigated completeness results that Kripke semantics faithfully abstracts away from our concrete semantics. We also derive the small model property for all of the axiom systems we provide, and discuss their inter-relationships in terms of a lattice of conservative extensions. Some complexity results are also provided.

§4 Finally. we shall look at applications of the framework developed. We show that imposing certain popular axioms leads to collapse results for concrete EvIL models and Kripke semantics. We then provide three embeddings of *intuitionistic logic* into EvIL, and discuss some of their connections to the philosophical literature.

§5 Finally, the framework developed shall be compared to other approaches.

We now turn to motivating the philosophical ideas that will be engaged in this text.

1.2 Explicit Justification

In this document, we shall propose a logic that enforces a form of *explicit justification* of propositions agents believe. The hunt for logics of explicit justification was initiated in [vB91][1]. One framework which has been proposed to achieve this is *Justification Logic* [AN05, Art07, Fit04, Fit05]. Alternative frameworks for reasoning about implicit/explicit information have also been proposed in [vBV09] and [Vel09]. We will discuss these other approaches in §5.

To model beliefs with justifications, we shall modify the semantics of modal logic to incorporate certain *basic beliefs*, which we should interpret as noninferentially justified. These basic beliefs then inferentially generate the rest of what the agent believes.

This perspective amounts to what is called *classical* or *old fashioned foundationalism* in the philosophical literature. Richard Fumerton describes the view as follows:

> [A] foundationalist is someone who claims that there are noninferentially justified beliefs and that all justified beliefs owe their justification, ultimately, in part, to the existence of noninferentially justified beliefs.[2] A belief is noninferentially justified if its justification is not constituted by the having of other justified beliefs. [DeP01, pg. 3]

To be completely explicit, our aim is to try to specifically modify the semantics of *basic modal logic*, and incorporate ingredients for a foundationalist analysis of knowledge. As will be demonstrated, this can be done without modifying the basic modal logic syntax.

1.3 Sketch

In this section we shall see a very informal presentation of the basic elements which shall compose the forthcoming analysis. A formal development of the ideas sketched in this section shall be given in §2.1. With this proviso, consider the basic modal language $\mathcal{L}_K(\Phi)$:

$$\varphi ::= p \in \Phi \mid \varphi \to \psi \mid \bot \mid \Box\varphi$$

Further, let $\mathfrak{M} \subseteq \wp\Phi \times \wp\mathcal{L}_K(\Phi)$, that is, let $\mathfrak{M}$ be pairs of sets of letters and formulae. Define the following truth predicate $\Vdash$ recursively:

Definition 1.3.1.

$\mathfrak{M}, (a, A) \Vdash p \iff p \in a$

$\mathfrak{M}, (a, A) \Vdash \varphi \to \psi \iff \mathfrak{M}, (a, A) \Vdash \varphi$ *implies* $\mathfrak{M}, (a, A) \Vdash \psi$

$\mathfrak{M}, (a, A) \Vdash \bot \iff False$

$\mathfrak{M}, (a, A) \Vdash \Box\varphi \iff$ *for all* $(b, B) \in \mathfrak{M}$, *if* $\mathfrak{M}, (b, B) \Vdash A$ *then* $\mathfrak{M}, (b, B) \Vdash \varphi$

[1] While this paper is considered seminal, it should be remarked that research into this subject began prior to it. Specifically, the phrase "explicit belief" appears to have its origins in [Lev84].

[2] In the proceeding discussion, we shall refer to *noninferentially justified beliefs* as *basic beliefs* or premises.

We now provide motivation for the above definition.

In the semantics presented, instead of thinking of every world individually, we think of every world as containing facts and a part of the agent's mind. This part of the agent's mind is represented by what a basis of propositions which she ascents to. We shall refer to these interchangeably as *premises*, *assumptions*, *basic beliefs*, *experiences*, or *evidence*. These sets of propositions may be thought of as representing the agent's frame of mind. The intended reading of $\Box\varphi$ also involves these basic beliefs. We shall interpret $\mathfrak{M}, (a, A) \Vdash \Box\varphi$ as expressing that the agent can produce a proof of φ on the basis of her evidence and background knowledge. In Proposition 1.3.4, we provide a proof sketch of this. We also provide a more rigorous proof of the correctness of this reading as Theorem 2.1.12 in §2.1. We will alternately read $\Box\varphi$ as *the agent believes phi*, *the agent can deduce* φ or *can compose an argument for* φ.

As in the original formulation of epistemic logic in [Hin69], we assume that agents are *doxastically omniscient* - that is they believe all of the consequences of their beliefs. Hence, when we say that the agent can deduce φ, we mean that they may derive an argument for φ given enough time and cognitive resources. Moreover, we will restrict ourselves to only considering *justified beliefs*, since intuitively an unjustified beliefs do not necessarily have logical arguments associated with them.

The view on epistemic and doxastic logic suggested above holds that beliefs are justified by deductive steps which are ultimately grounded in a basis of evidence. This view is present in the *evidentialist* view on epistemology, which may be summarized as follows:

> [E]videntialism is a supervenience thesis according to which facts about whether or not a person is justified in believing a proposition supervene on facts describing the evidence that the person has. [CF04], pg. 5

Evidentialism is a form of foundationalism, as we previously mentioned in §1.2. It is a flexible framework which is agnostic about the details in which inferences come to be justified. The semantics we propose is more focussed, by specifying that the basic mechanism that connects foundational evidence to justified beliefs is restricted to logic.

The intended reading of $\Diamond\varphi$ is *the agent can conceive that* φ or *the agent can imagine* φ *being possible*. The former is the standard reading in epistemic logic (see, for instance, [MvdH95]). In general, we shall prefer to read $\Diamond$ as in terms of *imagination*.

The above semantics demands a grammar restriction, because of a variation on Russell's paradox that emerges otherwise. Let

- $a := \varnothing$
- $A := \{\Box\bot\}$
- and $\mathfrak{M} := \{(a, A)\}$

Under this assignment, $\mathfrak{M}, (a, A) \Vdash \Box\bot$ has no determinate truth value. In light of this, we will restrict belief basis sets to $\mathcal{L}_0(\Phi)$, the propositional fragment of $\mathcal{L}_K$. Formally, $\mathcal{L}_0(\Phi)$ is defined as follows:

$$\varphi ::= p \in \Phi \mid \varphi \to \psi \mid \bot$$

We shall restrict our models in such a manner that $\mathfrak{M} \subseteq \Phi \times \wp\mathcal{L}_0(\Phi)$. This suffices to make every truth value of this logic determinate.

Basic modal logic is weakly sound and complete for the semantics we have presented. The following proposition establishes this:

Proposition 1.3.2 (Weak Soundness and Completeness)**.** *Assuming that the set of proposition letters Φ is infinite, for all $\mathfrak{M}$ we have:*

$$\vdash_K \varphi \text{ if and only if } \mathfrak{M}, (a, A) \Vdash \varphi \text{ for all finite } \mathfrak{M} \text{ for all } (a, A) \in \mathfrak{M}$$

Here K is basic modal logic, as defined in [BRV01, chapter 4, pg. 194].

Proof. Left to right is trivial, so we shall focus on right to left. Assume that $\nvdash_K \varphi$, then we know from completeness and the finite model property that there's some finite model $\mathbb{M} = \langle W, V, R\rangle$ and world $w \in W$ such that $\mathbb{M}, w \nVdash \varphi$ (see [BRV01], chapters 2 & 4 for details of these facts).

Now let $L(\varphi)$ be the proposition letters that occur as subformulae of φ, and let $p : W \hookrightarrow \Phi \backslash L(\varphi)$ be an injection. In other words p assigns *fresh letters* to worlds in the model [3]. Define $\theta : W^{\mathbb{M}} \to \wp\Phi \times \wp\mathcal{L}_0(\Phi)$ as follows [4]

$$\theta(x) := (\{q \in \Phi \mid \mathbb{M}, w \Vdash q\} \cup \{p_w\}, \{ \bigvee_{v \in R[w]} p_v\})$$

Now let $\Theta := \theta[W]$. An induction on the complexity of subformulae ψ of φ shows that $\mathbb{M}, w \Vdash \psi \iff \Theta, \theta(w) \Vdash \psi$ for all $w \in W^{\mathbb{M}}$. Since $\mathbb{M}, w \nVdash \varphi$ then we know that $\Theta, \theta(w) \nVdash \varphi$, which completes the proof. QED

The basic principle behind the above construction is to label all of the worlds with fresh letters, and then construct a special formula from these fresh letters for each world. The extension of each of these formulae is, in every case, exactly the worlds the agent could have accessed with the accessibility relation. A more elaborate construction, based on the idea in the above construction, will be presented in §3.7.

We will illustrate that the semantics we have proposed are adequate for modeling agents according to our declared intentions. Recall the following definitions from basic logic and modal model theory[5]:

Definition 1.3.3. *(1) For any model $\mathfrak{M}$, define $Th(\mathfrak{M})$:*

$$Th(\mathfrak{M}) := \{\varphi \in \mathcal{L}_K(\Phi) \mid \mathfrak{M}, (a, A) \Vdash \varphi \text{ for all } (a, A) \in \mathfrak{M}\}$$

*$Th(\mathfrak{M})$ is called **the theory of** $\mathfrak{M}$.*

(2) $A \subseteq_\omega B$ means that A is a finite subset of B

[3] In this vein, we shall abbreviate $p(w)$ as p_w. Note that because $L(\varphi)$ is *finite* and Φ is assumed to be *infinite*, such an injection always exists. This is a consequence of *The Axiom of Choice*.

[4] The invention of this particular function should properly be attributed to Johan van Benthem.

[5] This notation consciously imitates the notation employed in [BRV01].

(*3*) *Define* $\Gamma \vdash_K \varphi$ *to mean:*

$$\vdash_K \bigwedge \Delta \to \varphi \text{ for some } \Delta \subseteq_\omega \Gamma$$

If $\Gamma \vdash \varphi$*, we say that* φ *is* ***derivable from*** Γ*.*

The following theorem equates belief in at a world in a model with possession of a derivation. We shall return to this in §2.1, where it shall be formally referred to as the *Theorem Theorem*, Theorem 2.1.12 of §2.1:

Proposition 1.3.4. *For all* $A \subseteq_\omega \mathcal{L}_0(\Phi)$*, then* $\mathfrak{M}, (a, A) \Vdash \Box\varphi$ *if and only if* $Th(\mathfrak{M}) \cup A \vdash_K \varphi$*.*

Proof. The proof of the above hinges on two basic facts. The first is the *deduction theorem* (provided that Δ is finite):

$$\Gamma \cup \Delta \vdash_K \varphi \iff \Gamma \vdash_K \bigwedge \Delta \to \varphi \tag{1.3.1}$$

The above follows from Definition 1.3.3 part (*3*), and is one of the standard results in modal logic.

The next observation is also rather basic:

$$Th(\mathfrak{M}) \vdash_K \varphi \iff \varphi \in Th(\mathfrak{M}) \tag{1.3.2}$$

The proof of this follows from the fact that if $\vdash_K \varphi$ then $\varphi \in Th(\mathfrak{M})$, and $Th(\mathfrak{M})$ can be observed to be closed under modus ponens.

So assume that $A \subseteq_\omega \mathcal{L}_0$. With the above key facts we have the following chain of reasoning:

$$\begin{aligned}
Th(\mathfrak{M}) \cup A \vdash_K \varphi &\iff Th(\mathfrak{M}) \vdash \bigwedge A \to \varphi && \text{by (1.3.1)} \\
&\iff \bigwedge A \to \varphi \in Th(\mathfrak{M}) && \text{by (1.3.2)} \\
&\iff \mathfrak{M}, (b, B) \Vdash \bigwedge A \to \varphi \text{ for all } (b, B) \in Th(\mathfrak{M}) && \text{by Def. 1.3.3 part (1)} \\
&\iff \mathfrak{M}, (a, A) \Vdash \Box\varphi \text{ for any } a \text{ where } (a, A) \in \mathfrak{M} && \text{by Def. 1.3.1}
\end{aligned}$$

These equivalences suffice to prove the result. QED

A natural way to read $Th(\mathfrak{M})$ is the background knowledge the agent has about the universe she lives in. This approach presents an analysis of modal logic whereby an idealized agent is modeled as closed under deduction; this is the *doxastic omniscience* mentioned previously. Under this view, evidently the agent's beliefs correspond to those things for which she has proofs. This shall be the perspective we urge for the intuition we shall employ in future investigations.

1.4 Soundness

In this section we propose a notion of knowledge we wish to investigate in this thesis. It is natural to insist that if knowledge is based on beliefs generated via deduction from some set of premises, then those premises have to be *sound*.

As a first attempt, we might try to accomplish this by introducing a new operator $\circlearrowleft$ with the following semantics:

$$\mathfrak{M}, (a, A) \Vdash\!\!\models\circlearrowleft \iff \mathfrak{M}, (a, A) \Vdash\!\!\models A$$

Armed with these semantics, a first guess at what constitutes knowledge suggests it might be nothing more than possession of a belief based on a sound set of premises. So a first approximation of knowledge might be equated with the formula:

$$\circlearrowleft \wedge \Box\varphi.$$

Is this anything like an adequate analysis of knowledge? **No**. To illustrate why, let us consider a thought experiment.

Imagine that Charlotte suspects, correctly, that if John has tried to murder on Alex, then Alex has survived. She further learns, correctly, that John has indeed tried to murder Alex. But later, she "learns" some erroneous information asserting Vietnam is south of Malaysia. If we codify all of this as a set C, and let the real world be denoted c and the universe $\mathfrak{M}$, evidently we have $\mathfrak{M}, (c, C) \nVdash\!\!\models\circlearrowleft$, so this previous definition of knowledge fails.

This illustrates that the previously proposed definition of knowledge is inadequate. Charlotte's knowledge about John's unspeakable betrayal of Alex is correct, as well as her inference that Alex is tough as nails. Just because she has been deluded regarding irrelevant facts about geography shouldn't have any bearing on her knowledge about Alex.

1.5 Descartes

In reflection on the previous section, it should be remarked that philosophers have historically been concerned with defeasible experiential data, going back at least as early as Plato's *The Republic VII* [Pla98]. In answer to the problem faced by the above analysis of knowledge, guidance can be found in Descartes' *Meditations* [VMTD05]. In *Meditations I*, Descartes suggests that he might be in an enlightenment era version of *The Matrix* created by an all powerful demon. In *Meditations II*, he famously suggests how one might escape this trap:

> The Meditation of yesterday has filled my mind with so many doubts, that it is no longer in my power to forget them. Nor do I see, meanwhile, any principle on which they can be resolved; and, just as if I had fallen all of a sudden into very deep water, I am so greatly disconcerted as to be unable either to plant my feet firmly on the bottom or sustain myself by swimming on the surface. 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.[VMTD05, *Meditations II*]

This tactic proposes a natural solution to the problem the previous thought experiment: *Charlotte can know that Alex survives if she argues* ***only*** *from her experience involving Alex and John.* If

like Descartes she can forget some of what she has come to believe that is a little suspicious, she might be able to compose an argument with a sound basis that Alex is alive.

Taking Descartes as inspiration, we are presented with a novel semantic operation:

$$\mathfrak{M}, (a, A) \Vdash \boxminus\varphi \iff \text{for all } (b, B) \in \mathfrak{M} \text{ such that } a = b \text{ and } B \subseteq A \text{ then } \mathfrak{M}, (b, B) \Vdash \varphi$$

This mechanism lets Charlotte access subsets of her beliefs, which would then form the basis for various arguments she might compose. Provided that $(c, C') \in \mathfrak{M}$, where C' is the same as C but doesn't mention erroneous beliefs about geographical data, it might serve as a basis for Charlotte's knowledge that Alex survives. This suggests that the following equation can reasonably express a more adequate notion of knowledge:

$$\diamondminus(\circlearrowleft \wedge \Box\varphi)$$

The above formulation of knowledge expresses that knowledge is the existence of an argument drawn from a sound basis of premises that are a subset of one's experience. This is the principle formulation of knowledge to which our intuition will be made to return throughout this text.

1.6 Embracing Evidence

To recap, so far we have suggested adding a novel modality $\boxminus$ which corresponds to taking subsets of an agent's set of beliefs. In the context of conventional modal logic, this means a shift in perspective - instead of thinking of each world as a situation where the agent can imagine other situations, now each world corresponds to a network of beliefs ordered by inclusion. These networks of beliefs form a poset, or partially ordered set. We will visually represent them as *Hasse diagrams*, as we have done in Fig. 1.1. This follows the standard practice in lattice theory.

Furthermore, consider the following phenomenon: as higher nodes in a belief network are considered, the agent is employing more premises for the arguments they are composing, and using less pure logic to come to conclusions. This suggests that as we consider levels higher and higher in the poset of an agent's beliefs, this corresponds to embracing an agent's experience and interpretation of their sensory data. Arguments that rest on more premises are prima facie more fallible than arguments that rely on fewer assumptions.

The above idea of *embracing evidence* can be formalized in the semantics we are developing in the following fashion. Figure 1.1 naturally suggests that we might think of *going up* in a belief net, in a manner similar to how $\boxminus$ allows one to *go down* as we proposed in §1.5. Hence we have introduced a new operator $\boxplus$, with the following semantics:

$$\mathfrak{M}, (a, A) \Vdash \boxplus\varphi \iff \text{for all } (b, B) \in \mathfrak{M} \text{ if } b = a \text{ and } A \subseteq B \text{ then } \mathfrak{M}, (a, A) \Vdash \varphi$$

Just as $\boxminus$ corresponds to the agent casting assumptions into doubt, or disregarding her premises, $\boxplus$ corresponds to the agent embracing their experience, suspending disbelief and accepting her intuitions and senses.

This concludes the our introductory outline of our method of modelling knowledge.

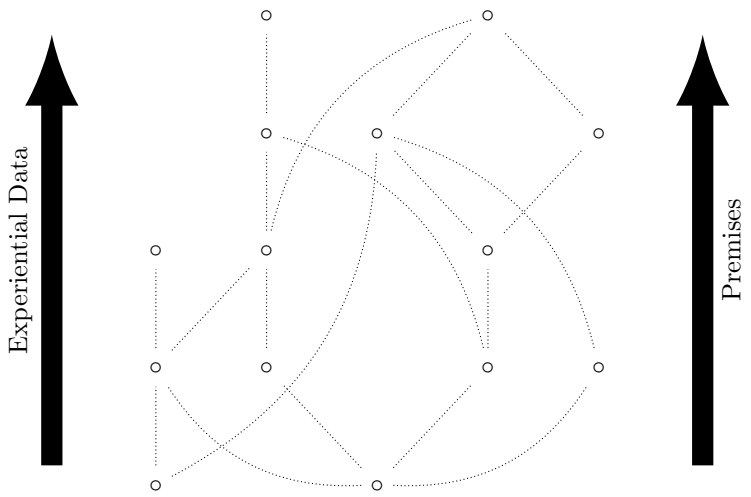


Figure 1.1: A network of beliefs

1.7 Closing Remarks

We may conclude that the logic we intend to develop in this essay shall satisfy the following criteria, based on the ideas provided in the previous sections:

§1.2 Agents shall be modelled with proofs for the things they believe.

For a set of beliefs A:

- **§1.4** It should be expressible whether everything in a basic set of beliefs A is sound
- **§1.5** Certain subsets $B \subseteq A$ of a set of beliefs should be accessible
- **§1.6** Certain extensions $B \supseteq A$ of a set of beliefs should also be accessible

In line with evidentialist epistemology, as mentioned in §1.3, we have decided to call the logic presented here *Evidentialist Logic*, or EvIL for brevity.

Chapter 2

Introduction to EviL

From with the philosophical intuitions and scaffolding provided from §1, we shall present a precise account of the previously developed ideas. This shall be done in three movements:

§2.1 In the first section we shall provide the basic grammar and semantics for EviL with a single agent; the presentation in this section will remain primarily philosophical and light.

§2.2 In the second section we develop several topics in the pure theory of EviL which are considered a bit beyond the bare essentials.

2.1 Elementary EviL

Grammar & Semantics

In this section we turn to developing the formal semantics for EviL with a single agent. We shall imagine the object of study in EviL is an agent, which we shall call the EviL agent. In §2.2, the semantic framework offered here is extended to incorporate multiple agents. In Appendix A, yet another framework is offered employing gamelike semantics, which avoids the grammar restriction suggested in §1.3.

The grammar restriction imposed on EviL was introduced to avoid paradoxes. That being the case, we shall discard the previous definition of ($\Vdash$) that was suggested in §1.3, in favor of demonstrably well-defined semantics. This shall be achieved in two steps.

Definition 2.1.5. *Let $\mathcal{L}_0(\Phi)$ be the language of classical propositional logic, defined by the following Backus-Naur form grammar:*

$$\varphi ::= p \in \Phi \mid \varphi \to \psi \mid \bot$$

Models for classical propositional logic can be thought of as sets $S \subseteq \Phi$; thus the truth predicate $(\models) : \wp\Phi \times \mathcal{L}_0(\Phi) \to \mathsf{bool}$ for classical propositional logic can be given recursively as follows:

Definition 2.1.6. *Define* $(\models)$ *such that*

$$\begin{aligned} S\models p &\iff p\in S \\ S\models\varphi\to\psi &\iff S\models\varphi \text{ implies } S\models\psi \\ S\models\bot &\iff False \end{aligned}$$

Further, observe that the language $\mathcal{L}_0$ is extended by EvIL

Definition 2.1.7. *Define* $\mathcal{L}(\Phi)$ *by the following Backus-Naur grammar:*

$$\varphi ::= p \in \Phi \mid \varphi \to \psi \mid \bot \mid \Box\varphi \mid \boxminus\varphi \mid \boxplus\varphi \mid \circlearrowleft$$

Unlike traditional modal logic, EvIL employs concrete models rather than Kripke structures. EvIL models are sets $\mathfrak{M} \subseteq \wp\Phi \times \wp\mathcal{L}_0(\Phi)$. Like classical propositional logic, semantics for EvIL are given recursively by a predicate $(\Vdash)$ which:

- Takes as input:
  - An EvIL model
  - A pair (a, A) where
    - $a \subseteq \Phi$ is a set of proposition letters
    - $A \subseteq \mathcal{L}_0(\Phi)$ is a set of propositional formulae.
  - A formula in the language $\mathcal{L}(\Phi)$
- Gives as output: a truth value in bool

More concisely, this may be written as

$$(\Vdash) \;:\; \wp(\wp\Phi \times \wp\mathcal{L}_0(\Phi)) \times (\wp\Phi \times \wp\mathcal{L}_0(\Phi)) \times \mathcal{L}(\Phi) \to \mathsf{bool}.$$

Definition 2.1.8. *Define* $(\Vdash)$ *recursively such that:*

$$\begin{aligned} \mathfrak{M},(a,A)\Vdash p &\iff p\in a \\ \mathfrak{M},(a,A)\Vdash\varphi\to\psi &\iff \mathfrak{M},(a,A)\Vdash\varphi \text{ implies } \mathfrak{M},(a,A)\Vdash\psi \\ \mathfrak{M},(a,A)\Vdash\bot &\iff False \\ \mathfrak{M},(a,A)\Vdash\Box\varphi &\iff \forall(b,B)\in\mathfrak{M}.(\forall\psi\in A.b\models\psi) \text{ implies } \mathfrak{M},(b,B)\Vdash\varphi \\ \mathfrak{M},(a,A)\Vdash\boxminus\varphi &\iff \forall(b,B)\in\mathfrak{M}.a=b \text{ and } B\subseteq A \text{ implies } \mathfrak{M},(b,B)\Vdash\varphi \\ \mathfrak{M},(a,A)\Vdash\boxplus\varphi &\iff \forall(b,B)\in\mathfrak{M}.a=b \text{ and } B\supseteq A \text{ implies } \mathfrak{M},(b,B)\Vdash\varphi \\ \mathfrak{M},(a,A)\Vdash\circlearrowleft &\iff \forall\psi\in A.a\models\psi \end{aligned}$$

Remark 2.1.9. *We will write* $\mathfrak{M} \Vdash \varphi$ *to mean* $\mathfrak{M},(a,A) \Vdash \varphi$ *for all* $(a,A) \in \mathfrak{M}$*. Further, we will write* $\Vdash \varphi$ *to mean* $\mathfrak{M} \Vdash \varphi$ *for all* $\mathfrak{M}$*.*

These semantics are well defined, since apart from relying on the semantics for propositional logic they may be observed to be compositional. Moreover, the following relationship can be observed:

Lemma 2.1.10 (Truthiness)**.** *Let $\varphi \in \mathcal{L}_0(\Phi)$. Then:*

$$a \models \varphi \Longleftrightarrow \mathfrak{M}, (a, A) \Vdash \varphi$$

for any $\mathfrak{M}$ and A.

Proof. This may be seen immediately by induction on φ. QED

With this, we have the following result:

Definition 2.1.11.

$$Th(\mathfrak{M}) := \{\varphi \in \mathcal{L}(\Phi) \mid \mathfrak{M} \Vdash \varphi\}$$

Theorem 2.1.12 (Theorem Theorem)**.** *If A is finite, then $\mathfrak{M}, (a, A) \Vdash \Box\varphi$ if and only if $Th(\mathfrak{M}) \cup A \vdash_{\text{EvIL}} \varphi$.*

Proof. I shall present $\vdash_{\text{EvIL}}$, the logical consequence turnstile for EvIL, in §3.1. For now, I will freely make use of important properties it possesses.

The proof proceeds the exactly as the proof of Proposition 1.3.4 from §1.3. We shall repeat it here. We shall assume that EvIL makes true the *deduction theorem* of classical logic:

$$\Gamma \cup \Delta \vdash_{\text{EvIL}} \varphi \iff A \vdash_{\text{EvIL}} \bigwedge B \to \varphi \tag{2.1.1}$$

Next, we use the following fact:

$$Th(\mathfrak{M}) \vdash_{\text{EvIL}} \varphi \iff \varphi \in Th(\mathfrak{M}) \tag{2.1.2}$$

This follows from the *soundness* of EvIL, and the fact that $Th(\mathfrak{M})$ is closed under modus ponens. We shall establish the soundness of EvIL in §3.7, in Theorem 3.3.44, the EvIL soundness and weak completeness theorem.

So assume that $A \subseteq_\omega \mathcal{L}_0$. Just as in Proposition 1.3.4, we have:

$$\begin{aligned}
Th(\mathfrak{M}) \cup A \vdash \varphi &\iff Th(\mathfrak{M}) \vdash \bigwedge A \to \varphi && \text{by (2.1.1)} \\
&\iff \bigwedge A \to \varphi \in Th(\mathfrak{M}) && \text{by (2.1.2)} \\
&\iff \mathfrak{M}, (b, B) \Vdash \bigwedge A \to \varphi \text{ for all } (b, B) \in Th(\mathfrak{M}) && \text{by Def. 2.1.11} \\
&\iff \mathfrak{M}, (a, A) \Vdash \Box\varphi \text{ for any } a \text{ where } (a, A) \in \mathfrak{M} && \text{by Def. 2.1.8}
\end{aligned}$$

QED

We have chosen the name "Theorem Theorem" because it means that for every belief the EvIL agent has, it is a theorem she has derived from her premises. Theorem 2.1.12 establishes one of the central desiderata outlined in §1.7 is achieved by EvIL. With this result the foundation is set for the the central intuition driving EvIL - that beliefs are the consequences of logical deductions.

It is a peculiarity of EvIL that these deductions are carried on in EvIL itself. This was achieved, primarily, by flirting heavily with paradox, as was illustrated in §1.3. As a consequence, we have tried to design EvIL to eat its own tail. It establishes that the EvIL agent is herself also a modeler just like us, using the same logic we are using to think about her herself, to think about the state space she lives in.

In summary theorem 2.1.12 is central to the conceptual backing behind EvIL. It provides the conceptual backbone of the perspective on epistemic logic this essay is intended to investigate.

Intuitions

In this section, we shall illustrate how we intuitively read the operators in EvIL, and provide a number of validities.

As per the traditional doxastic reading of $\Box\varphi$, we read this as asserting "The EvIL agent believes φ". Because of Theorem 2.1.12, the Theorem Theorem, we shall freely conflate this with the assertion "The EvIL agent has an argument for φ," which we take to be a kind of proof.

The intuition for how to read $\diamondsuit\varphi$ was first mentioned in §1.5 with respect to Descartes' Meditation II. It means "If the EvIL agent were to set aside some of her beliefs, or cast some of her beliefs into doubt, then φ would hold." Dually, we can read $\boxminus\varphi$ as saying something like "For all the ways that the EvIL agent might use her imagination, φ holds." One should recognize that these interpretations might seem inconsistent. These are not really an issue regard casting beliefs into doubt and embracing one's imagination as part of the same coin. For, naturally, when one doubts more things, then for a fleeting moment their dreams take flight as the inconceivable turns around into the conceivable, if only for a little while. To give an example, if Marta sets aside for a moment her belief that

$$\text{the law of gravity is an exceptionless regularity of the universe,} \tag{g}$$

then it seems natural that she might imagine that

$$\text{a propulsion device exploiting some exception to gravitation might be constructable.} \tag{p}$$

In the symbology of EvIL formulae, she would code this intuition as

$$\boxminus\,(\Box\neg g \to \Diamond p). \tag{2.1.3}$$

To give another example, if Marta pretends that it is not the case that:

$$\text{the canals of Amsterdam are filthy} \tag{f}$$

She might be able to imagine a scenario where

$$\text{she may swim comfortably in the Amstel river} \tag{r}$$

But not really. Marta really cannot really swim at ease in the Amstel, not just because it has tons of garbage, but also because

$$\text{she does not own a bathing suit,} \tag{b}$$

Frankly, Marta is not so bold that she could go skinny dipping in Amstel without that being awkward for her. In the language of EvIL, this thought experiment would be expressed as follows:

$$\neg \boxminus (\Box \neg f \to \Diamond r) \tag{2.1.4}$$

This is because Marta can cast into doubt the assumption of the filthiness of the canals of Amsterdam, while still retaining her belief that she does not have a bathingsuit, so swimming in Amstel would still be awkward for me. In symbols, she would write express this other sentiment as something of a refinement on (2.1.4), which is expressed as follows:

$$\circledcirc(\Diamond \neg f \wedge \Box b \wedge \neg \Diamond r) \tag{2.1.5}$$

Further, the intuition for how to read $\circledcirc\varphi$ is "If the EvIL agent were to remember something, then φ would hold." For instance, imagine a scenario where Marta wakes up and searches herself for her bike keys. To her horror, the keys are not there – and Marta immediately assumes that she might have left her keys in the lock on her bike, and figures there is a fair likelihood that

$$\text{the bike has been stolen because the keys were left in the lock.} \tag{s}$$

But once she recalls that

$$\text{she lent her bike to a friend,} \tag{l}$$

her fear subsides. Prior to remembering, while Marta thought it might be possible that her bike was stolen due to her own negligence, if she remembered what she had done then she no longer would have entertained this possibility. This observation is expressed as:

$$\Diamond s \wedge \boxplus(\Box l \to \Box \neg s) \tag{2.1.6}$$

We consider $\boxminus$ and $\boxplus$ to be inverse modalities of each other, in exactly the same way that *past* and *future* are inverse modalities in temporal logic. This is perhaps a little unusual; it is arguably more natural to think of *forgetting* as the inverse modality of remembering, and there does not appear to be an natural inverse operation corresponding to casting into doubt. Following the idea of the *web of belief* due to Quine, as presented in §1.6, we would extend a position asserting that remembering factive data is the same as embracing as much of one's evidence as possible.

Finally, reading considering the agent's web of beliefs as a network of related sets of premises, we may draw one final philosophical insight, which is reflective of a validity. A sound set of premises A intuitively will ensure any subset of them is also sound.

With this final intuition in mind, we turn to illustrating how the above philosophical intuitions give rise to an assortment of validities for EvIL.

Validities

The previous philosophical readings of EvIL immediately suggest certain validities will hold in the semantics. For instance, the assertion "A set of premises is sound if and only if all of its subsets are sound." would be expressed as

$$\Vdash \circlearrowleft \leftrightarrow \boxminus \circlearrowleft \tag{2.1.7}$$

Indeed, this is a validity of EvIL. Schematically, it may be tempting to think that maybe the same is true for $\boxplus$. However, we have that:

$$\not\Vdash \circlearrowleft \rightarrow \boxplus \circlearrowleft \tag{2.1.8}$$

Nor does this make much sense. It asserts "If the agent's basic beliefs are sound, then all extensions of her basic beliefs are sound too." Soundness is a fragile thing – it is rather easy to think of things to add to a sound set of basic beliefs which break soundness, such as "All logicians are centaurs" and other such demonstrably false nonsense.

Related to (2.1.7), there is another related validity associated with $\circlearrowleft$; namely that if the EvIL agent's assumptions are sound, then anything she concludes from them is true (employing the reading which naturally arises from Theorem 2.1.12). This is expressed as

$$\Vdash \circlearrowleft \rightarrow \Box\varphi \rightarrow \varphi \tag{2.1.9}$$

The formula (2.1.7) expresses that the soundness of one's premises is something *persistent* as the EvIL agent carries on casting doubt on assumptions and discarding them. Another thing that is persistent this way is the EviL agent's imagination:

$$\Vdash \Diamond\varphi \rightarrow \boxminus\Diamond\varphi \tag{2.1.10}$$

One may read (2.1.10) as saying something like "If the EvIL agent can imagine/conceive of something, then no matter what things she casts into doubt, she can still imagine it." One can also express something like the dual of this, namely

$$\Vdash \Box\varphi \rightarrow \boxplus\Box\varphi \tag{2.1.11}$$

We shall read the above as asserting "If the agent *can compose an argument* then she will still be able to compose that argument if she remembers more information and experiences she's had in the world." This should not be surprising – this is yet another expression of the Theorem 2.1.12, the Theorem Theorem, and the fact that the proof theory of EvIL is monotonic. In general, many of the assertions here exhibit interplay between $\boxplus$ and $\Box$, and dually $\boxminus$ and $\Diamond$ – further investigation of these relationships is taken up in §2.2.

For better or for worse, EvIL semantics make true the following assertion: if something is achievable by repeatedly casting assumptions into doubt, then it's achievable by casting assumptions into doubt only once:

$$\Vdash \diamondminus^{+}\varphi \rightarrow \diamondminus\varphi \tag{2.1.12}$$

Here $^{+}$ is taken from the syntax for *regular expressions* commonly used in computer science and UNIX programming to mean "one or more" [Fri06]. Similarly, we have assumed that discarding no assumptions is, in a way, vacuously casting assumptions into doubt. In light of this EvIL also makes true the following:

$$\Vdash \varphi \rightarrow \diamondminus\varphi \tag{2.1.13}$$

Furthermore, it is worth mentioning some harder to understand validities of this system. The first one is that when the agent believes something, they believe it regardless of the process of doubting or embracing their beliefs:

$$\Vdash \Box\varphi \rightarrow \Box\boxminus\varphi \tag{2.1.14}$$

$$\Vdash \Box\varphi \rightarrow \Box\boxplus\varphi \tag{2.1.15}$$

We can observe that this generalizes to multiple agents, as specified in §2.2.

Another more challenging validity is the fact that if some proposition φ holds, then for any restriction of the EvIL agent's beliefs (or dually, any extension), if those beliefs are sound, then φ must be conceivable (i.e., $\Diamond\varphi$ holds). This is expressed as the following two validities:

$$\Vdash \varphi \to \boxminus(\circlearrowleft\to \Diamond\varphi) \tag{2.1.16}$$

$$\Vdash \varphi \to \boxplus(\circlearrowleft\to \Diamond\varphi) \tag{2.1.17}$$

Finally, another a peculiarity of EvIL is that not all of its validities are *schematic*. For instance, there is a kind of *Cartesian dualism* present in the semantics, where the EvIL agent's deliberation on her evidence does not bear on brute matters of fact. For a world pair (a, A), A and a are basically separate - an EvIL agent's mind and the world they live are composed of different substance. This gives rise to the following four validities:

$$\Vdash p \to \boxminus p \tag{2.1.18}$$

$$\Vdash p \to \boxplus p \tag{2.1.19}$$

$$\Vdash \neg p \to \boxminus\neg p \tag{2.1.20}$$

$$\Vdash \neg p \to \boxplus\neg p \tag{2.1.21}$$

Hence, EvIL is not a *normal* logic.

On the other hand, it is by the same assumption of Cartesian dualism that underlies the non-normality that (2.1.7) as is a natural consequence. By accepting non-normality, and the grammar restriction we have imposed on *basic beliefs* to avoid paradoxes, it follows as a consequence that a belief set is sound if and only if all of its subsets are sound. Hence non-normality for EvIL part of the price that must be payed for the philosophically appealing features that the logic has to offer.

In the next section, we turn to a more systematic study of the validities of EvIL. We shall see that this gives rise to an *elimination theorem*.

2.2 EviL Elaborations

Elimination

In section §2.1, we saw some of the structural validities of EviL from a philosophical perspective. That being the case, the manner of presentation followed intuition, which did not follow an orthodox organization. In this section, we shall look at the validities of EvIL in a more systematic presentation. In doing so, we investigate an elimination theorem, which sits at the heart of EvIL.

To start, the following lemma summarizes the structural validities will be studied in the subsequent discussion:

Lemma 2.2.13. *The following validities hold for all* EvIL *models:*

$$
\begin{array}{ll}
\Vdash \boxminus p \leftrightarrow p & \Vdash \boxplus p \leftrightarrow p \\
\Vdash \boxminus \neg p \leftrightarrow \neg p & \Vdash \boxplus \neg p \leftrightarrow \neg p \\
\Vdash \boxminus \Diamond \varphi \leftrightarrow \Diamond \varphi & \Vdash \boxplus \Box \varphi \leftrightarrow \Box \varphi \\
\Vdash \oplus \Diamond \varphi \leftrightarrow \Diamond \varphi & \Vdash \ominus \Box \varphi \leftrightarrow \Box \varphi \\
\Vdash \boxminus \boxminus \varphi \leftrightarrow \boxminus \varphi & \Vdash \boxplus \boxplus \varphi \leftrightarrow \boxplus \varphi \\
\Vdash \boxminus \oplus \varphi \leftrightarrow \oplus \varphi & \Vdash \boxplus \ominus \varphi \leftrightarrow \ominus \varphi \\
\Vdash \boxminus \circlearrowleft \leftrightarrow \circlearrowleft & \Vdash \boxplus \neg \circlearrowleft \leftrightarrow \neg \circlearrowleft
\end{array}
$$

These validities suggest a definite interplay between the modalities of EvIL; they are highly suggestive of a general elimination theorem. To see what arises from Lemma 2.2.13, first observe that EvIL makes true the usual substitution rule:

Lemma 2.2.14. *If* $\Vdash \varphi \leftrightarrow \psi$ *is a validity, then* $\Vdash \chi \leftrightarrow \chi[\varphi/\psi]$ *is a validity for any* $\chi \in \mathcal{L}(\Phi)$.

Next, consider two sublanguages of the main language of EvIL:

Definition 2.2.15. *Define the following fragments:*[1]

$\mathcal{L}_A(\Phi)$:

$$\varphi ::= p \mid \neg p \mid \top \mid \bot \mid \circlearrowleft \mid \varphi \wedge \psi \mid \varphi \vee \psi \mid \Diamond \varphi \mid \boxminus \varphi \mid \oplus \varphi$$

$\mathcal{L}_B(\Phi)$:

$$\varphi ::= \neg p \mid p \mid \bot \mid \top \mid \neg \circlearrowleft \mid \varphi \vee \psi \mid \varphi \wedge \psi \mid \Box \varphi \mid \ominus \varphi \mid \boxplus \varphi$$

Definition 2.2.16. *Define two dualizing operations* $(\cdot)^A : \mathcal{L}_B(\Phi) \to \mathcal{L}_A(\Phi)$ *and* $(\cdot)^B : \mathcal{L}_A(\Phi) \to \mathcal{L}_B(\Phi)$*, using recursion, such that:*

$$
\begin{array}{ll}
\neg p^A := p & p^B := \neg p \\
p^A := \neg p & \neg p^B := p \\
\bot^A := \top & \top^B := \bot \\
\top^A := \bot & \bot^B := \top \\
\neg \circlearrowleft^A := \circlearrowleft & \circlearrowleft^B := \neg \circlearrowleft \\
(\varphi \vee \psi)^A := \varphi^A \wedge \psi^A & (\varphi \wedge \psi)^B := \varphi^B \vee \psi^B \\
(\varphi \wedge \psi)^A := \varphi^A \vee \psi^A & (\varphi \vee \psi)^B := \varphi^B \wedge \psi^B \\
(\Box \psi)^A := \Diamond(\psi^A) & (\Diamond \psi)^B := \Box(\psi^B) \\
(\ominus \psi)^A := \boxminus(\psi^A) & (\boxminus \psi)^B := \ominus(\psi^B) \\
(\boxplus \psi)^A := \oplus(\psi^A) & (\oplus \psi)^B := \boxplus(\psi^B)
\end{array}
$$

With the above definition in hand, it is straightforward to see the following duality theorem:

Theorem 2.2.17 (Duality)**.** *Observe that for all* $\varphi \in \mathcal{L}_A(\Phi)$ *and* $\psi \in \mathcal{L}_B(\Phi)$, $(\varphi^B)^A = \varphi$ *and* $(\psi^A)^B = \psi$. *Moreover, we have the following validities:* $\Vdash \neg(\varphi^B) \leftrightarrow \varphi$ *and* $\Vdash \neg(\psi^A) \leftrightarrow \psi$.

[1] We were inspired to look at the fragment $\mathcal{L}_A(\Phi)$ by thinking about the continuous fragment of μPML [Fon08].

The above duality is convenient, since it can be leveraged to transfer results proven for the fragment $\mathcal{L}_A(\Phi)$ to $\mathcal{L}_B(\Phi)$ and vice versa.

With the above machinery in place, we can observe a natural consequence of the logical equivalences given in Lemma 2.2.13:

Definition 2.2.18. *If $\varphi \in \mathcal{L}_A(\Phi) \cup \mathcal{L}_B(\Phi)$ then let φ^* be the same formula, with all instances of $\boxplus$, $\boxminus$, $\diamondminus$ and $\oplus$ eliminated. That is, $(\cdot)^*$ has the following recursive definition:*

$$\begin{array}{ll} p^* := p & (\neg p)^* := \neg p \\ \top^* := \top & \bot^* := \bot \\ \circlearrowleft^* := \circlearrowleft & (\neg \circlearrowleft)^* := \neg \circlearrowleft \\ (\varphi \vee \psi)^* := (\varphi^*) \vee (\psi^*) & (\varphi \wedge \psi)^* := (\varphi^*) \wedge (\psi^*) \\ (\Box\varphi)^* := \Box(\varphi^*) & (\Diamond\varphi)^* := \Diamond(\varphi^*) \\ (\boxminus\varphi)^* := \varphi^* & (\diamondminus\varphi)^* := \varphi^* \\ (\boxplus\varphi)^* := \varphi^* & (\oplus\varphi)^* := \varphi^* \end{array}$$

Theorem 2.2.19 (EvIL Elimination)**.** *For all $\varphi \in \mathcal{L}_A(\Phi)$ or $\varphi \in \mathcal{L}_B(\Phi)$, we have the following validity:*

$$\Vdash \varphi \leftrightarrow \varphi^*$$

Proof. The proof proceeds in three steps.

Step 1: First, use induction on $\varphi \in \mathcal{L}_A(\Phi)$, and show the following two facts simultaneously:

$$\Vdash \boxminus\varphi \leftrightarrow \varphi \qquad \Vdash \oplus\varphi \leftrightarrow \varphi$$

- Cases p, $\neg p$, $\bot$, $\top$, $\circlearrowleft$: In all of these situations, the result follows directly from the validities illustrated in Lemma 2.2.13.
- Cases $\wedge, \vee$: For $\boxminus$ the connective $\wedge$ is simple, and dually for $\oplus$ for the connective $\vee$. This is because in each case one may simply use distribution, such as can be done here:

$$\begin{aligned} \Vdash \boxminus(\varphi \wedge \psi) &\leftrightarrow \boxminus\varphi \wedge \boxminus\psi \\ &\leftrightarrow \varphi \wedge \psi \end{aligned}$$

  On the other hand, $\vee$ is more interesting for $\boxminus$, and dually $\wedge$ for $\oplus$. Using induction, Lemma 2.2.13, and subsitution, and distribution, we have the line of reasoning:

$$\begin{aligned} \Vdash \boxminus(\varphi \vee \psi) &\leftrightarrow \boxminus(\oplus\varphi \vee \oplus\psi) \\ &\leftrightarrow \boxminus\oplus(\varphi \vee \psi) \\ &\leftrightarrow \oplus(\varphi \vee \psi) \\ &\leftrightarrow \oplus\varphi \vee \oplus\psi \\ &\leftrightarrow \varphi \vee \psi \end{aligned}$$

- Case $\Diamond$: Once again, this follows immediately from the validities of Lemma 2.2.13, namely $\Vdash \boxminus\Diamond\varphi \leftrightarrow \Diamond\varphi$ and $\Vdash \oplus\Diamond\varphi \leftrightarrow \Diamond\varphi$

- Cases $\boxminus, \boxplus$: The final step follows from one more application of Lemma 2.2.13, namely by employing the following four validities

$$\Vdash \boxminus\boxplus\varphi \leftrightarrow \boxplus\varphi \quad \Vdash \boxplus\boxplus\varphi \leftrightarrow \boxplus\varphi$$
$$\Vdash \boxminus \boxminus \varphi \leftrightarrow \boxminus\varphi \quad \Vdash \boxplus \boxminus \varphi \leftrightarrow \boxminus\varphi$$

Step 2: With the above, we can prove for any $\varphi \in \mathcal{L}_A(\Phi)$ that $\Vdash \varphi \leftrightarrow \varphi^*$. Once again, the proof proceeds by induction, the only steps worth noting involve $\boxminus$ and $\boxplus$. In either case, these may be completed using Step 1. For instance, we know that $\Vdash \boxminus\varphi \leftrightarrow \varphi$, hence $\Vdash \boxminus\varphi \leftrightarrow \varphi^*$ by induction.

Step 3: With the result for $\mathcal{L}_A(\Phi)$ in hand, just observe that for $\psi \in \mathcal{L}_B(\Phi)$ we have that $(\psi^A)^* = (\psi^*)^A$. With this, substitution, and duality, we have the following chain of reasoning:

$$\begin{aligned}
\Vdash \psi &\leftrightarrow \neg(\psi^A) \\
&\leftrightarrow \neg((\psi^A)^*) \\
&\leftrightarrow \neg((\psi^*)^A) \\
&\leftrightarrow \neg(\neg(((\psi^*)^A)^B)) \\
&\leftrightarrow \neg\neg\psi^* \\
&\leftrightarrow \psi^*
\end{aligned}$$

QED

Example 2.2.20. *The following validities of* EvIL *are consequences of Theorem 2.2.19:*

$$\Vdash \boxminus\boxplus\Diamond t \vee \boxplus\Box\boxplus\neg t$$

$$\Vdash ((\boxplus \boxplus \Box q \wedge \boxplus\boxplus\Box q) \vee \Box \boxplus \boxplus q) \wedge ((\boxplus\Box\boxplus q \vee \Box\boxplus \boxplus q) \wedge \boxplus\Box \boxplus q) \leftrightarrow \Box q$$

$$\begin{array}{ll}
\Vdash & \boxplus \\
& \boxplus\boxplus\boxplus\boxplus\boxplus \boxplus\boxplus\boxplus\boxplus\boxplus\boxplus\boxplus\boxplus\boxplus\boxplus\boxplus\boxplus\boxplus\boxplus\boxplus\boxplus \\
& \boxplus\boxplus\boxplus\boxplus \boxplus\boxplus\Box\Box\Box\Box\Box\Box\Box\Box \boxplus\boxplus\boxplus\boxplus\boxplus \\
& \boxplus\boxplus \boxplus\boxplus\Box\Box\Box\Box\Box\Box\Box\Box\Box\Box \boxplus\boxplus\boxplus\boxplus \\
& \boxplus \boxplus \Box\Box\Box\Box\boxplus\boxplus\Box\Box\Box\Box\boxplus\boxplus\Box\Box\Box\Box \boxplus\boxplus \\
& \boxplus \boxplus \Box\Box\Box\Box\boxplus\boxplus\Box\Box\Box\Box\boxplus\boxplus\Box\Box\Box\Box \boxplus\boxplus \\
& \boxplus \boxplus \Box\Box\Box\Box\Box\Box\Box\Box\Box\Box\Box\Box\Box\Box\Box\Box \boxplus\boxplus \\
& \boxplus \boxplus \Box\Box\Box\boxplus\Box\Box\Box\Box\Box\Box\boxplus\Box\Box\Box \boxplus\boxplus \\
& \boxplus\boxplus \boxplus\boxplus\Box\Box\boxplus\boxplus\boxplus\boxplus\boxplus\boxplus\boxplus\boxplus\Box\Box \boxplus\boxplus\boxplus\boxplus \\
& \boxplus\boxplus\boxplus \boxplus\boxplus\Box\Box\Box\Box\Box\Box\Box\Box\Box \boxplus\boxplus\boxplus\boxplus \\
& \boxplus\boxplus\boxplus \boxplus\boxplus\boxplus\boxplus\boxplus\boxplus\boxplus\boxplus\boxplus\boxplus\boxplus\boxplus\boxplus\boxplus\boxplus\boxplus\boxplus\boxplus \\
& \boxplus\top
\end{array}$$

One way to read Theorem 2.2.19 is that $\boxminus$ and $\oplus$ are empty modalities on $\mathcal{L}_A(\Phi)$, and dually for $\mathcal{L}_B(\Phi)$ with $\diamond$ and $\boxplus$. Further, note that $\mathcal{L}_0(\Phi) = \mathcal{L}_A(\Phi) \cap \mathcal{L}_B(\Phi)$ (up to translation), which means that all four of $\boxplus$, $\boxminus$ along with their duals $\oplus$ and $\diamond$ vanish on the propositional language. Inspecting the semantics, this is to be expected, since neither $\boxplus$ nor $\boxminus$ interact with propositional truth values.

Finally, it should be mentioned that Theorem 2.2.19 reflects one of the basic themes of EvIL - the interplay between belief, reflected by $\Box$, and imagination, reflected by $\Diamond$. These two phenomena are just two sides of the same coin - furthermore, one could not have more natural opposites. Belief and imagination exemplify two warring forces dwelling within any EvIL agent's heart. Evidently soundness $\circlearrowleft$ is aligned with imagination and unsoundness $\neg \circlearrowleft$ is aligned with belief.

Multiple Agents

In this section we extend the semantics for EvIL from a single agent, as presented in §2.1, to accommodate multiple agents. This is primarily of interest since further results in EvIL, namely completeness, can naturally be abstracted beyond the single agent case.

The following provides the definition of the language of multi-agent EvIL:

Definition 2.2.21. *Define $\mathcal{L}(\Phi, \mathcal{A})$ by the following Backus-Naur grammar:*

$$\varphi ::= p \in \Phi \mid \varphi \to \psi \mid \bot \mid \Box_X \varphi \mid \boxminus_X \varphi \mid \boxplus_X \varphi \mid \circlearrowleft_X$$

Here $X \in \mathcal{A}$, and $\mathcal{A}$ is non-empty.

As in the single agent case, multi-agent EvIL models are sets $\mathfrak{M} \subseteq \wp\Phi \times (\wp\mathcal{L}_0(\Phi))^{\mathcal{A}}$ – that is, $\mathfrak{M}$ is a set of pairs of sets of proposition letters, and indexed sets of propositional formulae.

The semantic entailment relation for multi-agent EvIL is

$$(\Vdash) : \wp(\wp\Phi \times (\wp\mathcal{L}_0(\Phi))^{\mathcal{A}}) \times \wp\Phi \times (\wp\mathcal{L}_0(\Phi))^{\mathcal{A}} \times \mathcal{L}(\Phi, \mathcal{A}) \to \mathsf{bool}.$$

The input/output behavior of $(\Vdash)$ is just as it was defined before in §2.1, the only difference in this setting is that instead of taking a pair as an input, where the second element is a set, it takes an indexed set.

We now provide a formal definition of the semantics for the multi-agent $(\Vdash)$: [2]

Definition 2.2.22.

$$\begin{aligned}
\mathfrak{M}, (a, A) \Vdash p &\iff p \in a \\
\mathfrak{M}, (a, A) \Vdash \varphi \to \psi &\iff \mathfrak{M}, (a, A) \Vdash \varphi \text{ implies } \mathfrak{M}, (a, A) \Vdash \psi \\
\mathfrak{M}, (a, A) \Vdash \bot &\iff False \\
\mathfrak{M}, (a, A) \Vdash \Box_X \varphi &\iff \forall (b, B) \in \mathfrak{M}. (\forall \psi \in A_X . b \models \psi) \text{ implies } \mathfrak{M}, (b, B) \Vdash \varphi \\
\mathfrak{M}, (a, A) \Vdash \boxminus_X \varphi &\iff \forall (b, B) \in \mathfrak{M}. a = b \text{ and } B_X \subseteq A_X \text{ implies } \mathfrak{M}, (b, B) \Vdash \varphi \\
\mathfrak{M}, (a, A) \Vdash \boxplus_X \varphi &\iff \forall (b, B) \in \mathfrak{M}. a = b \text{ and } B_X \supseteq A_X \text{ implies } \mathfrak{M}, (b, B) \Vdash \varphi \\
\mathfrak{M}, (a, A) \Vdash \circlearrowleft_X &\iff \forall \psi \in A_X . a \models \psi
\end{aligned}$$

[2] Where $X \in \mathcal{A}$, we shall use A_X to denote $A(X)$ provided that $A : \mathcal{A} \to \wp\mathcal{L}_0(\Phi)$

Just as in §2.1, Lemma 2.1.10 and Theorem 2.1.12 can be seen to obtain for the new generalized semantics. Furthermore, all of the validities mentioned in §2.1 and §2.2 hold, along with Theorem 2.2.19, where $\Box$, $\Diamond$, $\boxminus$,$\boxplus$, $\diamondminus$, $\diamondplus$ and $\circlearrowleft$ are all replaced with $\Box_X$, $\Diamond_X$, $\boxminus_X$, $\boxplus_X$, $\diamondminus_X$, $\diamondplus_X$ and $\circlearrowleft_X$ respectively, for any fixed $X \in \mathcal{A}$.

Finally, there are two novel validities that arise in these semantics:

$$\Vdash \Box_X \varphi \rightarrow \Box_X \boxminus_Y \varphi$$
$$\Vdash \Box_X \varphi \rightarrow \Box_X \boxplus_Y \varphi$$

This is just to say, that the EvIL agent's deliberative process is opaque to other's beliefs, just as in the single agent case. This was expressed by (2.1.14) and (2.1.15) in §2.1. The agent cannot read anyone else's mind, nor anyone else hers.

Using the multi-agent semantics we have developed here, the proof theory for EvIL that shall be presented in §3 can now be given in higher generality.

Kripke Structures & Failure of Compactness

The language of EvIL is evidently modal, and in previous sections the semantics have largely suggested that there are clear connections to conventional Kripke semantics. In this section, we will demonstrate that every EvIL model corresponds to some highly structured Kripke model, with a minor modification on the standard definition. However, it will turn out that this correspondence is one way - the class of Kripke models for which EvIL is strongly complete do not, in general, possess corresponding EvIL models.

In order to understand EvIL models as Kripke models, we return to the visualization technique for EvIL models we introduced in §1.6. There we suggested that EvIL models can be thought of as *posets*, with the partial ordering corresponding to set containment. Here we make the added requirement that the worlds remain in this ordering. In addition, doxastic accessibility will be visualized with arrows. We can see examples of this visualization technique below in Figs. 2.1(a) and 2.1(b).

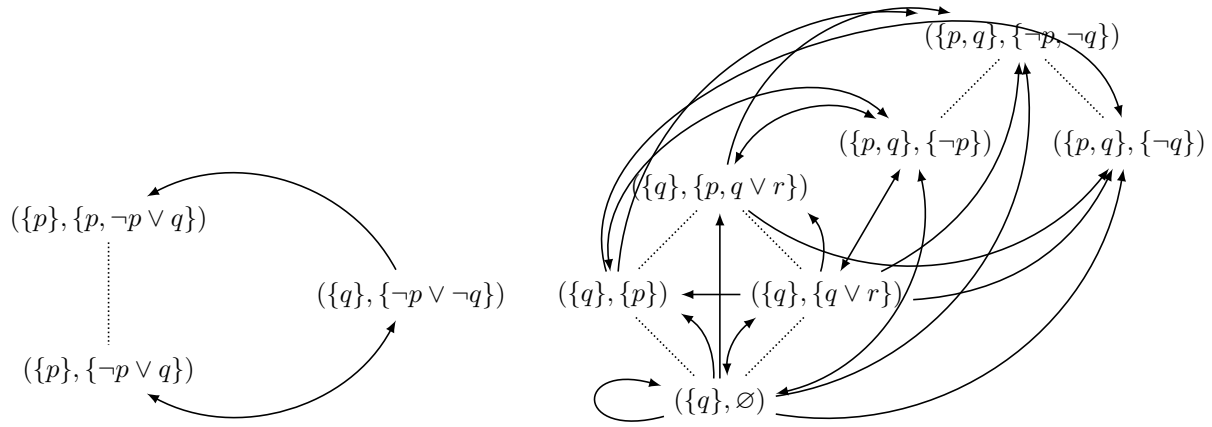
In the interest of being completely explicit, these figures should be read as follows:

- if one point (a, A) is above another point (b, B) and connected by a densely dotted line ⋮, this means that $a = b$ and $B \subset A$.
- if one point (a, A) is connected to another point (b, B) by a line with an arrow ↗, this means that $\mathfrak{M}, (b, B) \Vdash A$

In all of these depictions, the implicit relational structure of EvIL models is given visual expression. So it seems only natural that this graphically perceived structure could also find formal expression.

Following the modified semantics provided in §2.2, the developments this section will assume multiple agents.

Definition 2.2.23. *Let Φ be a set of letters and let $\mathcal{A}$ be a set of agents. A **Kripke structure** is*



(a) A fairly simple example

(b) A more complex example

Figure 2.1: EvIL model visualizations

a state transition system $\mathbb{M} = \langle W^{\mathbb{M}}, R^{\mathbb{M}}, \sqsubseteq^{\mathbb{M}}, \sqsupseteq^{\mathbb{M}}, V^{\mathbb{M}}, P \rangle$ *where*[3]*:*

- W *is a set of worlds*
- $R : \mathcal{A} \to \wp(W \times W)$*,* $\sqsubseteq\colon \mathcal{A} \to \wp(W \times W)$*, and* $\sqsupseteq\colon \mathcal{A} \to \wp(W \times W)$ *are* $\mathcal{A}$*-indexed sets of relations*[4]
- $V : \Phi \to \wp(W)$ *is a predicate letter valuation*
- $P : \mathcal{A} \to \wp(W)$ *are sets of worlds indexed by agents*

Let $\mathcal{K}_{\Phi,\mathcal{A},I}$ *denote the class of Kripke structures for letters* Φ*, agents* $\mathcal{A}$*, and where* $W \subseteq I$*.*

The Kripke semantics given by $(\Vdash\!\vdash) : \mathcal{K}_{\Phi,\mathcal{A},I} \to I \to \mathcal{L}(\Phi, \mathcal{A}) \to \mathsf{bool}$ for these models are defined recursively as usual, granting the exceptional behavior of P.

Definition 2.2.24. *Let* $\mathbb{M}$ *be a Kripke structure:*

$$
\begin{aligned}
\mathbb{M}, w \Vdash\!\vdash p &\iff w \in V(p) \\
\mathbb{M}, w \Vdash\!\vdash \varphi \to \psi &\iff \mathbb{M}, w \Vdash\!\vdash \varphi \textit{ implies } \mathbb{M}, w \Vdash\!\vdash \psi \\
\mathbb{M}, w \Vdash\!\vdash \bot &\iff \textit{False} \\
\mathbb{M}, w \Vdash\!\vdash \Box_X \varphi &\iff \forall v \in W. w R_X v \textit{ implies } \mathbb{M}, v \Vdash\!\vdash \varphi \\
\mathbb{M}, w \Vdash\!\vdash \boxminus_X \varphi &\iff \forall v \in W. w \sqsupseteq_X v \textit{ implies } \mathbb{M}, v \Vdash\!\vdash \varphi \\
\mathbb{M}, w \Vdash\!\vdash \boxplus_X \varphi &\iff \forall v \in W. w \sqsubseteq_X v \textit{ implies } \mathbb{M}, v \Vdash\!\vdash \varphi \\
\mathbb{M}, w \Vdash\!\vdash \circlearrowleft_X &\iff w \in P_X
\end{aligned}
$$

[3]Where the context is clear, we shall drop the superscript $\mathbb{M}$.

[4]We shall abbreviate $R(X)$, $\sqsubseteq(X)$, $\sqsupseteq(X)$, $P(X)$ as R_X, $\sqsubseteq_X$, $\sqsupseteq_X$ and P_X respectively

Kripke structures can be observed to typically have a lot less structure than EvIL models. On the other hand, EvIL models can be understood as Kripke structures in disguise. To illustrate this, observe the following lemma:

Definition 2.2.25 ($\mho^{\mathfrak{M}}$ Translation)**.** *Let $\mathfrak{M}$ be an* EvIL *model. Define $\mho^{\mathfrak{M}} := \langle \mathfrak{M}, R^{\mathfrak{M}}, \sqsubseteq^{\mathfrak{M}}, \sqsupseteq^{\mathfrak{M}}, V^{\mathfrak{M}}, P^{\mathfrak{M}} \rangle$, where*

- $(a, A) R_X^{\mathfrak{M}} (b, B) \iff \forall \psi \in A_X . b \models \psi$
- $(a, A) \sqsubseteq_X^{\mathfrak{M}} (b, B) \iff a = b$ *and* $A_X \subseteq B_X$
- $(a, A) \sqsupseteq_X^{\mathfrak{M}} (b, B) \iff a = b$ *and* $A_X \supseteq B_X$
- $(a, A) \in P^{\mathfrak{M}}(X) \iff \forall \psi \in A_X . a \models \psi$
- $V(p) := \{(a, A) \in \mathfrak{M} \mid \mathfrak{M}, (a, A) \Vdash p\}$

Lemma 2.2.26. *For all $\mathfrak{M}$ and all $(a, A) \in \mathfrak{M}$,*

$$\mathfrak{M}, (a, A) \Vdash \varphi \text{ if and only if } \mho^{\mathfrak{M}}, (a, A) \Vdash \varphi.$$

Proof. This follows from a straightforward induction on φ. QED

The following summarizes the structural properties of EvIL models, when transformed into Kripke structures:

Proposition 2.2.27. *For any* EvIL *model $\mathfrak{M}$, $\mho^{\mathfrak{M}}$ has the following properties, for all agents $\{X, Y\} \subseteq \mathcal{A}$: :*

(I) $\sqsubseteq_X^{\mathfrak{M}}$ *is reflexive*

(II) $\sqsubseteq_X^{\mathfrak{M}}$ *is transitive*

(III) $w \sqsubseteq_X^{\mathfrak{M}} v$ *if and only if* $v \sqsupseteq_X^{\mathfrak{M}} w$

(IV) *If* $w \sqsubseteq_X^{\mathfrak{M}} v$ *then* $(w \in V(p)$ *if and only if* $v \in V(p))$

(V) $(R_X^{\mathfrak{M}} \circ \sqsubseteq_X^{\mathfrak{M}}) \subseteq R_X^{\mathfrak{M}}$

(VI) $(\sqsubseteq_Y^{\mathfrak{M}} \circ R_X^{\mathfrak{M}}) \subseteq R_X^{\mathfrak{M}}$ *and*
$(\sqsupseteq_Y^{\mathfrak{M}} \circ R_X^{\mathfrak{M}}) \subseteq R_X^{\mathfrak{M}}$

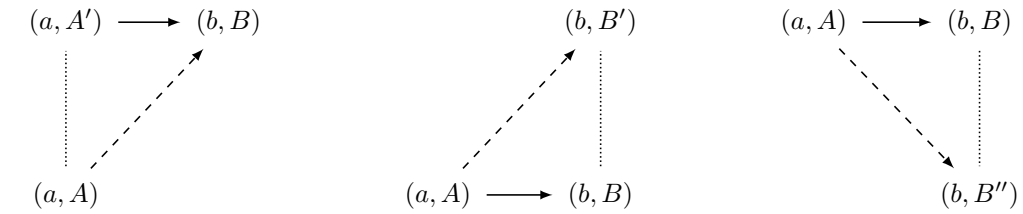
(VII) $w \in P^{\mathfrak{M}}(X)$ *if and only if* $w R_X^{\mathfrak{M}} w$

The situation in (V) can be visualized in a commutative diagram depicted in 2.2(a), while (VI) can be split into Figs. 2.2(b) and 2.2(c).

Proof. Everything except (V) and (VI) follows directly immediately from the definitions:

(V) We must show $(R_X^{\mathfrak{M}} \circ \sqsubseteq_X^{\mathfrak{M}}) \subseteq R_X^{\mathfrak{M}}$.

Assume that $(a, A) \sqsubseteq_X^{\mathfrak{M}} (b, B) R_X^{\mathfrak{M}} (c, C)$, then evidently $\forall \psi \in B_X . c \models \psi$. Since $A_X \subseteq B_X$ then evidently $\forall \psi \in A_X . c \models \psi$. This means that $(a, A) R_X^{\mathfrak{M}} (c, C)$, which suffices the claim.



(a) – (V) –
$(R_X^{\mathfrak{M}} \circ \sqsubseteq_X^{\mathfrak{M}}) \subseteq R_X^{\mathfrak{M}}$

(b) – (VI) –
$(\sqsubseteq_Y^{\mathfrak{M}} \circ R_X^{\mathfrak{M}}) \subseteq R_X^{\mathfrak{M}}$

(c) – (VI) –
$(\sqsupseteq_Y^{\mathfrak{M}} \circ R_X^{\mathfrak{M}}) \subseteq R_X^{\mathfrak{M}}$

Figure 2.2: Visualizations of the relationships in Proposition 2.2.27

(*VI*) We must show:

$$(\sqsubseteq_Y^{\mathfrak{M}} \circ R_X^{\mathfrak{M}}) \subseteq R_X^{\mathfrak{M}}$$
$$\&$$
$$(\sqsupseteq_Y^{\mathfrak{M}} \circ R_X^{\mathfrak{M}}) \subseteq R_X^{\mathfrak{M}}$$

So assume either of the following:

$$(a, A) R_X^{\mathfrak{M}} (b, B) \sqsubseteq_X^{\mathfrak{M}} (c, C)$$
$$\text{or}$$
$$(a, A) R_X^{\mathfrak{M}} (b, B) \sqsupseteq_X^{\mathfrak{M}} (c, C)$$

In either case we have that $\forall \psi \in A_X.b \models \psi$, and moreover $b = c$. Hence $\forall \psi \in A_X.c \models \psi$, which means that $(a, A) R_X^{\mathfrak{M}} (c, C)$, which was what was to be shown.

QED

Definition 2.2.28. *A Kripke structure is called* EvIL *if it makes true the above properties* (I) *through* (VII).

The Kripke semantics provide proper intuition behind EvIL models. We think of the defined relations given as follows:

- If $x R_X^{\mathfrak{M}} y$, then at world x the agent X can imagine y is true, since y is compatible with what the agent believes
- If $x \sqsubseteq_X^{\mathfrak{M}} y$, then agent X's assumptions at world x (or the experiences they are taking under consideration) are contained in her evidence at y

Given this perspective, the proof of (V) can be understood in the following way - if the agent assumes fewer things, more things are imaginable, since it is easier for a world to be incompatible with an agent's evidence.

Finally, while Prop. 2.2.27 presents itself as a sort of representation lemma, the relationship between EvIL semantics and Kripke semantics is not reciprocal. Proposition 2.2.30 shows that not every

Kripke model can be represented as an EvIL model, by presenting an elementary example of this failure of representation. It turns on the following observation:

Lemma 2.2.29. *For a given* EvIL *model* $\mathfrak{M}$*, for any* $\{(a, A), (b, B), (c, C)\} \subseteq \mathfrak{M}$*, if* $a = b$ *then* $a \models C$ *if and only if* $b \models C$*.*

Proof. Recall that the semantics for $\models$, as defined in Definition 2.1.6 in §2.1 are the usual semantics for classical propositional logic. Remembering this, the above is an elementary result in basic logic. QED

Proposition 2.2.30 (Failure of Representation)**.** *Not every* EvIL *Kripke structure has a representative* EvIL *model.*

Proof. Consider a single agent EvIL Kripke structure $\mathbb{M} := \langle W, R, \sqsubseteq, \sqsupseteq, V, P\rangle$ where

$$W := \{w, v\} \qquad \sqsubseteq := \sqsupseteq := \{(w, w), (v, v)\}$$
$$R := \{(w, v)\} \qquad V(p) := \varnothing \text{ for all } p \in \Phi$$
$$P := \varnothing$$

This structure is depicted in Fig. 2.3. We shall show that $\mathbb{M}$ is not represented by any EvIL model.

Observe that $\mathbb{M}$ makes true the following:

$$\mathbb{M}, w \Vdash \Diamond\top \tag{2.2.22}$$
$$\mathbb{M}, w \Vdash \Box\neg p \text{ for all } p \in \Phi \tag{2.2.23}$$
$$\mathbb{M}, w \Vdash \neg p \text{ for all } p \in \Phi \tag{2.2.24}$$
$$\mathbb{M}, w \Vdash \neg\Diamond\Diamond\top \tag{2.2.25}$$

Armed with these observations, we can assert that it is impossible for there to be an EvIL structure $\mathfrak{M}$ with a world (a, A) such that $\mathbb{M}, w \Vdash \varphi$ if and only if $\mathfrak{M}, (a, A) \Vdash \varphi$.

For suppose there were, then we could deduce the following facts, using the observations above:

(*1*) From (2.2.22), there must be some pair $(b, B) \in \mathfrak{M}$ such that $b \models A$. Hence, A must be *consistent*.

(*2*) From (2.2.23), we know that for the b mentioned above it must be that $b = \varnothing$. This is a direct consequence of Lemma 2.1.10, the Truthiness Lemma.

(*3*) From (2.2.24), evidently $a = \varnothing$

(*4*) From (2.2.25), it must be that $a \not\models A$. Otherwise by the semantics of EvIL as defined in §2.1 we would have $\mathfrak{M}, (a, A) \Vdash \Diamond\Diamond\top$

Since $a = b = \varnothing$ and $b \models A$ then by Lemma 2.2.29 it must be that $a \models A$. But this clearly is absurd! $\lightning$ QED

The above one way correspondence is admittedly inconvenient - it means that while EvIL only enjoys some features from traditional epistemic logic, it is denied others. Despite this, EvIL enjoys

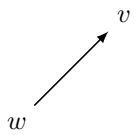


Figure 2.3: A Kripke structure $\mathbb{M}$ with no EvиL representation

most of the benefits of basic modal logic. Indeed, we shall see in §3.3 that EvиL is strongly complete for EvиL Kripke models.

Perhaps the most important formal feature that EvиL semantics lacks in comparison to abstract Kripke semantics is that, as a consequence of the observations made in Proposition 2.2.30, EvиL is not compact.

Theorem 2.2.31 (Failure of Compactness)**.** *If the set of proposition letters* Φ *is infinite, then* EvиL *is not compact for* EvиL *semantics.*

Proof. We shall prove this result for the single agent case (the multiple agent case is an obvious generalization). Consider the function $\tau : \Phi \to \mathcal{L}(\Phi)$, defined as follows:

$$\tau(p) := (\Diamond\top) \wedge (\Box\neg p) \wedge (\neg p) \wedge (\neg\Diamond\Diamond\top)$$

We shall see that $\tau[\Phi]$ is finitely satisfiable, but not in its entirety.

Clearly not all of $\tau[\Phi]$ is satisfiable in EvиL semantics, by the arguments presented in the proof of Proposition 2.2.30.

Now consider some finite subset of $S \subseteq_\omega \tau[\Phi]$. We shall construct a model that makes S true. Since τ is injective, we know there is some $\Psi \subseteq \Phi$ such that $S = \tau[\Psi]$. Since Φ is infinite, we know there is some $\rho \in \Phi \backslash \Psi$. Now consider a model $\mathfrak{M} = \{(\{\rho\}, \{\neg\rho\}), (\varnothing, \{\bot\})\}$. This is depicted in Fig. 2.4. It is straightforward to verify that $\mathfrak{M}, (\{\rho\}, \{\neg\rho\}) \Vdash \tau[\Psi]$, so $\mathfrak{M}$ is a suitable witness. QED

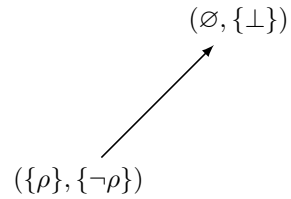


Figure 2.4: A model $\mathfrak{M}$ where $\mathfrak{M}, (\{\rho\}, \{\neg\rho\}) \Vdash \tau[\Psi]$ for $\Psi \subseteq_\omega \Phi$ and $\rho \notin \Psi$

A consequence of the failure of compactness, while strong completeness can be obtained for EvиL using Kripke semantics, to achieve completeness for EvиL semantics a finitary proof must be carried out. Recall that this was exactly the strategy used in our original sketch of EvиL that we gave in Proposition 1.3.2 in §1.3.

The next section is devoted to studying completeness for EvиL.

Chapter 3

EviL Completeness

In this section we provide a complete axiomatization of multi-agent EviL. In addition to axiomatics, we shall also look at subsystems and supersystems of EviL, and provide complexity bounds on EviL decision procedures.

We have organized this section in the following manner:

§3.1 We first present a sound axiom system for EviL.

§3.2 Next, we give a definition of the class of *partly* EviL Kripke structures.

We then reveal that EviL is sound and strongly complete for the class of partly EviL structures. Completeness rests on the observation that the axioms of EviL are all in the *Sahlqvist fragment*, or have obvious meanings in terms of the traditional canonical model construction for Modal Logic. This abstract completeness for EviL can be understood as an elementary application of van Benthem's *correspondence theory* for modal logic.

§3.3 In this section we recall the definition of an EviL *Kripke structure*, as we gave in Definition 2.2.28 from §2.2, and show that every *partly* EviL Kripke structure may be "completed" by constructing a bisimilar EviL structure.

This has, as a consequence, that EviL is complete for EviL Kripke structures.

§3.4 In this section, we discuss why the abstract completeness proof developed in the previous sections, while important, is not adequate in light of the developments in §2.1 and the intuitions we saw in that section. We shall sketch what further needs to be shown to give the desired completeness theorem for EviL.

§3.5 In this section we show that EviL has a small model property for *partly* EviL Kripke structure. This is accomplished by constructing a Kripke structure consisting of finite maximally consistent sets in the manner of the Fischer-Ladner closure style completeness proof of PDL [BRV01, chapter 4, pgs. 241–248]. We also discuss how our results give rise to complexity observations for the EviL decision problem.

§3.6 In this section, we introduce the concept of an *island*, which are special equivalence classes for EvIL structures. We shall prove several properties, which take the form of various irreducibilities.

§3.7 Following the proof of Proposition 1.3.2 in §1.3, we shall show that every finite EvIL Kripke structure is *(almost)-homomorphic* to another, EvIL structure we shall call $\bigstar$, provided that we have an infinite number of letters in Φ. We shall show that there is a map ϑ of worlds in $\mathbb{M}$ to worlds in $\bigstar$ that preserves formulae in a language $\mathcal{L}(\Psi, \mathcal{A})$ where $\Psi \subseteq_\omega \Phi$.

The construction of $\bigstar$ shall make use of the island structures introduced in the previous section, and here we will also introduce the concept of *names* and *surnames*.

We have, as a consequence of the above, we shall be able to show that EvIL is weakly complete for EvIL structures.

§3.8 In this section, we once again pause to take stock of the results we have established in previous sections. Here we discuss the relationship between the abstract completeness we have previously established and our concrete completeness.

§3.9 We next introduce two subsystems of EvIL, corresponding to the $\boxplus$ and $\boxminus$ only fragments. We briefly go over the completeness theorems and the small model property for these systems, as well as decidability results. In each case, a special bisimulation theorem is in order to achieve completeness.

§3.10 We provide an extension to EvIL and its subsystems, introducing the universal modality U. We sketch completeness and the finite model property, and mention how translation may be extended to this system.

§3.11 In this section we give a lattice of EvIL systems, and discuss the known complexity bounds of various levels of this lattice.

3.1 Axioms of EviL

In this section, we shall present the axiomatics for multi-agent EvIL. The axioms of EvIL are presented in Table 3.1, which comprises a Hilbert systems for $\vdash_{\text{EvIL}}$[1]. In addition to giving each axiom, we provide a philosophical reading of what each axiom says. As remarked in §2.1, EvIL is not *normal*, that is it is not closed under variable substitution.

Needless to say, these axioms shall be the focus of study for all of our future investigations in §3.

As a final remark, we should mention that EvIL trivially makes true the *deduction* theorem. This is because we follow the convention set forth in [BRV01, Definition 4.4, pg. 192], and make the subsequent stipulation:

Definition 3.1.32.

$$\Gamma \vdash \varphi \iff \exists \Delta \subseteq_\omega \Gamma. \vdash \bigwedge \Delta \to \varphi$$

[1]By abuse of notation, we shall omit subscripts where they are thought to not be ambiguous

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

Table 3.1: A Hilbert style axiom system for EvIL

While trivial, the deduction theorem plays a key role in the proof of the Theorem Theorem, Theorem 2.1.12 from §2.1. Specifically, it is this definition that justifies equation 2.1.1.

3.2 Partly EviL Kripke Structures & Strong Completeness

In this section, we define what it means for a model to be *partly* EviL. We should note that partly EviL models are exactly the same as EviL models as defined in Definition 2.2.28 from §2.2, only property (VII) has been weakened to $(VIII)'$ and $(V)'$.

Definition 3.2.33 (Partly EviL)**.** *A Kripke structure* $\mathbb{M} = \langle W, R, \sqsubseteq, \sqsupseteq, V, P\rangle$ *is called* ***partly EviL*** *whenever it makes true the following properties, for all agents* $\{X, Y\} \subseteq \mathcal{A}$*:*

$(I)'$ $\sqsubseteq_X$ *is reflexive*

$(II)'$ $\sqsubseteq_X$ *is transitive*

$(III)'$ $w \sqsubseteq_X v$ *if and only if* $v \sqsupseteq_X w$

$(IV)'$ *If* $w \sqsubseteq_X v$ *then* $(w \in V(p)$ *if and only if* $v \in V(p))$

$(V)'$ *If* $w \in P_X$ *and* $w \sqsupseteq_X v$ *then* $v \in P_X$

$(VI)'$ $(R_X \circ \sqsubseteq_X) \subseteq R_X$

$(VII)'$ $(\sqsubseteq_Y \circ R_X) \subseteq R_X$ *and*
$(\sqsupseteq_Y \circ R_X) \subseteq R_X$

$(VIII)'$ *If* $w \in P_X$ *then* wR_Xw

Note that it is elementary that every EviL structure is partly EviL as well.

These properties are exactly the properties enforced by the axioms of EviL given in Table 3.1 in §3.1. We can see this by observing the following theorem:

Definition 3.2.34. *We shall write*

$$\Gamma \Vdash_{\text{pEviL}} \varphi$$

to mean that for all partly EviL *Kripke structures* $\mathbb{M} = \langle W, R, \sqsubseteq, \sqsupseteq, V, P\rangle$*, for all worlds* $w \in W$ *if* $\mathbb{M}, w \Vdash \Gamma$ *then* $\mathbb{M}, w \Vdash \varphi$*.*

Theorem 3.2.35 (Partly EviL Strong Soundness and Completeness)**.**

$$\Gamma \vdash_{\text{EviL}} \varphi \text{ if and only if } \Gamma \Vdash_{\text{pEviL}} \varphi$$

Proof. The left to right direction, *soundness*, is trivial (one should use induction). So we shall focus on the right to left direction; to do this we shall consider the contrapositive. We shall make heavy use of *correspondence* theory, namely the *Sahlqvist Correspondence Theorem* [BRV01, Theorem 4.42, pg. 212]. We note that the axioms (3), (4), (7), (8), (9), (10), (11), (12) and (13) are all

Sahlqvist formulae. When we say that a particular fact corresponds to an axiom, we mean that from the Sahlqvist correspondence theorem and first order logic one may be employed to show the fact in question.

So assume $\Gamma \nvdash \varphi$, we must show that $\Gamma \nVdash \varphi$. To see this, we carry out the canonical model construction as described in [BRV01, chapter 4, pgs. 198–422][2]. Let $\mathcal{E}$ be the set of maximally consistent sets of formulae for EvIL. Define the *canonical model*

$$\mathscr{E} := \langle \mathcal{E}, R, \sqsubseteq, \sqsupseteq, V, P \rangle$$

where, for all $\{w, v\} \subseteq \mathcal{E}$:

- wR_Xv if and only if $\{\varphi \mid \Box_X\varphi \in w\} \subseteq v$
- $w \sqsubseteq_X v$ if and only if $\{\varphi \mid \boxplus_X \varphi \in w\} \subseteq v$
- $w \sqsupseteq_X v$ if and only if $\{\varphi \mid \boxminus_X \varphi \in w\} \subseteq v$
- $V(p) := \{w \mid p \in w\}$
- $P_X := \{w \mid \circlearrowleft_X \in w\}$

We know from the *Lindenbaum Lemma* that Γ may be extended to some maximally consistent γ such that $\Gamma \subseteq \gamma$, $\gamma \in \mathcal{E}$ and $\varphi \notin \gamma$ [BRV01, Lemma 4.17, pg. 199]. By the *Truth Lemma* we may establish $\mathscr{E}, \gamma \nVdash \varphi$ and $\mathscr{E}, \gamma \Vdash \Gamma$ [BRV01, Lemma 4.21, pgs. 201]. So it suffices to establish that $\mathscr{E}$ is partly EvIL, by establishing that it satisfies the properties given in Definition 3.2.33.

(*I*)′ "$\sqsubseteq_X$ is reflexive" corresponds to axiom (3).

(*II*)′ "$\sqsubseteq_X$ is transitive" corresponds to axiom (4).

(*III*)′ "$\sqsubseteq_X$ is the reverse $\sqsupseteq_X$" corresponds to axioms (12) and (13).

(*IV*)′ Assume $w \sqsubseteq_X v$, we shall show that

$$w \in V(p) \text{ if and only if } v \in V(p)$$

Now assume that $w \in V(p)$, then $\mathscr{E}, w \Vdash p$. By axiom (6) and the Truth Lemma we have that $\boxplus_X p \in w$, whence $p \in v$ by definition. The other direction is similar, however it uses axiom (5) instead.

(*VI*)′ The assertion

$$(R_X \circ \sqsubseteq_X) \subseteq R_X$$

corresponds to axiom (7) (noting that one should reason given $(III)'$).

[2]It is important to note that the results obtained in [BRV01] are technically for any *normal* modal logic, but they may be generalized to non-normal logics such as the EvIL logic under consideration.

(VII)′ Given $(III)'$, the fact that

$$(\sqsubseteq_Y \circ R_X) \subseteq R_X$$
$$\&$$
$$(\sqsupseteq_Y \circ R_X) \subseteq R_X$$

corresponds to axioms (8) and (9).

(V)′ "If $w \in P_X$ and $w \sqsupseteq_X v$ then $v \in P_X$" corresponds to axiom (11).

(VIII)′ "If $w \in P(X)$ then wR_Xw" corresponds to axiom (10).

QED

3.3 Bisimulation & EviL Strong Completeness

In this section, we show that for every partly EviL Kripke structure, we may "complete" it by constructing a bisimilar EviL structure; this amounts to enforcing the EviL property (VII), namely that a world is reflexive for R_X if and only if it models $\circlearrowleft_X$. This will allow us to establish that EviL is sound and strongly complete for EviL Kripke models.

We shall first review the definition of *bisimulation*, which we shall have to modify somewhat given our modified definition of Kripke structures. We follow [BRV01, Definition 2.16, pg. 64] in our presentation:

Definition 3.3.36. *Let $\mathbb{M} = \langle W, R, \sqsubseteq, \sqsupseteq, V, P\rangle$ and $\mathbb{M}' = \langle W', R', \sqsubseteq', \sqsupseteq', V', P'\rangle$. A non-empty binary relation $Z \subseteq W \times W'$ is a called a **bisimulation between** $\mathbb{M}$ **and** $\mathbb{M}'$ (denoted $Z : \mathbb{M} \leftrightarrows \mathbb{M}'$) if the following are satisfied:*

(i) If wZw' then w and w' satisfy the same proposition letters, along with the special letters P_X.

*(ii) **Forth** – If wZw' and $w \rightsquigarrow v$, then there exists a $v' \in W'$ such that vZv' and $w' \rightsquigarrow' v'$.*

*(iii) **Back** – If wZw' then $w' \rightsquigarrow' v'$, then there exists a $v \in W$ such that vZv' and $w \rightsquigarrow v$.*

... where $\rightsquigarrow$ is any of R_X, $\sqsubseteq_X$, or $\sqsupseteq_X$, where X is any agent in the class of agents $\mathcal{A}$.

We now recall one of the most crucial theorems in all of modal logic:

Theorem 3.3.37 (The Fundamental Theorem of Bisimulations)**.** *If $Z : \mathbb{M} \leftrightarrows \mathbb{M}'$ and wZw', then for all formulae φ we have that*

$$\mathbb{M}, w \Vdash \varphi \text{ if and only if } \mathbb{M}', w' \Vdash \varphi$$

Proof. This is Theorem 2.20 in [BRV01, pg. 67] QED

We now introduce a Backus-Naur form grammar for the Either type constructor. This will give us precise notation for manipulating the disjoint union of a Kripke structure $\mathbb{M}$ with itself.

Definition 3.3.38.

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

We now make use of this grammar to express an operation for making bisimilar models:

Definition 3.3.39 (Bisimulator)**.** *Let $\mathbb{M}$ be a Kripke model, then define a new Kripke model:*

$$\text{ʘ}^{\mathbb{M}} := \langle W^{\text{ʘ}}, R^{\text{ʘ}}, \sqsubseteq^{\text{ʘ}}, \sqsupseteq^{\text{ʘ}}, V^{\text{ʘ}}, P^{\text{ʘ}} \rangle$$

where:

$$\begin{array}{rcl}
W^{\text{ʘ}} & := & \{w_l, w_r \mid w \in W^{\mathbb{M}}\} \\
V^{\text{ʘ}}(p) & := & \{w_l, w_r \mid w \in V^{\mathbb{M}}(p)\} \\
P_X^{\text{ʘ}} & := & \{w_l, w_r \mid w \in P_X^{\mathbb{M}}\} \\
R_X^{\text{ʘ}} & := & \{(w_l, v_r), (w_r, v_l) \mid wR_X^{\mathbb{M}}v\} \cup \\
 & & \{(w_l, v_l), (w_r, v_r) \mid wR_X^{\mathbb{M}}v \ \&\ w \in P_X^{\mathbb{M}}\} \\
\sqsupseteq_X^{\text{ʘ}} & := & \{(w_l, v_l), (w_r, v_r) \mid w \sqsupseteq_X^{\mathbb{M}} v\} \\
\sqsubseteq_X^{\text{ʘ}} & := & \{(w_l, v_l), (w_r, v_r) \mid w \sqsubseteq_X^{\mathbb{M}} v\}
\end{array}$$

It is instructive to review how ʘ operates on Kripke models it takes as input. One idea is that ʘ causes every world w in $\mathbb{M}$ to undergo *mitosis*, and split into two identical copies named w_l and w_r. These copies obey three rules:

(*1*) The *left copy* of a world w, denoted w_l, can see the *right copy* of a world v, denoted v_r, provided that $wR^{\mathbb{M}}v$ originally. This is similarly true for right copies, only reflected.

(*2*) If $\mathbb{M}, w \Vdash \circlearrowleft_X$, then the copies w_l and w_r of w can see both v_l and v_r provided that $wR^{\mathbb{M}}v$ to begin with.

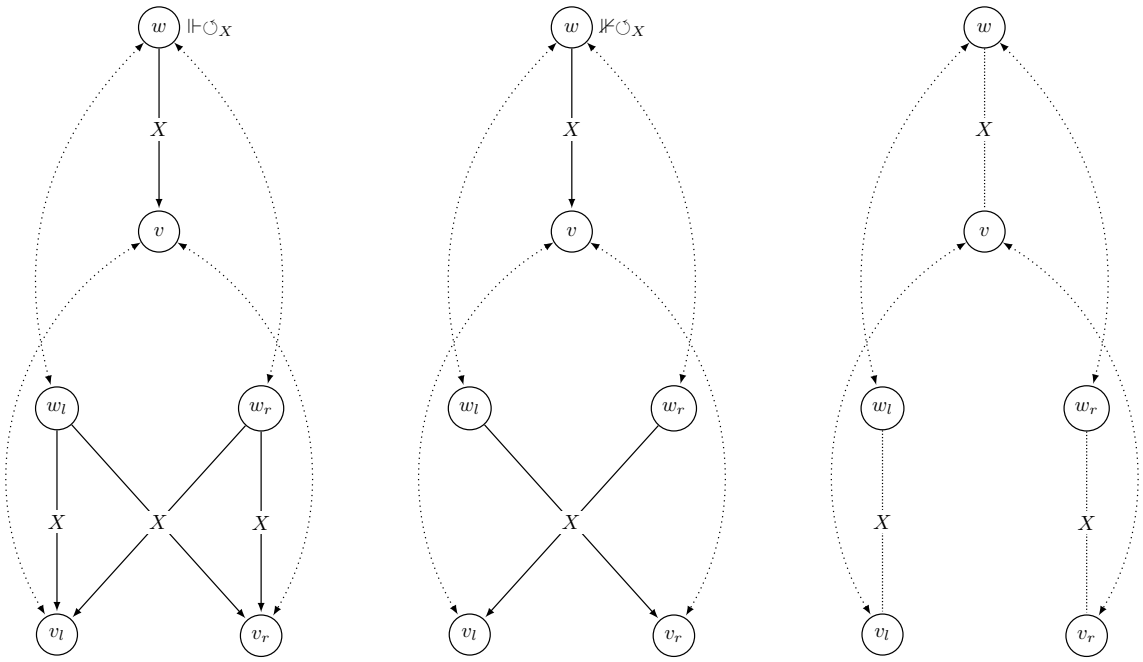
(*3*) If $w \sqsubseteq_X^{\mathbb{M}} v$ then $w_l \sqsubseteq_X^{\text{ʘ}} v_l$ and $w_r \sqsubseteq_X^{\text{ʘ}} v_r$, but never $w_l \sqsubseteq_X^{\text{ʘ}} v_r$ or $w_r \sqsubseteq_X^{\text{ʘ}} v_l$.

The reason ʘ duplicates everything in this manner is because we are preventing R_X reflexivity whenever $w \notin P_X$. This means that when $wR^{\mathbb{M}}v$, there are two situations, which are depicted in Figs. 3.1(a) and 3.1(b). The third rule is depicted in 3.1(c).

For clarity, here is how one should read these three diagrams:

- If one point w is connected to another point w' by a dotted lines with arrows at both ends ◂···▸, then those points are bisimilar.
- If one point w is connected to another point v by a solid line with an arrow and a label X $\overset{X}{\rightarrow}$, then wR_Xv, taking care to note which model we are reasoning in.
- If one point w is connected and *above* to another point v by a densely line with no arrow and a label X $\overset{X}{\cdots}$, then $w \sqsupseteq_X v$.

We summarize the mechanics of ʘ in the following proposition:



(a) $\mathbb{M}, w \Vdash \circlearrowleft_X$ (b) $\mathbb{M}, w \nVdash \circlearrowleft_X$ (c) Always

Figure 3.1: Visualizations of ↀ's operation

Proposition 3.3.40. *Let $\{w, v\} \subseteq W^{\text{ↀ}}$, and let $\{w^\circ, v^\circ\} \subseteq W^{\mathbb{M}}$ such that $w^\circ_l = w$ or $w^\circ_r = w$ and similarly for v°.*

(1) If w and v have different handedness, then

$$wR^{\text{ↀ}}_X v \text{ if and only if } w^\circ R^{\mathbb{M}}_X v^\circ$$
$$\&$$
$$w \sqsubseteq^{\text{ↀ}}_X v \text{ or } v \sqsubseteq^{\text{ↀ}}_X w \text{ never holds}$$

(2) If w and v have the same handedness, then

$$wR^{\text{ↀ}}_X v \text{ if and only if } w \in P^{\mathbb{M}}_X \ \&\ w^\circ R^{\mathbb{M}}_X v^\circ$$
$$\&$$
$$w \sqsubseteq^{\text{ↀ}}_X v \text{ if and only if } w^\circ \sqsubseteq^{\mathbb{M}}_X v^\circ$$

We shall now provide proof that ↀ gives rise to a bisimulation:

Lemma 3.3.41. *For any Kripke model $\mathbb{M} = \langle W, V, P_X, R_{\Box_X}, R_{\boxminus_X}, R_{\boxplus_X}\rangle$, we have the following bisimulation $Z : \mathbb{M} \leftrightarroweq \text{ↀ}^{\mathbb{M}}$:*

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

Proof. It follows directly from the definition of ↀ that the truth of the letters are preserved, along with the back and forth conditions for the $\sqsubseteq_X$ and $\sqsupseteq_X$ relations. The proof of the back and forth

conditions for R_X involves elementary reasoning by cases on whether $\mathbb{M}, w \Vdash \circlearrowleft_X$ or not. This simple argumentation suffices the rest of the proof. QED

We now turn to proving that this bisimulation completes a partially EvIL Kripke structure. We shall make use the mechanics of the construction of $\circlearrowleft$ heavily.

Theorem 3.3.42 (EvIL Completion)**.** *If* $\mathbb{M}$ *is partly* EvIL *Kripke structure then* $\circlearrowleft^{\mathbb{M}}$ *is an* EvIL *Kripke structure.*

Proof. We must verify that $\circlearrowleft^{\mathbb{M}}$ makes true all of the EvIL properties. We may observe that (I) through (IV) and (VI) follow by construction, and the fact that since $\mathbb{M}$ is partly EvIL by hypothesis it makes true $(I)'$ through $(IV)'$ and $(VII)'$. All that is left to show is (V) and (VII).

(V) We must show

$$(R_X^{\circlearrowleft} \circ \sqsubseteq_X^{\circlearrowleft}) \subseteq R_X^{\circlearrowleft}$$

So assume that $w \sqsubseteq_X^{\circlearrowleft} uR_X^{\circlearrowleft}v$. We know that since $w \sqsubseteq_X^{\circlearrowleft} u$ then by construction they must have the same handedness. Without loss of generality assume that both w and u are *left*, that is there is some $\{w^\circ, u^\circ\} \subseteq W^{\mathbb{M}}$ such that $w = w_l^\circ$ and $u = u_l^\circ$. By construction we have that $w^\circ \sqsubseteq^{\mathbb{M}} u^\circ$. It suffices to show that $wR_X^{\circlearrowleft}v$; to do this we shall reason by cases on the handedness of v.

Opposite – Assume that v has the opposite handedness, hence $v = v_r^\circ$ for some $v^\circ \in W^{\mathbb{M}}$. Then by construction we have that $u^\circ R_X^{\mathbb{M}} v^\circ$. This means that since $\mathbb{M}$ is partly EvIL, it makes true $(VI)'$, so $w^\circ R_X^{\mathbb{M}} v^\circ$. Hence by construction we have $wR_X^{\circlearrowleft}v$.

Same – Now assume that v has the same handedness as w and u, so $v = v_l^\circ$ for some $v^\circ \in W^{\mathbb{M}}$. Since $u_l^\circ R_X^{\circlearrowleft} v_l^\circ$ by assumption, we know from the definition of $\circlearrowleft$ that $u^\circ \in P_X^{\mathbb{M}}$. Since $\mathbb{M}$ is partly EvIL by hypothesis, and $u \sqsupseteq_X w$, then from $(V)'$ we have that $w \in P_X^{\mathbb{M}}$. But then we know by $(VI)'$ that $w^\circ R_X^{\mathbb{M}} v^\circ$, hence by construction we have $wR_X^{\circlearrowleft}v$ as desired.

(VII) We must show that "$w \in P_X^{\circlearrowleft}$ if and only if $w \sqsupseteq_X^{\circlearrowleft} w$". We know that the left to right direction holds by construction, since by assumption $\mathbb{M}$ is partly EvIL and hence makes true $(VIII)'$.

Now assume that $w \sqsupseteq_X^{\circlearrowleft} w$, then since w can see something with the same handedness as itself (namely itself), by construction we know that $w^\circ \in P_X^{\mathbb{M}}$ where $w_l^\circ = w$ or $w_r^\circ = w$. Whence $w \in P_X^{\circlearrowleft}$, which completes the argument.

QED

With these observations, we may now strengthen Theorem 3.2.35 from partly EvIL models to the fully EvIL models originally defined in Definition 2.2.28 from §2.2:

Definition 3.3.43. *As in Definition 3.2.34, we shall write*

$$\Gamma \Vdash_{\text{pEvIL}} \varphi$$

to mean that for all partly EvIL *Kripke structures* $\mathbb{M} = \langle W, R, \sqsubseteq, \sqsupseteq, V, P\rangle$*, for all worlds* $w \in W$ *if* $\mathbb{M}, w \Vdash \Gamma$ *then* $\mathbb{M}, w \Vdash \varphi$*.*

Theorem 3.3.44 (Strong EvIL Soundness and Completeness)**.**

$$\Gamma \vdash_{\text{EvIL}} \varphi \text{ if and only if } \Gamma \Vdash_{\text{EvIL}} \varphi$$

Proof. Note that every EvIL Kripke model is partly EvIL, so soundness follows immediately from Theorem 3.2.35.

Now assume that $\Gamma \nvdash_{\text{EvIL}} \varphi$, we must show that there is some witnessing EvIL model with a world that makes this false. We know from Theorem 3.2.35 that there is some partly EvIL model $\mathscr{E}$ and some w in $\mathscr{E}$ such that $\mathscr{E}, w \nVdash \varphi$ and $\mathscr{E}, w \Vdash \Gamma$. We know from Lemma 3.3.41 that $\mathscr{E} \leftrightarrows \mathfrak{S}^{\mathscr{E}}$, hence by Theorem 3.3.37, *The Fundamental Theorem of Bisimulations*, we know that $\mathfrak{S}^{\mathscr{E}}, w_l \nVdash \varphi$ and $\mathfrak{S}^{\mathscr{E}}, w_l \Vdash \Gamma$. From Theorem 3.3.42 we may observe that $\mathfrak{S}^{\mathscr{E}}$ is indeed EvIL, which means that we have found a suitable witness for completeness as desired. QED

This completes the strong, abstract completeness proof of EvIL in Kripke semantics. We shall now turn to taking stock of what we have shown so far, and discuss why we must go further to give a true proof of *completeness* of EvIL.

3.4 Taking Stock I

In the previous sections, we showed that the logic of EvIL we presented in 3.1 was complete for EvIL Kripke models. In this section we pause for a moment to discuss why we must go further, and reason what further needs to be shown in order to establish completeness.

First, recall the semantics we developed in Definition 2.1.8 in §2.1. These semantics were carefully crafted to make true the mystical Theorem 2.1.12, the *Theorem Theorem*. This equated $\Box\varphi$ with a proof of φ, in the following manner:

$$\mathfrak{M}, (a, A) \Vdash \Box_X \varphi \iff Th(\mathfrak{M}) \cup A_X \vdash_{\text{EvIL}} \varphi$$

In the above, we assume that A_X is finite. This above property of EvIL was enforced to accommodate the *Justification Principle* from §1.2, which says that when an agent believes something, she must have a reason.

This critical insight, driven by our philosophical perspective on the nature of knowledge, is lost in the abstracta of Kripke semantics. The Kripke semantics perspective on EvIL is basically meaningless on its own; for why would anyone ever care about EvIL Kripke structures, without the light of the fact that they somehow abstract EvIL semantics? We know that not every EvIL

Kripke structure can be represented by a EvIL structure by Proposition 2.2.30. How do we know that EvIL Kripke structures are faithfully abstracting our concrete semantics at all?

The connection of EvIL Kripke structures to EvIL has not yet been entirely revealed, but it is as follows:

> EvIL *models are finitary, concrete objects, and* EvIL *Kripke structures are their potentially infinite, abstract idealizations.*

Succinctly, this relationship is expressed as follows:

$$\Gamma \Vdash_{\text{EvIL}} \varphi \iff \Gamma \Vdash \varphi \tag{3.4.1}$$

... for all φ and finite Γ.

The relationship is important, since Kripke Semantics are the natural semantics for modal logics, and hence enable one to rapidly reason about them. Equation (3.4.1) allows us to see that, when thinking about EvIL, one may freely employ strong completeness and neglect concerns about failure of compactness, with the understanding that when we restrict ourselves to finitary circumstances the abstract semantics and the concrete semantics coincide.

Sections §3.5 through §3.7 shall be devoted to establishing this relationship between the abstract and concrete semantics for EvIL. Since we know that the logic of EvIL models is not compact from Theorem 2.2.31, we shall establish a *small model property* for partly EvIL and EvIL Kripke structures in §3.5. By modifying the translation system for finite Kripke structures to EvIL modifying we originally gave in the proof of Proposition 1.3.2 from §1.3, we shall show how to translate finite EvIL Kripke structures into EvIL models in §3.7. We shall find need to make use of the concept of *islands*, which we shall introduce in §3.6.

After the above developments, we shall once again take stock of our observations in §3.8. We shall prove the equation (3.4.1), and make use of our previous results to establish complexity bounds on the decision procedure for EvIL.

3.5 Small Model Construction

In this section we provide definitions and lemmas related to the subformula construction $\heartsuit^{\varphi}$. We follow [Boo95, chapter 5, pgs. 78–84] in our approach, as well as the "Fischer-Ladner Construction" used in the completeness theorem of PDL [BRV01, chapter 4, pgs. 241–248].

We first recall the definition of *pseudo-negation* from the Fischer-Ladner construction for the completeness of PDL [BRV01, chapter 4, pg. 243]. We shall also introduce *pseudo-boxes*. All of these are defined as follows:

Definition 3.5.45.

$$\sim \varphi := \begin{cases} \psi & \text{if } \varphi = \neg\psi \\ \neg\varphi & \text{o/w} \end{cases} \qquad \boxdot_X \varphi := \begin{cases} \varphi & \text{if } \varphi = \boxminus_X \psi \\ \boxminus_X \varphi & \text{o/w} \end{cases} \qquad \boxtimes_X \varphi := \begin{cases} \varphi & \text{if } \varphi = \boxplus_X \psi \\ \boxplus_X \varphi & \text{o/w} \end{cases}$$

Like pseudo-negation, the idea of pseudo-boxes is that they raise the semantic behavior of operators to the syntactic level. This is summarized in the following lemma:

Lemma 3.5.46.

$$\vdash \sim \varphi \leftrightarrow \neg\varphi \qquad \vdash \boxdot_X \varphi \leftrightarrow \boxminus_X \varphi \qquad \vdash \boxtimes_X \varphi \leftrightarrow \boxplus_X \varphi$$

$$\boxdot_X \varphi = \boxdot_X \boxdot_X \varphi \qquad \boxtimes_X \varphi = \boxtimes_X \boxtimes_X \varphi$$

Proof. We remind the reader that $\vdash$ here abbreviates $\vdash_{\text{EvIL}}$.

$\vdash \sim \varphi \leftrightarrow \neg\varphi$ – Assume that φ is unnegated, then $\sim \varphi = \neg\varphi$ and hence we know that $\vdash \neg\varphi \leftrightarrow \neg\varphi$, which suffices. If $\varphi = \neg\psi$, then we know that $\sim \varphi = \psi$, and since in classical logic we have that $\vdash \psi \leftrightarrow \neg\neg\psi$ we have the result.

$\vdash \boxdot_X \varphi \leftrightarrow \boxminus_X \varphi$ – If φ is not boxed with $\boxminus_x$, the result is trivial. So assume that $\varphi = \boxminus_X \psi$, then $\boxdot_X \varphi = \boxminus_X \psi$.

Note that for EvIL Kripke structures, for which $\boxminus_X$ corresponds to $\sqsupseteq_X$, then from EvILness we know that $\sqsupseteq_X$ is transitive and reflexive, hence $\Vdash_{\text{EvIL}} \boxminus_X \psi \leftrightarrow \boxminus_X \boxminus_X \psi$. By completeness, we know that $\vdash \boxminus_X \psi \leftrightarrow \boxminus_X \boxminus_X \psi$. But this suffices exactly what we wanted to prove.

$\vdash \boxtimes_X \varphi \leftrightarrow \boxplus_X \varphi$ – This result follows using exactly the same reasoning as above, only it uses the fact that the *dual* of $\sqsupseteq_X$, which is $\sqsubseteq_X$, is reflexive and transitive too.

$\boxdot_X \varphi = \boxdot_X \boxdot_X \varphi$ – First assume that φ is a $\boxminus_X$ boxed formula. Then $\boxdot_X \varphi = \boxdot_X \boxdot_X \varphi = \varphi$. Next assume that φ is not a $\boxminus_X$ boxed formula, then $\boxdot_X \varphi = \boxminus_X \varphi$, and hence

$$\begin{aligned} \boxdot_X \boxdot_X \varphi &= \boxdot_X \boxminus_X \varphi \\ &= \boxminus_X \varphi \\ &= \boxdot_X \varphi \end{aligned}$$

$\boxtimes_X \varphi = \boxtimes_X \boxtimes_X \varphi$ – The proof of this assertion is exactly the same as the proof for $\boxdot_X$.

QED

We shall use these operations above in the subformula construction we will carry out. Next, we introduce an operation which will allow us to restrict our subformulae to precisely the finitary number of agents that shall be relevant.

Definition 3.5.47. *Let $\delta(\varphi) \subseteq \mathcal{A}$ be the set of agents that occur in φ*

We now employ primitive recursion to define the finite set of formulae that we shall use in our construction, which we have labeled Σ. This operation behaves as follows:

- Σ takes as input:

- A set of agents Δ
- A formula φ where $\varphi \in \mathcal{L}(\mathcal{A}, \Phi)$

- Σ outputs a set S of $\mathcal{L}(\mathcal{A}, \Phi)$ formulae (that is, $S \subseteq \mathcal{L}(\mathcal{A}, \Phi)$)

We may summarize this concisely as the following type signature:

$$\Sigma : (\wp\mathcal{A}) \times \mathcal{L}(\mathcal{A}, \Phi) \to \wp\mathcal{L}(\mathcal{A}, \Phi)$$

Definition 3.5.48. *Define* $\Sigma(\Delta, \varphi)$ *using primitive recursion as follows:*

$$\begin{aligned}
\Sigma(\Delta, p) :=& \{p, \neg p, \bot, \neg\bot\} \cup \\
& \{\boxminus_X p, \neg \boxminus_X p, \boxplus_X p, \neg \boxplus_X p \mid X \in \Delta\} \\
\Sigma(\Delta, \bot) :=& \{\bot, \neg\bot\} \\
\Sigma(\Delta, \circlearrowleft_X) :=& \{\circlearrowleft_X, \neg \circlearrowleft_X, \boxminus_X \circlearrowleft_X, \neg\boxminus_X \circlearrowleft_X, \bot, \neg\bot\} \\
\Sigma(\Delta, \varphi \to \psi) :=& \{\varphi \to \psi, \neg(\varphi \to \psi)\} \cup \Sigma(\Delta, \varphi) \cup \Sigma(\Delta, \psi) \\
\Sigma(\Delta, \Box_X \varphi) :=& \{\Box_X \varphi, \neg\Box_X \varphi, \boxplus_X \Box_X \varphi, \neg \boxplus_X \Box_X \varphi\} \cup \\
& \{\Box_X \boxslash_Y \varphi, \neg\Box_X \boxslash_Y \varphi, \\
& \Box_X \boxtimes_Y \varphi, \neg\Box_X \boxtimes_Y \varphi, \\
& \boxslash_Y \varphi, \neg\boxslash_Y \varphi, \\
& \boxtimes_Y \varphi, \neg\boxtimes_Y \varphi \mid Y \in \Delta\} \cup \\
& \Sigma(\Delta, \varphi) \\
\Sigma(\Delta, \boxminus_X \varphi) :=& \{\boxminus_X \varphi, \neg \boxminus_X \varphi\} \cup \Sigma(\Delta, \varphi) \\
\Sigma(\Delta, \boxplus_X \varphi) :=& \{\boxplus_X \varphi, \neg \boxplus_X \varphi\} \cup \Sigma(\Delta, \varphi)
\end{aligned}$$

To understand how the above operates, we assume that the reader has some background in recursive programming. Recall how "subformulae" are defined for the Fischer-Ladner construction for the completeness proof of PDL. We can see that authors describe the set of subformulae $\neg FL(\Sigma)$ as follows:

> We defined $\neg FL(\Sigma)$, the *closure of* Σ, as the smallest set containing which is Fischer-Ladner closed and closed under single negations [BRV01, pg. 243].

Here Fischer-Ladner closed means the construction satisfies certain subformula properties, such as "if $\langle\pi_1;\pi_2\rangle\varphi \in \neg FL(\Sigma)$ then $\langle\pi_1\rangle\langle\pi_2\rangle\varphi \in \neg FL(\Sigma)$." We ask the reader who knows a little about computers, how would one go about programming the Fischer-Ladner closure? The easiest way to program the Fischer-Ladner closure, in languages like Haskell or OCaml, would be to use pattern recognition and (primitive) recursion. This is just as we have done, informally, for the EvIL subformulae construction.

We argue that writing a concise, programmatic recursive characterization as we have done for Σ is the easiest way to express the set with the features we desire. For one thing, the closure properties we shall want depend at the top-level on a constant set of agents, which are calculated at the

beginning of the construction. Moreover, as we shall see, we shall need some formulae boxed in certain ways to ensure certain partly EviL properties and certain formulae boxed in other ways to ensure other partly EviL properties. Worse yet – since we have multiple kinds of pseudo operators, we cannot enforce closure for all of them. Managing the priorities of when a formula should be closed for which operations roughly amounts to giving the algorithmic characterization we have carried out above.

In our subsequent proofs, we shall capitalize on combinatoric properties that our subformulae operation obeys. Some of these features are summarized in the following proposition.

Proposition 3.5.49. $\Sigma(\delta(\varphi), \varphi)$ *is finite. Moreover, we have the following:*

- $\varphi \in \Sigma(\delta(\varphi), \varphi)$
- *If* $\psi \in \Sigma(\delta(\varphi), \varphi)$ *and* χ *is a subformula of* ψ*, then* $\chi \in \Sigma(\delta(\varphi), \varphi)$
- *If* $\psi \in \Sigma(\delta(\varphi), \varphi)$ *then* $\sim \psi \in \Sigma(\delta(\varphi), \varphi)$
- *If* $\boxminus_X \varphi \in \Sigma(\delta(\varphi), \varphi)$ *then* $\boxslash_X \varphi \in \Sigma(\delta(\varphi), \varphi)$
- *If* $\boxplus_X \varphi \in \Sigma(\delta(\varphi), \varphi)$ *then* $\boxtimes_X \varphi \in \Sigma(\delta(\varphi), \varphi)$

We follow [BRV01, pg. 243] in our definition of the set of (relativized) maximally consistent sets:

Definition 3.5.50 (Atoms)**.** *Let* $At(\Psi)$ *denote the maximally consistent subsets of* Ψ

We next have the Finitary Lindenbaum Lemma:

Lemma 3.5.51 (Finitary Lindenbaum Lemma)**.** *If* $\Gamma \nvdash \varphi$ *and* $\Gamma \subseteq \Sigma(\delta(\varphi), \varphi)$*, then there is a* $\gamma \in At(\Sigma(\delta(\varphi), \varphi))$ *such that* $\Gamma \subseteq \gamma$ *and* $\gamma \nvdash \varphi$

Proof. The proof of this assertion follows the same proof of the finitary Lindenbaum Lemma offered in [BRV01, Lemma 4.83, pg. 244] and [Boo95, chapter 5, pg. 79]. QED

We now turn to defining the EvIL subformula model we shall use in the subsequent finitary completeness theorem:

Definition 3.5.52. *Define*

$$\circledcirc^{\varphi} := \langle W, R, \sqsubseteq, \sqsupseteq, V, P \rangle$$

Where:

$$
\begin{aligned}
W :=& At(\Sigma(\delta(\varphi), \varphi)) \\
V(p) :=& \{w \in W \mid p \in w\} \\
P_X :=& \{w \in W \mid \circlearrowleft_X \in w\} \cup \{w \in W \mid X \notin \delta(A)\} \\
R_X :=& \begin{cases} \{(w,v) \in W \times W \mid \{\psi \mid \Box_X \psi \in w\} \subseteq v\} & \text{when } X \in \delta(\varphi) \\ \varnothing & \text{o/w} \end{cases} \\
\sqsupseteq_X :=& \begin{cases} \{(w,v) \in W \times W \mid \{\psi, \boxdot_X \psi \mid \boxdot_X \psi \in w\} \subseteq v \ \& \\ \qquad\qquad \{\psi, \boxtimes_X \psi \mid \boxtimes_X \psi \in v\} \subseteq w\}\} & \text{when } X \in \delta(\varphi) \\ \{(w,w) \mid w \in W\} & \text{o/w} \end{cases} \\
\sqsubseteq_X :=& \begin{cases} \{(v,w) \in W \times W \mid \{\psi, \boxdot_X \psi \mid \boxdot_X \psi \in w\} \subseteq v \ \& \\ \qquad\qquad \{\psi, \boxtimes_X \psi \mid \boxtimes_X \psi \in v\} \subseteq w\}\} & \text{when } X \in \delta(\varphi) \\ \{(w,w) \mid w \in W\} & \text{o/w} \end{cases}
\end{aligned}
$$

In the above construction, we note that the definition of $V(p)$, P_X and R_X are defined as usual. Only $\sqsupseteq_X$ and $\sqsubseteq_X$ have are unusual. Here we are consciously imitating the completeness techniques given in [Boo95, chapter 5, pgs. 78–84], and using them for EvIL.

We shall now show that ☺^{φ} satisfies the *Truth lemma*. Once again, the method of the proof of the following theorem is adapted from [Boo95, chapter 5, pgs. 78–84].

Lemma 3.5.53 (Truth Lemma)**.** *For any subformula $\psi \in \Sigma(\delta(\varphi), \varphi)$ and any $w \in W^{\text{☺}}$, we have that*

$$\text{☺}^{\varphi}, w \Vdash \psi \iff \psi \in w$$

Proof. The proof proceeds by induction on ψ.

$p \in \Phi$**,** $\circlearrowleft_X$**,** $\bot$ – These steps are elementary.

$\varphi \to \psi$ – Since we know that $\Sigma(\delta(\varphi), \varphi)$ is closed under subformulae, from the inductive hypothesis we have that $\text{☺}^{\varphi}, w \Vdash \varphi \iff \varphi \in w$ and $\text{☺}^{\varphi}, w \Vdash \varphi \iff \varphi \in w$. The rest of the step involves reasoning by cases, using the fact that $\Sigma(\delta(\varphi), \varphi)$ is closed under pseudo-negation, w is maximal and pseudo-negation logically equivalent to negation.

$\Box_X \varphi$ – The right to left direction follows by the fact that $\Sigma(\delta(\varphi), \varphi)$ is closed under subformulae, and the inductive hypothesis. Hence we shall concern ourselves with the left to right direction.

So assume that $\Box_X \psi \notin w$ we shall show that there is a v such that wR_xv and $\text{☺}^{\varphi}, v \nVdash \psi$. Consider the set

$$\Xi := \{\sim \psi\} \cup \{\chi \mid \Box_X \chi \in w\}$$

Note that $\Xi \subseteq \Sigma(\delta(\varphi), \varphi)$. If Ξ is consistent, then $\Xi \nvdash \varphi$ and we know from the Lindenbaum Lemma that Ξ may be extended to the v we desire.

So suppose towards a contradiction that Ξ is not consistent. Then $\vdash \neg \bigwedge \Xi$, which means by classical logic that:

$$\vdash \left(\bigwedge_{\Box_X \chi \in w} \chi \right) \to \psi$$

But then we know by modal logic that:

$$\vdash \left(\bigwedge_{\Box_X \chi \in w} \Box_X \chi \right) \to \Box_X \psi$$

This means that since $w \vdash \bigwedge_{\Box_X \chi \in w} \Box_X \chi$, then we have that $\Box_X \psi \in w$ by maximality after all. This is ridiculous! $\lightning$

$\boxminus_X \varphi$ – This case is similar to the case for $\Box_X \varphi$, but harder to understand.

We shall demonstrate the left to right direction, since right to left is elementary. Assume that $\boxminus_X \psi \notin w$, then we shall find a v such that $w \sqsupseteq_x v$ and $\circledcirc^{\varphi}, v \nVdash \varphi$. Since w is maximal and $\vdash \boxdot_X \psi \leftrightarrow \boxminus_X \psi$, we have that $w \nvdash \boxdot_X \psi$.

Now abbreviate:

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

As before, if $\{\sim \psi\} \cup A \cup B$ consistent then it extends to the desired v.

So suppose towards a contradiction that $\{\sim \psi\} \cup A \cup B \vdash \bot$. Then $A \cup B \vdash \psi$, and furthermore by the equivalences in Lemma 3.5.46 and rule (3) and the axioms we have that[3]

$$\boxdot_X A \cup \boxdot_X B \vdash \boxdot_X \psi.$$

So let

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

Since $\boxdot_X \boxdot_X \chi = \boxdot_X \chi$ by Lemma 3.5.46, we have $A' = \boxdot_X A$. Moreover, by Lemma 3.5.46, axiom (12), and classical logic we can see that

$$\vdash \sim \chi \to \boxdot_X \sim \boxtimes_X \chi$$

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

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

However, evidently $A' \cup B' \subseteq w$, hence $w \vdash \boxdot_X \psi$, which is a contradiction! $\lightning$

To complete the argument, we must show that $w \sqsupseteq_X v$. Since $A \subseteq v$ we just need to check that $\{\psi, \boxtimes_X \psi \mid \boxtimes_X \psi \in v\} \subseteq w$. Suppose that $\boxtimes_X \psi \in v$ but $\psi \notin w$. Since w is maximally

[3]Here $\boxdot_X S$ is shorthand for $\{\boxdot_X \chi \mid \chi \in S\}$.

consistent we have then that $\sim \psi \in w$, hence $\sim \boxtimes_X \in B$ by definition and thus $\sim \boxtimes_X \psi \in v$, since $B \subseteq v$. This contradicts that v is consistent. $\lightning$

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 3.5.46 that $\boxtimes_X \boxtimes_X \psi = \boxtimes_X \psi$, which once again implies that v is inconsistent. $\lightning$

$\boxplus_X \varphi$ – This is exactly as in the case of $\boxminus_X \varphi$, only the appropriate dual assertions of all of the statements used are employed.

QED

We shall now turn to establishing that our finite model is indeed a partly EvIL Kripke structure, following the manner that we used to show the same result for the canonical EvIL model $\mathscr{E}$.

Lemma 3.5.54 (☺^φ is Partly EvIL)**.** ☺^φ *is a finite partly* EvIL *Kripke structure.*

Proof. The fact that ☺^φ is finite follows from the fact that $W^{\text{☺}} \subseteq \Sigma(\delta(\varphi), \varphi)$ and $\Sigma(\delta(\varphi), \varphi)$ is itself finite, as we established in Lemma 3.5.46.

The remainder of the proof is devoted to establishing the partly EvIL properties for ☺^φ.

(I)′ We must show "$\sqsubseteq_X$ is reflexive." We first note that if $X \notin \delta(\varphi)$, then this is true trivially by the construction of ☺^φ.

So assume that $X \in \delta(\varphi)$. We need to show two facts:

$$\{\psi, \boxdot_X \psi \mid \boxdot_X \psi \in w\} \subseteq w$$
$$\&$$
$$\{\psi, \boxtimes_X \psi \mid \boxtimes_X \psi \in w\} \subseteq w$$

However, these facts follow from Lemma 3.5.46, the maximality of w and the fact that we know EvIL logic proves:

$$\vdash \boxminus_X \psi \to \psi$$
and
$$\vdash \boxplus_X \psi \to \psi$$

We know that EvIL proves these facts from our previous completeness theorem, Theorem 3.3.44, and the fact that $\sqsubseteq_X$ is reflexive for EvIL models.

(II)′ A quick glance at the definition of ☺^φ reveals that "$\sqsubseteq_X$ is transitive" is immediate by construction.

(III)′ Once again, the assertion "$\sqsubseteq_X$ is the reverse $\sqsupseteq_X$" follows immediately by construction.

(*IV*)′ Assume $w \sqsubseteq_X v$, we shall show that

$$w \in V(p) \text{ if and only if } v \in V(p)$$

If $X \notin \delta(\varphi)$ then we know that $w = v$, so the above is true.

So assume that $X \in \delta(\varphi)$. Now assume that $w \in V(p)$, then $\mathscr{E}, w \Vdash p$. This means that $p \in w$, whence $p \in \Sigma(\delta(\varphi), \varphi)$. By Definition 3.5.48 we have that $\boxplus_X p = \boxtimes_X p \in \Sigma(\delta(\varphi), \varphi)$. By axiom (6) and the Truth Lemma (Lemma 3.5.53), we have that $\boxtimes_X p \in w$, whence $p \in v$ by construction of $\heartsuit^{\varphi}$. The other direction is similar, however it uses axiom (5) instead.

(*VI*)′ To prove the assertion

$$(R_X \circ \sqsubseteq_X) \subseteq R_X$$

We first note that if $X \notin \delta(\varphi)$, then $R_X \circ \sqsubseteq_X = R_X = \varnothing$ by construction. Hence we may safely assume $X \in \delta(\varphi)$.

Next assume that $w \sqsubseteq_X u$ and uR_Xv, we need to show that wR_Xv. In order to do this, by construction we need to show that if $\Box_X\varphi \in w$ then $\varphi \in v$. However, we know that $\vdash \Box_X\varphi \rightarrow \boxtimes_X\Box_X\varphi$ by the logic of EvIL, and by the definition of 3.5.48 we know that if $\Box_X\varphi \in \Sigma(\delta(\varphi), \varphi)$ then $\boxtimes_X\Box_X\varphi \in \Sigma(\delta(\varphi), \varphi)$ too, so $\boxtimes_X\Box_X\varphi \in w$ by maximality. However, we then have that $\Box_X\varphi \in u$ by construction, and thus $\varphi \in v$ as desired.

(*VII*)′ We must show:

$$(\sqsubseteq_Y \circ R_X) \subseteq R_X$$
$$\&$$
$$(\sqsupseteq_Y \circ R_X) \subseteq R_X$$

We shall only show that $\sqsubseteq_Y \circ R_X \subseteq R_X$, since $\sqsupseteq_Y$ is analogous.

First, observe that we may assume that $Y \in \delta(\varphi)$, since if not then $\sqsubseteq_Y = id$, the identity relation, so $\sqsubseteq_Y \circ R_X = R_X$ which suffices. Now assume that $\Box_X\varphi \in w$, wR_Xv and $v \sqsubseteq_Y u$; we must show that $\varphi \in u$. Since $Y \in \delta(\varphi)$ then by the construction of Σ, along with the EvIL fact that $\vdash \Box_X\varphi \rightarrow \Box_X\boxtimes_X\varphi$, we know that $Nec_X\boxtimes_Y\varphi \in w$, whence $pBP_Y\varphi \in v$. Since $v \sqsubseteq_Y u$, we have that $\varphi \in u$ as desired.

(*VIII*)′ To show "If $w \in P(X)$ then wR_Xw," assume $\circlearrowleft_X \in w$ and $\Box_X\varphi \in w$. Then we know by EvIL and maximality that $\varphi \in w$, which suffices to show that wR_Xw.

(*V*)′ We must show "If $w \in P_X$ and $w \sqsupseteq_X v$ then $v \in P_X$." If $w \in P_x$ then by construction we know that $\circlearrowleft_X \in w$. Note that this can only happen if $X \in \delta(\varphi)$. As in the case of $(IV)'$ we know by the definition of Σ, EvIL logic and maximality we have that $\boxminus_X \circlearrowleft_X \in w$. Whence we have $\circlearrowleft_X \in v$ as desired.

QED

We may combine the results above with what we have shown previously in §3.3 to give the following series of results:

Theorem 3.5.55 (Weak EvIL Soundness and Completeness)**.**

EvIL *is weakly sound and complete for finite* EvIL *Kripke structures.*

Proof. Since soundness is straightforward, we only prove completeness.

Assume that $\nvdash \varphi$, in other words we have $\varnothing \vdash \varphi$. This means by the Lindenbaum Lemma that $\varnothing$ may be extended to some $w \in At(\delta(\varphi), \varphi)$ such that $\heartsuit^{\varphi}, w \nVdash \varphi$. We know that $\heartsuit^{\varphi}$ is partly EvIL from Theorem 3.5.54, so from Theorem 3.3.42 we have that $\heartsuit^{\heartsuit}$ is an EvIL model, and since it is bisimilar to $\heartsuit$ we know that $\heartsuit^{\heartsuit}, w_l \nVdash \varphi$. Now note that for $\heartsuit$ obeys the following rule: If $|W|$ is finite then $|W^{\heartsuit}| = 2 \times |W|$ (since all that $\heartsuit$ is doing is making duplicates of all of the worlds). Hence we know that $\heartsuit^{\heartsuit}$ is indeed finite, which means it is a suitable witness. QED

Theorem 3.5.56 (EvIL Small Model Property)**.** *If φ is satisfiable by some* EvIL *Kripke structure, then φ is satisfied by some finite* EvIL *Kripke structure $\mathbb{M}$ where $|\mathbb{M}| \in O(EXP2(|\varphi|))$*

Proof. Assume that φ is satisfiable in some EvIL Kripke structure, then we know by soundness that $\nvdash \varphi$, hence $\heartsuit^{\varphi}, w \nVdash \varphi$ for some w extending $\varnothing$. So it suffices to show that $|\heartsuit^{\varphi}| \in O(EXP2(|\varphi|))$.

Note that $\heartsuit^{\varphi} \subseteq \wp(\Sigma(\delta(\varphi), \varphi))$. We shall show that $|\Sigma(\delta(\varphi), \varphi)| \in O(EXP(|\varphi|))$, then we would know that $|\wp(\Sigma(\delta(\varphi), \varphi))| \in O(EXP2(|\varphi|))$, which would suffice to show the result.

First observe that $|\delta(\varphi)| \in O(EXP(|\varphi|))$. This is because in a worst case scenario, φ is constantly branching into $\psi \to \chi$ formulae, and adding two new agents every time it branches. However, it cannot have more than $3^{|\varphi|}$ agents even in this worst case scenario, so we know that $|\delta(\varphi)| \in O(EXP(|\varphi|))$.

To finish the argument, we again perform a worst-case analysis. Every non-branching step (ie. every time Σ processes a formula not of the form $\psi \to \chi$) of $\Sigma(\delta(\varphi), \zeta)$ introduces at worst $O(|\delta(\varphi)|) \in O(EXP(|\varphi|))$ many formulae. In a worst case scenario $\Sigma(\delta(\varphi), \varphi)$ must branch $O(EXP(|\varphi|))$ times and each time perform a $O(EXP(|\varphi|))$ operation. Even in this worst case scenario, the complexity is still $O(EXP(|\varphi|))$, which is as we claimed. QED

Theorem 3.5.57 (EvIL Decidability)**.** EvIL *is decidable and the time complexity of the decision problem for* EvIL *is bounded above by $O(EXP3(|\varphi|))$*

Proof. We know that $\heartsuit^{\varphi} \in \wp(\wp(\Sigma(\delta(\varphi), \varphi)))$, we know that φ is not a tautology of EvIL if and only if there is a suitable EvIL witnessing Kripke structure in $\wp(\wp(\Sigma(\delta(\varphi), \varphi)))$, defined in the manner of $\heartsuit$. So a decision procedure to check if φ is an EvIL tautology is to check every member of $\wp(\wp(\Sigma(\delta(\varphi), \varphi)))$ to see it gives rise to an EvIL model with some world which disproves φ. Since, as we saw in the proof of Theorem 3.5.56, we know that $|\Sigma(\delta(\varphi), \varphi)| = O(EXP(|\varphi|))$, this procedure takes $O(EXP3(|\varphi|))$ many steps to complete. QED

We shall now move on to showing how we may recover a concrete EvIL model from a finite EvIL Kripke structure. The above results shall ensure completeness of EvIL for its intended semantics. Before proceeding we shall first need to introduce the concept of a *island*.

3.6 Islands

In this section, we discuss *islands* in EvIL models and present crucial features they make true. These also help one to understand how to visualize EvIL models.

Definition 3.6.58. *Let* $\mathbb{M}$ *be a partly* EvIL *Kripke structure. Define:*

$$\lceil w \rfloor := \left\{ v \;\middle|\; w \left(\bigcup_{X \in \mathcal{A}} \sqsubseteq_X \cup \sqsupseteq_X \right)^* v \right\}$$

Here $\rightsquigarrow^*$ *is the reflexive transitive closure of a relation* $\rightsquigarrow$*. We say that* $\lceil w \rfloor$ *is the* ***island*** *that* w *belongs to.*

Islands are a rather important concept in EvIL, which have implicitly played a role in our intuitions prior to this point. Before carrying on, we shall go over several ways to think about islands before proceeding.

(*i*) One way to understand $\lceil w \rfloor$ is that this is the set of worlds that are graph reachable from w using $\sqsubseteq_X$ and $\sqsupseteq_X$ for any agent X. Since we are considering both $\sqsubseteq_X$ and $\sqsupseteq_X$, then we know that we are thinking about graph reachability on undirected graphs. This means that $\lceil w \rfloor$ gives rise to *equivalence classes* over the worlds in an EvIL Kripke model.

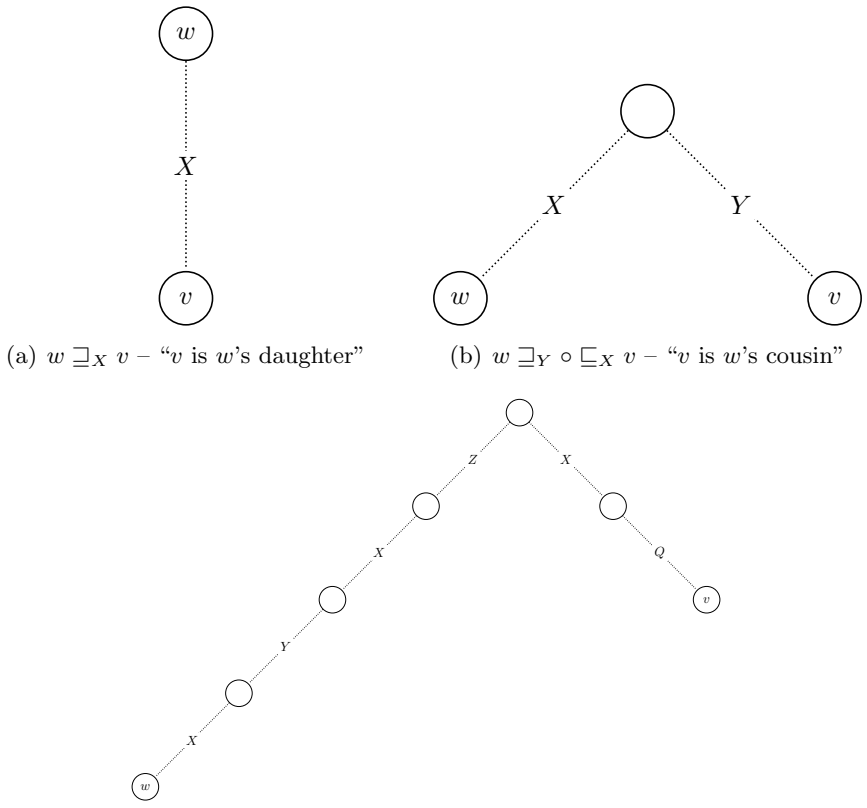
(*ii*) Another way to understand $\lceil w \rfloor$, which we shall return to in §3.7 with the idea of *surnames*, is that it represents w's *extended family*. For instance, we might think that if $w \sqsupseteq_X v$ then v is w's daughter, while $w \sqsupseteq_Y \circ \sqsubseteq_X v$ means that v is w's cousin. These sorts of relationships are depicted in Figs. 3.2(a), 3.2(b), and 3.2(c). Of course, this analogy is perhaps most pleasant to think about in the case of one agent – if there are multiple agents, then complicated "inbreeding" situations can happen where $w \sqsubseteq_X v$ and $w \sqsupseteq_X v$ but $w \neq v$.

(*iii*) A final way to think about islands is to remember the discussion we originally presented in §1.6. This way of thinking about islands makes the most sense in the single agent case. Every island is a can be thought of as a connected poset. For instance we can see that in both Figs. 2.1(a) and 2.1(b), there are two islands in each graph. Note that as we asserted in §1.6, as one travels down a belief poset, one can imagine more things. EvIL Kripke models are good abstractions on this intuition; indeed, property (V) and axiom 7 reflect exactly this idea. Anticipating what we shall reveal in lemma 3.6.59, we have pictured an agent's island in Fig. 3.3. If we try to think about how many worlds a node in a poset can access as its "width", we can imagine islands as *Christmas trees*, since they are fatter for lower nodes and thinner for upper nodes. We have depicted the Christmas tree analogy in this in Fig. 3.4.

In a multi-agent setting, we might think of an island as combined belief networks of agents, glued together yet still independent.

In many ways, *islands* behave as a single entity; this is precisely in accordance with reading (*iii*) above. We summarize the ways they behave in the following lemma:

Lemma 3.6.59 (Island Lemma)**.** *The following hold if* $\mathbb{M}$ *is partly* EvIL*:*



(a) $w \sqsupseteq_X v$ – "v is w's daughter"

(b) $w \sqsupseteq_Y \circ \sqsubseteq_X v$ – "v is w's cousin"

(c) "v is w's second cousin twice removed"

Figure 3.2: Family relationships within columns

(*1*) *For all* w *we have* $w \in \lceil w \rceil$

(*2*) *If* $w \in \lceil v \rceil$ *then* $\lceil w \rceil = \lceil v \rceil$

(*3*) *If* wR_Xv *then for all* $u \in \lceil v \rceil$ *we have* wR_Xu

(*4*) wR_Xv *if and only if* $wR_X\lceil v \rceil$[4]

(*5*) *If* $w \in \lceil v \rceil$ *then* $w \in V(p)$ *if and only if* $v \in V(p)$ *for all* $p \in \Phi$

Proof.

(*1*) This follows since by $(I)'$ we know that $\sqsubseteq_X$ is reflexive.

(*2*) This follows from the general fact that if w is graph-reachable from v, then u is graph reachable from w if and only if u is graph reachable from v.

(*3*) Assume that wR_Xv, and assume that $v \left(\bigcup_{X \in \mathcal{A}} \sqsubseteq_X \cup \sqsupseteq_X\right)^* u$. To show that wR_Xu, use $(VII)'$, that $(\sqsubseteq_Y \circ R_X) = R_X = (\sqsupseteq_Y \circ R_X)$, and induction on the path length from v to u.

(*4*) With (*1*), this is a equivalent to (*3*).

[4]By a minor abuse of notation, $wR_X\lceil v \rceil$ means that for all $u \in \lceil v \rceil$, wR_Xu.

(*5*) Assume that $w \in \lceil v \rfloor$, and that $w \in V(p)$. Then we may induct on path length, and use $(IV)'$ to see that $v \in V(p)$. From (*1*) and (*2*) above, we know that $w \in \lceil v \rfloor$ implies that $v \in \lceil w \rfloor$, so we can see hat the converse holds true too.

QED

The above lemma asserts that islands are to be thought of as worlds in of themselves - for they make true the same letters and can only be accessed as a unit. Moreover, we know from $(VI)'$, we can see in the single agent case that as an agent "ascends" in an island, they can access fewer worlds, which may be equated with holding more beliefs.

We hope that the above discussion provides some insight into how to think of islands in an intuitive manner. In the next section, we shall show how to leverage the concept of an island to show how we may translate a finite EvIL Kripke structures witnessing a formula φ into corresponding EvIL models.

3.7 Translation & Evil Completeness

In this section, we turn to showing that every finite EvIL Kripke structure $\mathbb{M}$ has a corresponding EvIL model $\bigstar$ which is an (almost)-homomorphic projection[5]. Assuming that Φ is infinite and $\Psi \subseteq_\omega \Phi$, then we shall show that $\mathbb{M}$ and $\bigstar$ agree on the language $\mathcal{L}(\Psi, \mathcal{A})$. The method of the proof of this correspondence generalizes the elementary argument presented in Proposition 1.3.2 from §1.3. From this correspondence, we shall obtain a weak completeness theorem for EvIL and its intended semantics.

Recall that in the proof of Proposition 1.3.2 we assumed an infinite store of unused letters, and assigned them to worlds in order to control the accessibility in the EvIL model we constructed. This was embodied by a function $p : W \rightarrowtail \Phi \backslash L(\varphi)$; for each world w, p_w was the *name* we assigned to it. In our construction here, we shall extend this metaphor, using a generic finite set $\Psi \subseteq_\omega \Phi$.

Recall that among the three principle ways we described for thinking about think about islands, one way to think of $\lceil w \rfloor$ was as w's extended family. So along with *personal names*, we shall also want to assign family names or *surnames*.

With these above considerations in mind, we offer the following definition:

Definition 3.7.60. *Assume the set of letters Φ is infinite, and fix a finite $\Psi \subseteq_\omega \Phi$, a finite* EvIL *Kripke model $\mathbb{M}$*

- *Let Ψ be a finite set of proposition letters.*
- *Let*
$$\Lambda := \{\{w\}, \lceil w \rfloor \mid w \in W^{\mathbb{M}}\}$$
That is, Λ is the set of worlds and islands.

[5]Note that we shall not provide a formal definition of what it means for a map to be (almost)-homomorphic, since we consider this concept more intuitive than formal. Intuitively, two objects are *(almost)-homomorphic* when they are homomorphic for all intents and purposes.

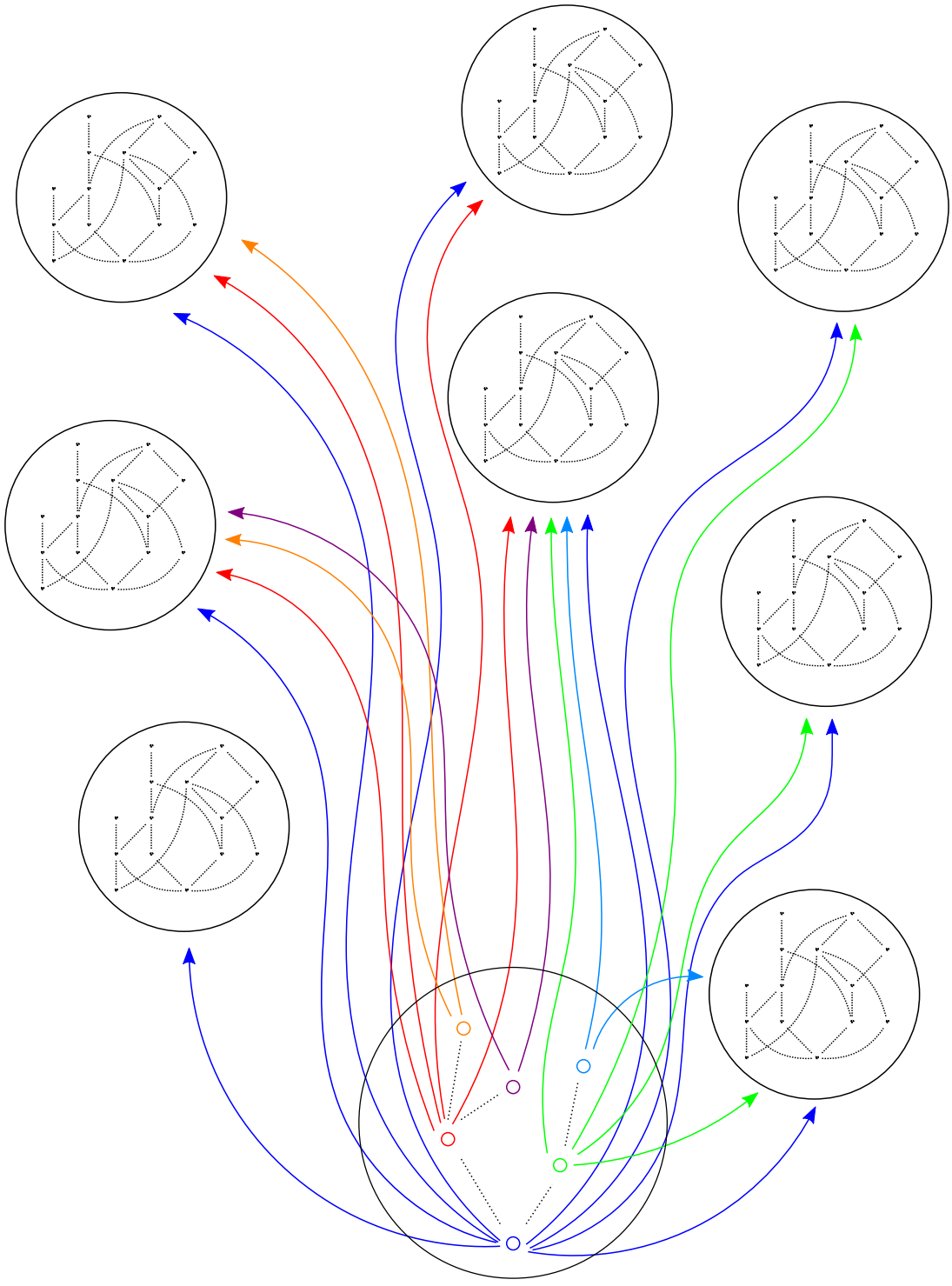


Figure 3.3: The inner functioning of an island and its relations to other islands

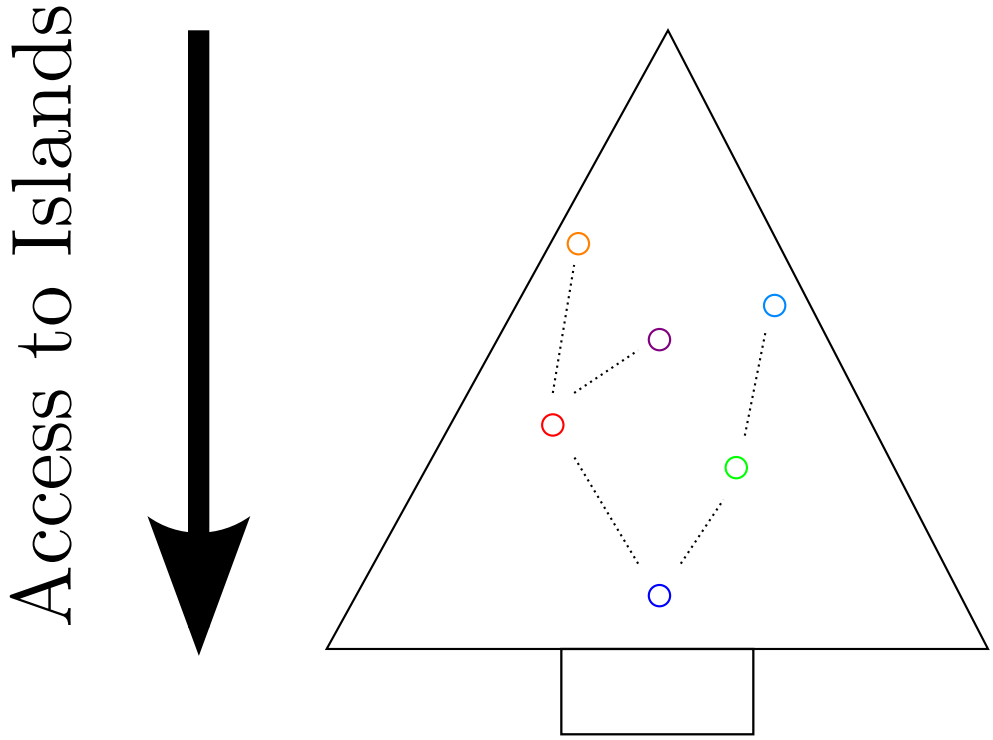


Figure 3.4: An island is like a *Christmas tree*

- *Let* $p : \Lambda \rightarrowtail \Phi \backslash \Psi$ *be an injection, assigning names to worlds and surnames to islands*[6].
- *Let* $\vartheta : W^{\mathbb{M}} \to \wp\Phi \times (\wp(\mathcal{L}_0(\Phi)))^{\mathcal{A}}$ *be defined such that:*

$$\vartheta(w) := (\vartheta_1(w), \vartheta_2(w))$$

Where:

  - $\vartheta_1 : W^{\mathbb{M}} \to \wp\Phi$ *is defined to be:*

$$\vartheta_1(w) := \{q \in \Psi \mid \mathbb{M}, w \Vdash q\} \cup \{p_{\ulcorner w \urcorner}\}$$

    We may understand ϑ_1 *as providing a propositional valuation to worlds in* $W^{\mathbb{M}}$

  - $\vartheta_2 : W^{\mathbb{M}} \to \wp(\mathcal{L}_0(\Phi))$ *is defined to be:*

$$\vartheta_2(w) := \prod_{X \in \mathcal{A}} \vartheta_{2A}(w, X) \cup \vartheta_{2B}(w, X)$$

    Where:

    - $\vartheta_{2A} : W^{\mathbb{M}} \times \mathcal{A} \to \wp(\mathcal{L}_0(\Phi))$ *is defined to be:*

$$\vartheta_{2A}(w, X) := \{\neg p_{\ulcorner v \urcorner} \mid \neg w R_X v\}$$

[6]Subsequently, we shall abbreviate $p(\{w\})$ as p_w and $p(\ulcorner w \urcorner)$ as $p_{\ulcorner w \urcorner}$

◇ $\vartheta_{2B} : W^{\mathbb{M}} \times \mathcal{A} \to \wp(\mathcal{L}_0(\Phi))$ *is defined to be:*

$$\vartheta_{2B}(w, X) := \{\bot \to p_v \mid w \sqsupseteq_X v\}$$

We may understand ϑ_2 as providing, for each agent, a corresponding set of propositional formulae. These formulae constitute their set of basic beliefs, as we originally introduced in §1.3.

ϑ_{2A} and ϑ_{2B} each constitute a component that goes into the basic belief set we assign to a particular agent.

- *Let* $\maltese := \vartheta[W]$

Certain remarks must be made regarding the above definition.

For one, note that we are ensured by the axiom of choice that p_w is well defined, since by hypothesis we have that W is finite, whence $\wp W$ is finite and since $\Lambda \subseteq \wp W$ we know that Λ is finite as well. Since we know that Φ is infinite then $\Phi \backslash \Psi$ is infinite as well, and there always exists an embedding of a finite set into an infinite set.

To be completely explicit about our intentions, $\maltese$ is an EvIL model we are constructing which shall preserve the truth of φ for all of the worlds in $\mathbb{M}$. Our goal is that $\maltese$ should be an *(almost)-homomorphic projection of* $\mathbb{M}$ *under* ϑ with respect to a language $\mathcal{L}(L, \mathcal{A})$, where L is a finite set of letters. This is precisely why we have set $\maltese$ to be the image of W under ϑ. Permit us to explain what "almost homomorphic" means exactly.

Recall the definition of $\mho^{\maltese}$ from Definition 2.2.25 from §2.2. This defines $\sqsubseteq^{\maltese}$, $\sqsupseteq^{\maltese}$, and $R^{\maltese}$. To ensure that $\maltese$ is (almost)-homomorphic to $\mathbb{M}$, we shall want to enforce the following relationships:

$$q \in \Psi \Longrightarrow (\mathbb{M}, w \Vdash q \iff \maltese, \vartheta(w) \Vdash q) \tag{3.7.2}$$

$$\mathbb{M}, w \Vdash \circlearrowleft_X \iff \maltese, \vartheta(w) \Vdash \circlearrowleft_X \tag{3.7.3}$$

$$w \sqsubseteq_X^{\mathbb{M}} v \iff \vartheta(w) \sqsubseteq_X^{\maltese} \vartheta(v) \tag{3.7.4}$$

$$w \sqsupseteq_X^{\mathbb{M}} v \iff \vartheta(w) \sqsupseteq_X^{\maltese} \vartheta(v) \tag{3.7.5}$$

$$v \in \ulcorner w \urcorner^{\mathbb{M}} \iff \vartheta(v) \in \ulcorner \vartheta(w) \urcorner^{\maltese} \tag{3.7.6}$$

$$wR_X^{\mathbb{M}} v \iff \vartheta(w) R_X^{\maltese} \vartheta(v) \tag{3.7.7}$$

So in order for $\maltese$ to "solve" the above equations, we have various logical constraints on our definitions, which we have used to determined the design choices we have made. We shall show that $\maltese$ solves the above equations in Lemma 3.7.61.

Before we go ahead and prove results about $\maltese$, we shall try to brush up certain natural questions one may naturally ask about $\maltese$.

- *Why does $\vartheta_1(w)$ encode w's surname but not her full name? That is, why is it that $p_{\ulcorner w \urcorner} \in \vartheta_1(w)$ but $p_w \notin \vartheta_1(w)$?*

Note that in our construction of $\maltese$ we have been trying to enforce that (3.7.4), (3.7.5) and (3.7.6). From definition 2.2.25 from §2.2 we know that if $\vartheta(w) \sqsubseteq_X^{\maltese} \vartheta(v)$ then $\vartheta_1(w) = \vartheta_1(v)$. Hence we must define ϑ_1 in such a manner where if $p_w \in \vartheta_1(w)$ then $p_w \in \vartheta_1(v)$. In fact, since we are enforcing (3.7.6), then we know that we cannot encode any information in $\vartheta_1(w)$ without putting it into $\vartheta_1(v)$ for any $v \in \lceil w \rfloor^{\mathbb{M}}$. However, knowing that we intend to preserve columns in our construction, we may safely encode information about column membership in ϑ_1, as we have done.

- *Why does $\vartheta_2(w)$ encode $\neg p_{\lfloor v \rfloor}$, that is the negation of v's surname, when $\neg w R_X v$, as opposed to her full name?*

  Recall that we want to enforce that (3.7.7). We want to make sure that $\vartheta(w)$ can "see" $\vartheta(v)$ in all and only those situations when it is supposed to. We accomplish this by encoding surname information into $\vartheta_1(v)$, and "blacklisting" certain surnames in $\vartheta_2(w)$ we do not want w to "see" using $R_X^{\maltese}$. Here we are very consciously exploiting the Lemma 3.6.59(*4*), which asserts if one member of an island is not accessible to w then nobody on that island is.

- *Why does $\vartheta_2(w)$ encode the "vacuous" information that $\bot \to p_v$ when $w \sqsupseteq_X v$?* In order to enforce (3.7.4) and (3.7.5), we need to encode information regarding $\sqsubseteq_X^{\mathbb{M}}$ and $\sqsupseteq_X^{\mathbb{M}}$ somewhere. We cannot encode this information in ϑ_1, for the reason that it can only safely encode information at the island level using surnames. Hence we must encode this information in ϑ_2; it is for this reason that we have chosen to include $\bot \to p_v \in \vartheta_{2B}(w, X)$.

  However, we do not want the information we encode in $\vartheta_{2B}(w, X)$ to interfere with $R_X^{\maltese}$, so one way to ensure that "harmless" information is encoded is to use tautologies, as we have done.

Hopefully the reader has some intuition about the engineering choices we made in the construction of $\maltese$. We now turn to proving that $\maltese$ satisfies our design criteria.

Lemma 3.7.61. *Provided that $\mathbb{M}$ is* EviL*, our definition of $\maltese$ suffices* (3.7.2) *through* (3.7.7)*.*

Proof.

- (3.7.2)

$$q \in \Psi \Longrightarrow (\mathbb{M}, w \Vdash q \iff \maltese, \vartheta(w) \Vdash q)$$

  Let $q \in \Psi$. We have two directions we must reason:

  $\Longrightarrow$ First assume that $\mathbb{M}, w \Vdash q$. We know that

$$\begin{aligned} \maltese, \vartheta(w) \Vdash q &\iff q \in \vartheta_1(w) \\ &\iff q \in \{q \in \Psi \mid \mathbb{M}, w \Vdash q\} \cup \{p_{\lfloor w \rfloor}\} \end{aligned}$$

  Hence $\maltese, \vartheta(w) \Vdash q$ as desired.

  $\Longleftarrow$ Assume that $\maltese, \vartheta(w) \Vdash q$, we to show $\mathbb{M}, w \Vdash q$. By our assumption we have either $q \in \{q \in L \mid \mathbb{M}, w \Vdash q\}$ or $q \in \{p_{\lfloor w \rfloor}\}$. In the former case we are done, and the latter

case is impossible since $p_{\ulcorner w \urcorner} \in \Phi \backslash \Psi$ by definition, hence it is impossible for $q \in \{p_{\ulcorner w \urcorner}\}$ by hypothesis.

- (3.7.3)

$$\mathbb{M}, w \Vdash \circlearrowleft_X \iff \maltese, \vartheta(w) \Vdash \circlearrowleft_X$$

Since $\mathbb{M}$ is EviL, and $\maltese, \vartheta(w) \Vdash \circlearrowleft_X$ if and only if $\vartheta(w) R_X^{\maltese} \vartheta(w)$, by virtue of property (VI) of EviL Kripke models it suffices to prove (3.7.7) below.

- (3.7.4)

$$w \sqsubseteq_X^{\mathbb{M}} v \iff \vartheta(w) \sqsubseteq_X^{\maltese} \vartheta(v)$$

We have two directions to show:

$\Longrightarrow$ Assume that $w \sqsubseteq_X^{\mathbb{M}} v$. To ensure $\vartheta(w) \sqsubseteq_X^{\maltese} \vartheta(v)$ we need to ensure two things:

(i) $\vartheta_1(w) = \vartheta_1(v)$ – In order for this to be the case, we must have:

$$\underbrace{\{q \in \Psi \mid \mathbb{M}, w \Vdash q\}}_{A} \cup \underbrace{\{p_{\ulcorner w \urcorner}\}}_{B} = \underbrace{\{q \in \Psi \mid \mathbb{M}, v \Vdash q\}}_{C} \cup \underbrace{\{p_{\ulcorner v \urcorner}\}}_{D}$$

Note that by hypothesis, w and v are on the same island, which means that $B = D$. Since if two worlds in an Evil model are on the same island, then by the Island Lemma they make the same proposition letters true, hence $A = C$, which suffices.

(ii) $(\vartheta_2(w))_X \subseteq (\vartheta_2(v))_X$ – Since $(\vartheta_2(u))_X = \vartheta_{2A}(u, X) \cup \vartheta_{2B}(u, X)$, it suffices to show that $\vartheta_{2A}(w, X) \subseteq \vartheta_{2A}(v, X)$ and $\vartheta_{2B}(w, X) \subseteq \vartheta_{2B}(v, X)$:

- $\vartheta_{2A}(w, X) \subseteq \vartheta_{2A}(v, X)$ – Assume that $x \in \vartheta_{2A}(w, X)$. Then $x = \neg p_{\ulcorner u \urcorner}$ for some $u \in W$ where $\neg w R_X^{\mathbb{M}} u$. It suffices to show that $\neg v R_X^{\mathbb{M}} u$.

  Suppose towards a contradiction that $v R_X^{\mathbb{M}} u$, then by hypothesis we have that $w R_X^{\mathbb{M}} \circ \sqsubseteq_X^{\mathbb{M}} u$. However, we know that since $\mathbb{M}$ is EviL then by (V) we have that $R_X^{\mathbb{M}} \circ \sqsubseteq_X^{\mathbb{M}} \subseteq R_X^{\mathbb{M}}$, which means that $w R_X^{\mathbb{M}} u$ after all. $\lightning$

- $\vartheta_{2B}(w, X) \subseteq \vartheta_{2B}(v, X)$ – Assume that $x \in \vartheta_{2B}(w, X)$, then $x = \bot \to p_u$ for some u such that $u \sqsubseteq_X^{\mathbb{M}} w$. Then by transitivity we have that $u \sqsubseteq_X^{\mathbb{M}} v$, which means that $\bot \to p_u \in \vartheta_{2B}(v, X)$ as desired.

$\Longleftarrow$ Assume that $\vartheta(w) \sqsubseteq_X^{\maltese} \vartheta(v)$. We know that since $\mathbb{M}$ is EviL then $\sqsubseteq_X^{\mathbb{M}}$ is reflexive, so $w \sqsubseteq_X^{\mathbb{M}} w$, whence $\bot \to p_w \in \vartheta_{2B}(w)$. Thus $\bot \to p_w \in (\vartheta_2(v))_X$, which means that either $\bot \to p_w \in \vartheta_{2A}(v, X)$ or $\bot \to p_w \in \vartheta_{2B}(v, X)$. We can see that $\bot \to p_w \neq \neg p_{\ulcorner u \urcorner}$ for all u since these formulae are of different forms, so it must be that $\bot \to p_w \in \vartheta_{2B}(v)$. This means that $w \sqsubseteq_X^{\mathbb{M}} v$, as desired.

- (3.7.5)

$$w \sqsupseteq_X^{\mathbb{M}} v \iff \vartheta(w) \sqsupseteq_X^{\maltese} \vartheta(v)$$

This follows from (3.7.4) and the fact that both $\mathbb{M}$ and $\maltese$ are EviL, hence $x \sqsubseteq_X y \iff y \sqsupseteq_X x$ for both structures.

- (3.7.6)

$$v \in \lceil w \rfloor^{\mathbb{M}} \iff \vartheta(v) \in \lceil \vartheta(w) \rfloor^{\maltese}$$

The fact that islands in both structures correspond follows from the correspondences between $\sqsubseteq_X$ and $\sqsupseteq_X$, as we already saw in (3.7.4) and (3.7.5).

- (3.7.7)

$$wR_X^{\mathbb{M}}v \iff \vartheta(w)R_X^{\maltese}\vartheta(v)$$

$\Longrightarrow$ First assume that $wR_X^{\mathbb{M}}v$, we want to show $\vartheta(w)R_X^{\maltese}\vartheta(v)$. This means that we must show $\vartheta_1(v) \models (\vartheta_2(w))_X$. Since $(\vartheta_2(w))_X = \vartheta_{2A}(w, X) \cup \vartheta_{2B}(w, X)$, we have two steps:

$\vartheta_1(v) \models \vartheta_{2A}(w, X)$ – Assume that $\vartheta_1(v) \not\models \vartheta_{2A}(w, X)$, then it must be that there is some $u \in W^{\mathbb{M}}$ where $p_{\lceil u \rfloor} \in \vartheta_1(u)$ and $\neg p_{\lceil u \rfloor} \in \vartheta_{2A}(w)$, which means that $\neg wR_X^{\mathbb{M}}u$. Since $p_{\lceil u \rfloor} \notin L(\Phi)$ it must be that $p_{\lceil u \rfloor} = p_{\lceil v \rfloor}$, hence $\lceil u \rfloor = \lceil v \rfloor$. Then by the Island Lemma we have $\neg wR_X^{\mathbb{M}}v$ after all. $\lightning$

$\vartheta_1(v) \models \vartheta_{2B}(w, X)$ – Simply note that everything in $\vartheta_{2B}(w, X)$ is a tautology, by construction, so this step follows vacuously.

$\Longleftarrow$ Assume $\vartheta(w)R_X^{\maltese}\vartheta(v)$, in other words $\vartheta_1(v) \models (\vartheta(w))_X$. We shall show $wR_X^{\mathbb{M}}v$. So suppose to the contrary that $\neg wR_X^{\mathbb{M}}v$, then $\neg p_{\lceil v \rfloor} \in \vartheta_{2A}(w)$. However we know that $p_{\lceil v \rfloor} \in \vartheta_1(v)$, hence $\vartheta_1 \models p_{\lceil v \rfloor}$, which means that $\vartheta_1(v) \not\models (\vartheta(w))_X$, which contradicts our assumption. $\lightning$

QED

Having established that $\maltese$ is indeed (almost)-homomorphic to $\mathbb{M}$, we may use this to show that $\mathbb{M}$ and $\maltese$ are logically the same over $\mathcal{L}(\Psi, \mathcal{A})$.

Lemma 3.7.62 (EvIL Translation)**.** *Let $\mathbb{M}$ be an* EvIL *Kripke structure. For any formula $\varphi \in \mathcal{L}(\Psi)$, and any $w \in W$, we have*

$$\mathbb{M}, w \Vdash \varphi \iff \maltese, \vartheta(w) \Vdash \varphi$$

Proof. Using induction, and Lemma 3.7.61, the result follows from the fact that $\maltese$ and $\mathbb{M}$ correspond in all of the ways relevant to $\mathcal{L}(\Psi, \mathcal{A})$. QED

Hence, from the above, we may prove a central result of EvIL:

Theorem 3.7.63 (EvIL Soundness and Weak Completeness)**.**

$$\vdash_{\text{EvIL}} \varphi \iff \Vdash \varphi$$

Proof. Soundness is trivial, so we shall only prove completeness.

Assume that $\nvdash \varphi$. We know from Theorem 3.5.55 that there is a finite $\mathbb{M}$ such that $\mathbb{M}, w \nVdash \varphi$ for some $w \in W$.

Now let $\Psi = L(\varphi)$, where $L(\varphi)$ is the letters that occur in φ, just as we originally defined in the proof of Proposition 1.3.2 from §1.3. Since $\varphi \in \mathcal{L}(L(\varphi), \mathcal{A})$, from lemma 3.7.62 we have that $\bigstar, \vartheta(w) \nVdash \varphi$. Then evidently $\bigstar$ is our desired counter model, hence we have the theorem. QED

With this, we may conclude the proof of completeness of EvIL.

3.8 Taking Stock II

In this section, we take stock of what we have illustrated so far in our investigations into the completeness of EvIL. We discuss how the nature of the abstract semantics for EvIL in relationship to its concrete semantics, and we view this relationship from a wider mathematical perspective.

We shall begin by substantiating the relationship we established in §3.4, which we expressed in (3.4.1).

Lemma 3.8.64. *For* $\Gamma \subseteq_\omega \mathcal{L}(\Phi, \mathcal{A})$ *and infinite* Φ*:*

$$\Gamma \Vdash_{\text{EvIL}} \varphi \iff \Gamma \Vdash \varphi$$

Proof. We may observe that since Γ is finite, then by classical logic and our previous completeness theorems we have the following chain of reasoning:

$$\begin{aligned} \Gamma \Vdash_{\text{EvIL}} \varphi &\iff \Vdash_{\text{EvIL}} \bigwedge \Gamma \to \varphi \\ &\iff \vdash_{\text{EvIL}} \bigwedge \Gamma \to \varphi \\ &\iff \Vdash \bigwedge \Gamma \to \varphi \\ &\iff \Gamma \Vdash \varphi \end{aligned}$$

QED

As a further remark, we feel the need to discuss the nature of the relationship between the *concrete* and *abstract* semantics that EvIL exhibits. We began with EvIL models, which were intended to model intuitions we had regarding the nature of epistemology. In so doing, we used the language of traditional epistemic logic, even though we modified the semantics heavily. We found this gave rise to relational models that are the traditional object of study of modal logic, however we found that while we could abstract to traditional Kripke structures, this was not symmetric – we could not abstract back.

We argue that this particular relationship is commonplace in mathematics. For instance, it is natural to think of the integers as a concrete object. After all, every mathematics student at some

point learns Kronicker's legendary mantra "God created the integers, all else is the creation of man" [Bel86, pg. 477]. However, it is by these concrete origins, we may recognize the integers as a concrete Noetherian ring, and carry on understanding mathematics on a more abstract and general fashion. For instance, the fact that every ideal in a Noetherian ring is equal to a finite intersection of primary ideals is a pure abstraction of Euclid's prime decomposition theorem[AM94, Lemmas 7.11 and 7.12, pg. 83]. This is part of the character of mathematics; abstraction is guided by intuition drawn from more concrete objects. In the same manner we may regard *Stone Representation Theorem* as an infinitary abstraction of *Birkoff's Theorem* [DP02, chapters 11 and 5, respectively], and the *Yoneda Lemma* as an abstraction of Cayley's theorem [SR99, chapters 4 and 1, respectively].

Despite the order of presentation given here, we should make things clear: we did not derive the abstract completeness theorem in §3.2 until we were convinced that EviL Kripke structures generalized our concrete structures. The process by which EviL was developed involved finding the results in §3.5 first, and letting the defining properties of concrete Evil models define the logic. Abstract completeness was an afterthought. Of course, just as in the case of complex analysis and trigonometry, our abstract formalism is far easier to manipulate than our original EviL models. Moreover, since the abstract completeness theorem is far simpler than the concrete completeness theorem, our presentation has followed suit.

EviL Kripke structures really are abstract idealizations of concrete EviL models, as we have illustrated. This puts us in a powerful position. On the one hand, we have concrete semantics by which we may sharpen our intuition. On the other hand, we have well behaved abstract semantics which faithfully provide an idealized domain for us to carry out formal work with relative ease. We feel the situation is analogous to the one in *complex analysis*, where the easiest way to prove trigonometric theorems is to employ complex idealizations such as *Euler's Formula* (the reader will recall this is the assertion $e^{i\theta} = \cos\theta + i\sin\theta$). Perhaps more appropriately, we can think of the space of EviL Kripke structures as a *compactification*, in a sense, of the concrete EviL models, since the logic is indeed compact over the Kripke structure abstraction.

The subsequent sections shall go on to illustrate how we may use abstract Kripke semantics to easily understand properties of EviL, and show that we may use the correspondence exhibited in (3.4.1) to transfer these results to EviL models.

3.9 Subsystems of EviL

In this section, we shall investigate two subsystems of single agent EviL.

We first consider the following two fragments of the main grammar.

Definition 3.9.65. *Define* $\mathcal{L}^{\boxminus}(\Phi)$ *as the fragment:*

$$\varphi ::= p \in \Phi \mid \varphi \to \psi \mid \bot \mid \Box\varphi \mid \boxminus\varphi \mid \circlearrowleft$$

Define $\mathcal{L}^{\boxplus}(\Phi)$ *as the fragment:*

$$\varphi ::= p \in \Phi \mid \varphi \to \psi \mid \bot \mid \Box\varphi \mid \boxplus\varphi \mid \circlearrowleft$$

EvIL$^{\boxminus}$		EvIL$^{\boxplus}$	
(1)	$\vdash \varphi \to \psi \to \varphi$	(1)	$\vdash \varphi \to \psi \to \varphi$
(2)	$\vdash (\varphi \to \psi \to \chi) \to (\varphi \to \psi) \to \varphi \to \chi$	(2)	$\vdash (\varphi \to \psi \to \chi) \to (\varphi \to \psi) \to \varphi \to \chi$
(3)	$\vdash (\neg\varphi \to \neg\psi) \to \psi \to \varphi$	(3)	$\vdash (\neg\varphi \to \neg\psi) \to \psi \to \varphi$
(4)	$\vdash \boxminus\varphi \to \varphi$	(4)	$\vdash \boxplus\varphi \to \varphi$
(5)	$\vdash \boxminus\varphi \to \boxminus\boxminus\varphi$	(5)	$\vdash \boxplus\varphi \to \boxplus\boxplus\varphi$
(6)	$\vdash p \to \boxminus p$	(6)	$\vdash p \to \boxplus p$
(7)	$\vdash \neg p \to \boxminus\neg p$	(7)	$\vdash \neg p \to \boxplus\neg p$
(8)	$\vdash \Diamond\varphi \to \boxminus\Diamond\varphi$	(8)	$\vdash \Box\varphi \to \boxplus\Box\varphi$
(9)	$\vdash \Box\varphi \to \Box\boxminus\varphi$	(9)	$\vdash \Box\varphi \to \Box\boxplus\varphi$
(10)	$\vdash \varphi \to \boxminus(\circlearrowleft\!\!\to \Diamond\varphi)$	(10)	$\vdash \varphi \to \boxplus(\circlearrowleft\!\!\to \Diamond\varphi)$
(11)	$\vdash \circlearrowleft\!\!\to \boxminus\circlearrowleft$	(11)	$\vdash \neg\circlearrowleft\!\!\to \boxplus\neg\circlearrowleft$
(12)	$\vdash \Box(\varphi \to \psi) \to \Box\varphi \to \Box\psi$	(12)	$\vdash \Box(\varphi \to \psi) \to \Box\varphi \to \Box\psi$
(13)	$\vdash \boxminus(\varphi \to \psi) \to \boxminus\varphi \to \boxminus\psi$	(13)	$\vdash \boxplus(\varphi \to \psi) \to \boxplus\varphi \to \boxplus\psi$
(I)	$\dfrac{\vdash \varphi \to \psi \qquad \vdash \varphi}{\vdash \psi}$	(I)	$\dfrac{\vdash \varphi \to \psi \qquad \vdash \varphi}{\vdash \psi}$
(II)	$\dfrac{\vdash \varphi}{\vdash \Box\varphi}$	(II)	$\dfrac{\vdash \varphi}{\vdash \Box\varphi}$
(III)	$\dfrac{\vdash \varphi}{\vdash \boxminus\varphi}$	(III)	$\dfrac{\vdash \varphi}{\vdash \boxplus\varphi}$

Table 3.2: Axiom systems EvIL$^{\boxminus}$ and EvIL$^{\boxplus}$ respectively

It is natural to wonder about EvIL restricted to thee two fragments. After all, while ideas from Cartesian skepticism naturally leads one to think about $\mathcal{L}^{\boxminus}(\Phi)$, as we saw in §1.5, it is harder to motivate $\mathcal{L}^{\boxplus}(\Phi)$. However, we regard each fragment as worthy of study in its own right.

Table 3.2 gives the axioms systems for the two fragments in question. The system corresponding to the $\mathcal{L}^{\boxminus}$ fragment is referred to as EvIL$^{\boxminus}$, and similarly the fragment corresponding to the $\mathcal{L}^{\boxplus}$ fragment is referred to as EvIL$^{\boxplus}$.

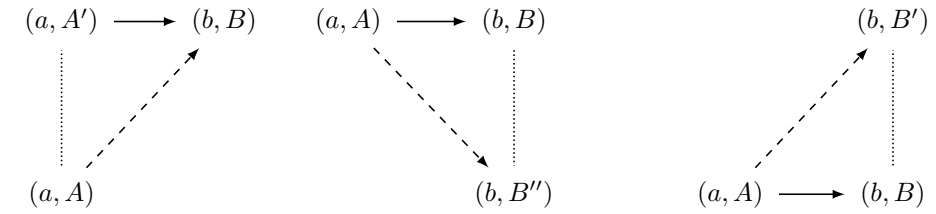
From these axioms, we shall define two sorts of EvIL models the correspond to the properties defined by the above axiom systems.

Definition 3.9.66. *The following properties specify* ⊟**EviL** *and* ⊞**EviL** *Kripke structures:*

	⊟**EviL**	⊞**EviL**
(I)	$\sqsupseteq$ *is reflexive*	$\sqsubseteq$ *is reflexive*
(II)	$\sqsupseteq$ *is transitive*	$\sqsubseteq$ *is transitive*
(III)	$w \sqsupseteq v$ *if and only if* $v \sqsubseteq w$	$w \sqsubseteq v$ *if and only if* $v \sqsupseteq w$
(IV)	*If* $w \sqsupseteq v$ *then* $(w \in V(p)$ *if and only if* $v \in V(p))$	*If* $w \sqsubseteq v$ *then* $(w \in V(p)$ *if and only if* $v \in V(p))$
(V)	$(R \circ \sqsubseteq) \subseteq R$	$(R \circ \sqsubseteq) \subseteq R$
(VI)	$(\sqsupseteq \circ R) \subseteq R$	$(\sqsubseteq \circ R) \subseteq R$
(VII)	*If* $w \sqsupseteq v$ *and* $v \in P$ *then* vRw	*If* $w \sqsubseteq v$ *and* $v \in P$ *then* vRw
$(VIII)$	*If* $w \in P$ *and* $w \sqsupseteq v$ *then* $v \in P$	*If* $w \notin P$ *and* $w \sqsubseteq v$ *then* $v \notin P$

Exactly as in the case of the EviL Kripke structures we introduced in §2.2, we may naturally visualize certain properties in commutative diagrams:

- Properties $(V)^{\boxminus}$ and $(V)^{\boxplus}$ are depicted as Fig. 3.5(a), which is the same as Fig. 2.2(a)
- Property $(VI)^{\boxminus}$ is depicted as Fig. 3.5(b), which is the same as Fig. 2.2(b)
- Property $(VI)^{\boxplus}$ is depicted as Fig. 3.5(c), which is the same as Fig. 2.2(c)



(a) – $(VI)^{\boxminus}$ & $(VI)^{\boxplus}$ – (b) – $(VI)^{\boxminus}$ – (c) – $(VI)^{\boxplus}$ –

Figure 3.5: Visualizations of the relationships in Proposition 2.2.27

We may recall that the commutative diagrams depicted split the original EviL property (V); this is not coincidental. By elementary reasoning we may observe that every partly EviL Kripke structure (and hence, every EviL Kripke structure) is both ⊟EviL and ⊞EviL. In fact, the logical differences between partly EviL, ⊟EviL, and ⊞EviL properties respectively may be summarized as follows:

- ⊟EviL and ⊞EviL Kripke structures *strengthen* $(VIII)'$ to $(VII)^{\boxminus}$ and $(VII)^{\boxplus}$. Note that in the presence of the other properties of ⊟EviL and ⊞EviL Kripke structures, we may observe that $(VII)^{\boxminus}$ and $(VII)^{\boxplus}$ are logically equivalent.

- ⊟EvIL and ⊞EvIL Kripke structures *weaken* $(VII)'$ to $(VI)^{\boxminus}$ and $(VI)^{\boxplus}$, respectively.
- With the exception of $(VI)^{\boxminus}$ and $(VI)^{\boxplus}$, the ⊟EvIL properties are logically equivalent to the ⊞EvIL properties.

Hence, just as the proof of abstract completeness of EvIL involved producing EvIL bisimilar completions of partly Evil Kripke structures using the operator $\mathfrak{S}$, the proof of the abstract completeness of EvIL$^{\boxminus}$ and EvIL$^{\boxplus}$ shall involve producing partly EvIL bisimilar completions

Before turning to bisimulation, we shall first prove abstract completeness for EvIL$^{\boxminus}$ and EvIL$^{\boxplus}$ and their respective classes of Kripke structures.

Definition 3.9.67. *We shall write*

$$\Gamma \Vdash_{\boxminus\text{EvIL}} \varphi$$

to mean that for all ⊟EvIL *Kripke structures* $\mathbb{M} = \langle W, R, \sqsubseteq, \sqsupseteq, V, P\rangle$*, for all worlds* $w \in W$ *if* $\mathbb{M}, w \Vdash \Gamma$ *then* $\mathbb{M}, w \Vdash \varphi$*.*

Moreover, we shall write

$$\Gamma \Vdash_{\boxplus\text{EvIL}} \varphi$$

to mean the same for all ⊞EvIL *Kripke structures.*

Theorem 3.9.68 (⊟/⊞EvIL Strong Soundness and Completeness)**.**

$$\Gamma \vdash_{\text{EvIL}^{\boxminus}} \varphi \text{ if and only if } \Gamma \Vdash_{\boxminus\text{EvIL}} \varphi$$
$$\&$$
$$\Gamma \vdash_{\text{EvIL}^{\boxplus}} \varphi \text{ if and only if } \Gamma \Vdash_{\boxplus\text{EvIL}} \varphi$$

Proof. The proof of these two propositions, in each case, proceeds exactly as in the proof of Theorem 3.2.35 from §3.2. In each case we perform the usual canonical model construction that is used in modal logic. Rather than rehash that proof, here we simply list how we may infer the desired properties we attribute to these canonical models. At the risk of being slightly redundant, we have chosen to present the arguments for both logics, even though they are highly symmetric:

⊟EviL	⊞EviL
$(I)^{\boxminus}$ corresponds to the EvIL$^{\boxminus}$ axiom (4)	$(I)^{\boxplus}$ corresponds to the EvIL$^{\boxplus}$ axiom (4)
$(II)^{\boxminus}$ corresponds to the EvIL$^{\boxminus}$ axiom (5)	$(II)^{\boxplus}$ corresponds to the EvIL$^{\boxplus}$ axiom (5)
$(III)^{\boxminus}$ – Since the canonical model construction in this case does not specify $\sqsubseteq$, we shall define $w \sqsubseteq v$ if and only if $v \sqsupseteq w$	$(III)^{\boxplus}$ – Since the canonical model construction in this case does not specify $\sqsupseteq$, we shall define $w \sqsupseteq v$ if and only if $v \sqsubseteq w$
$(IV)^{\boxminus}$ corresponds to the EvIL$^{\boxminus}$ axioms (6) and (7)	$(IV)^{\boxplus}$ corresponds to the EvIL$^{\boxplus}$ axioms (6) and (7)
$(V)^{\boxminus}$ corresponds to the EvIL$^{\boxminus}$ axiom (8)	$(V)^{\boxplus}$ corresponds to the EvIL$^{\boxplus}$ axiom (8)
$(VIII)^{\boxminus}$ corresponds to the EvIL$^{\boxminus}$ axiom (11)	$(VIII)^{\boxplus}$ corresponds to the EvIL$^{\boxplus}$ axiom (11)
$(VI)^{\boxminus}$ corresponds to the EvIL$^{\boxminus}$ axiom (9)	$(VI)^{\boxplus}$ corresponds to the EvIL$^{\boxplus}$ axiom (9)
$(VII)^{\boxminus}$ corresponds to the EvIL$^{\boxminus}$ axiom (10)	$(VII)^{\boxplus}$ corresponds to the EvIL$^{\boxplus}$ axiom (10)
	QED

With the previous completeness theorem, we shall give two constructions which provide bisimular partly EvIL completions of both ⊟EvIL and ⊞EvIL Kripke structures.

Definition 3.9.69 ($\ominus$ and $\oplus$ Bisimulators)**.** *Let $\mathbb{M}$ be a Kripke model, then define:*

$$\ominus^{\mathbb{M}} := \langle W^{\ominus}, R^{\ominus}, \sqsubseteq^{\ominus}, \sqsupseteq^{\ominus}, V^{\ominus}, P^{\ominus} \rangle$$

where:

$$\begin{aligned}
W^{\ominus} &:= \sqsupseteq^{\mathbb{M}} \\
V^{\ominus}(p) &:= \{(w,v) \in W^{\ominus} \mid v \in V^{\mathbb{M}}\} \\
P^{\ominus} &:= \{(w,v) \in W^{\ominus} \mid v \in V^{\mathbb{M}}\} \\
R^{\ominus} &:= \{((w,v),(t,u)) \in (W^{\ominus})^2 \mid vR^{\mathbb{M}}t \ \& \ vR^{\mathbb{M}}u\} \\
\sqsupseteq^{\ominus} &:= \{((w,v),(w,u)) \in (W^{\ominus})^2 \mid v \sqsupseteq^{\mathbb{M}} u\} \\
\sqsubseteq^{\ominus} &:= \{((w,v),(w,u)) \in (W^{\ominus})^2 \mid v \sqsubseteq^{\mathbb{M}} u\}
\end{aligned}$$

$$\oplus^{\mathbb{M}} := \langle W^{\oplus}, R^{\oplus}, \sqsubseteq^{\oplus}, \sqsupseteq^{\oplus}, V^{\oplus}, P^{\oplus} \rangle$$

where:

$$\begin{aligned}
W^{\oplus} &:= \sqsubseteq^{\mathbb{M}} \\
V^{\oplus}(p) &:= \{(w,v) \in W^{\oplus} \mid v \in V^{\mathbb{M}}\} \\
P^{\oplus} &:= \{(w,v) \in W^{\oplus} \mid v \in V^{\mathbb{M}}\} \\
R^{\oplus} &:= \{((w,v),(t,u)) \in (W^{\oplus})^2 \mid vR^{\mathbb{M}}t \ \& \ vR^{\mathbb{M}}u\} \\
\sqsupseteq^{\oplus} &:= \{((w,v),(w,u)) \in (W^{\oplus})^2 \mid v \sqsupseteq^{\mathbb{M}} u\} \\
\sqsubseteq^{\oplus} &:= \{((w,v),(w,u)) \in (W^{\oplus})^2 \mid v \sqsubseteq^{\mathbb{M}} u\}
\end{aligned}$$

Our intuition for the above construction comes from the following proposition:

Definition 3.9.70. *Let $\downarrow w := \{v \in W \mid w \sqsupseteq v\}$, which is the **downset** of w.*

*Let $\uparrow w := \{v \in W \mid w \sqsupseteq v\}$, which is the **upset** of w.*

Lemma 3.9.71 (Downset/Upset Lemma)**.**

Let $\mathbb{M}$ *be a partly* EviL$^{\boxminus}$ *Kripke structure. We have:*

$(1)^{\boxminus}$ *For all* w, $w \in \downarrow w$

$(2)^{\boxminus}$ wRv *if and only if* $wR \downarrow v$

$(3)^{\boxminus}$ *if* $w \in \downarrow v$ *then* $w \in V(p)$ *if and only if* $v \in V(p)$

$(4)^{\boxminus}$ $v \sqsupseteq w$ *and* $w \in P$ *implies* $wR \downarrow v$

Let $\mathbb{M}$ *be a partly* EviL$^{\boxplus}$ *Kripke structure. We have:*

$(1)^{\boxplus}$ *For all* w, $w \in \uparrow w$

$(2)^{\boxplus}$ wRv *if and only if* $wR \uparrow v$

$(3)^{\boxplus}$ *if* $w \in \uparrow v$ *then* $w \in V(p)$ *if and only if* $v \in V(p)$

$(4)^{\boxplus}$ $v \sqsupseteq w$ *and* $w \in P$ *implies* $wR \uparrow v$

Proof.

$(1)^{\boxminus}$ Follows from property $(I)^{\boxminus}$

$(2)^{\boxminus}$ Follows from properties $(I)^{\boxminus}$ and $(VI)^{\boxminus}$

$(3)^{\boxminus}$ Follows from property $(IV)^{\boxminus}$

$(4)^{\boxminus}$ Follows from properties $(VI)^{\boxminus}$ and $(VII)^{\boxminus}$

$(1)^{\boxplus}$ Follows from property $(I)^{\boxplus}$

$(2)^{\boxplus}$ Follows from properties $(I)^{\boxplus}$ and $(VI)^{\boxplus}$

$(3)^{\boxplus}$ Follows from property $(IV)^{\boxplus}$

$(4)^{\boxplus}$ Follows from properties $(VI)^{\boxplus}$ and $(VII)^{\boxplus}$

QED

We now can state the central idea behind $\ominus$ and $\oplus$. Essentially, in $\ominus$ and $\oplus$ we shall force the islands of each of the structures we construct to correspond to the downsets and upsets of the original $\boxminus$EviL and $\boxplus$EviL structures, respectively. We note that in many respects, Lemma 3.9.71 illustrates that downsets and upsets are similar in many respects to islands, just as we saw in Lemma 3.6.59, the Island Lemma, from §3.6. In $\ominus$, each world has as its first coordinate which downset it belongs to, and similarly for $\oplus$. As a consequence, $\sqsupseteq$ and $\sqsubseteq$ end up being the set of worlds in each of these constructions.

To see that $\ominus$ and $\oplus$ are the completions we want, we shall see that $\oplus$ and $\ominus$ give rise to special bisimulations, which each preserve the formulae of $\mathcal{L}^{\boxminus}(\Phi)$ and $\mathcal{L}^{\boxplus}(\Phi)$. We will next see that if $\mathbb{M}$ is $\boxminus$EviL then $\ominus$ is partly EviL, and likewise if $\mathbb{M}$ is $\boxplus$EviL then $\ominus$ is partly EviL. Since in the case of the logics EviL$^{\boxminus}$ and EviL$^{\boxplus}$ we are only concerned with fragments of the main language, the limited bisimulations we will deploy will be sufficient for our purposes.

Definition 3.9.72. *A relation* Z *is called a* $\boxminus$*-bisimulation between* $\mathbb{M}$ *and* $\mathbb{M}'$*, with type annotation* $Z : \mathbb{M} \leftrightarrows^{\boxminus} \mathbb{M}'$*, if it satisfies the letter conditions, as well as the back and forth conditions for* R *and* $\sqsupseteq$*, but not necessarily* $\sqsubseteq$*.*

A Z *relation is called a* $\boxplus$*-bisimulation is the same, with type annotation* $Z : \mathbb{M} \leftrightarrows^{\boxplus} \mathbb{M}'$*, only in this case weenforce the back and forth conditions for* R *and* $\sqsubseteq$*, but not necessarily* $\sqsupseteq$*.*

The following is a trivial consequence of *The Fundamental Theorem of Bisimulations*, Theorem 3.3.37, that we gave in 3.3:

Proposition 3.9.73. *If* $Z : \mathbb{M} \leftrightarrows^{\boxminus} \mathbb{M}'$*, then if* $\varphi \in \mathcal{L}^{\boxminus}(\Phi)$ *and* wZv *then* $\mathbb{M}, w \Vdash \varphi \iff \mathbb{M}', v \Vdash \varphi$*.*

If $Z : \mathbb{M} \leftrightarrows^{\boxplus} \mathbb{M}'$*, then if* $\varphi \in \mathcal{L}^{\boxplus}(\Phi)$ *and* wZv *then* $\mathbb{M}, w \Vdash \varphi \iff \mathbb{M}', v \Vdash \varphi$.

As asserted, $\ominus$ and $\oplus$ each give rise to $\boxminus$-bisimilar and $\boxplus$-bisimilar Kripke structures, respectively:

Lemma 3.9.74. *Let* $\mathbb{M}$ *be a Kripke structure.*

If $\sqsupseteq$ *is reflexive and transitive, then* $Z : \mathbb{M} \leftrightarrows^{\boxminus} \ominus^{\mathbb{M}}$ *where* $wZ(u, w)$ *for all* $u \in \mathbb{M}$

In the same vain, if $\sqsupseteq$ *is reflexive and transitive, then again* $Z : \mathbb{M} \leftrightarrows^{\boxplus} \ominus^{\mathbb{M}}$ *where* $wZ(u, w)$ *for all* $u \in \mathbb{M}$

Proof. The two proofs are entirely analogous, so we shall only prove for $\ominus$.

- Proposition letters and P: Simply note that we defined

  $$(u, w) \in V^{\ominus}(p) \iff w \in V^{\mathbb{M}}(p)$$

  And similarly for P

- R forth: Assume that $wR^{\mathbb{M}}v$ and $wZ(u, w)$. Since we are assume that $\sqsubseteq$ is reflexive then we know that $(v, v) \in W^{\ominus}$, hence $vZ(v, v)$ and $(u, w)R^{\mathbb{M}}(v, v)$ as per definition.

- R back: Assume that $(u, w)R^{\ominus}(t, v)$ and $wZ(u, w)$; by construction it must be that $wR^{\mathbb{M}}v$.

- $\sqsupseteq$ forth: Assume that $w \sqsupseteq^{\mathbb{M}} v$ and $wZ(u, w)$. It must then be that $u \sqsupseteq^{\mathbb{M}} w$ by construction, whence by transitivity we know that $u \sqsupseteq^{\mathbb{M}} v$ and thus $(u, v) \in W^{\ominus}$. So $vZ(u, v)$ and moreover $(u, w) \sqsupseteq^{\ominus} (u, v)$ by construction, which suffices.

- $\sqsupseteq$ back: As with the case for R back, this follows by construction.

QED

We now show that $\ominus$ and $\oplus$ really are adequate completions:

Lemma 3.9.75 ($\boxminus/\boxplus$ Completion)**.**
If $\mathbb{M}$ *is* $\boxminus$Evil*, then* $\ominus$ *is partly* Evil

Likewise, if $\mathbb{M}$ *is* $\boxplus$Evil*, then* $\oplus$ *is partly* Evil

Proof. Since the proof involves many steps, we shall economize on space by only proving for $\ominus$, since $\oplus$ is similar.

$(I)'$ asserts $\sqsubseteq^{\ominus}$ is reflexive. This follows by construction since $\sqsubseteq^{\mathbb{M}}$ is reflexive, as per $(I)^{\boxminus}$ and $(III)^{\boxminus}$.

$(II)'$ asserts $\sqsubseteq^{\ominus}$ is transitive. This follows by construction since $\sqsubseteq^{\mathbb{M}}$ is transitive, as per $(II)^{\boxminus}$ and $(III)^{\boxminus}$.

$(III)'$ follows from the fact that $\mathbb{M}$ makes true $(II)^{\boxminus}$, hence $\sqsubseteq^{\ominus}$ is transitive by construction, as desired.

$(IV)'$ asserts that $\sqsupseteq^{\ominus}$ is the reverse of $\sqsubseteq^{\ominus}$, which follows directly by construction

$(V)'$ asserts that "if $w \sqsubseteq^{\ominus} v$ then $(w \in V(p)$ if and only if $v \in V(p))$." This follows by construction and $(IV)^{\boxminus}$ for $\mathbb{M}$.

$(VI)'$ asserts

$$(R^{\ominus} \circ \sqsubseteq^{\ominus}) \subseteq R^{\ominus}.$$

As above, $\ominus$ inherits this property from $\mathbb{M}$, which obeys $(V)^{\boxminus}$.

$(VII)'$ asserts

$$(\sqsubseteq^{\ominus} \circ R) \subseteq R^{\ominus}$$
$$\&$$
$$(\sqsupseteq^{\ominus} \circ R^{\ominus}) \subseteq R^{\ominus}.$$

Since $\mathbb{M}$ obeys $(VI)^{\boxminus}$, we have that

$$(\sqsupseteq^{\ominus} \circ R^{\ominus}) \subseteq R^{\ominus}$$

So it suffices to show that if $(s,w)R^{\ominus}(t,u) \sqsupseteq^{\ominus} (t,v)$ then $wR^{\ominus}v$. We can see by the definition of $\ominus$ that all we have to show is that $wR^{\mathbb{M}}v$. With our assumptions, we can see by construction that we know the following two things:

$$wR^{\mathbb{M}}t$$
$$\&$$
$$v \in\downarrow t$$

However, we know from Lemma 3.9.71, the Downset/Upset Lemma, that if $wR^{\mathbb{M}}t$ then $wR^{\mathbb{M}} \downarrow t$, which means that $wR^{\mathbb{M}}v$ as desired.

$(VIII)'$ We must show "If $(v,w) \in P^{\ominus}$ then $(v,w)R^{\ominus}(v,w)$." So assume that $(v,w) \in P^{\ominus}$. By construction we must show that $wR^{\mathbb{M}}v$ and $wR^{\mathbb{M}}w$

Since $(v,w) \in W^{\ominus}$ then by construction we know $w \sqsubseteq^{\mathbb{M}} v$. Since $(v,w) \in P^{\ominus}$, then evidently $w \in P^{\mathbb{M}}$. So by $(VII)^{\boxminus}$ we know that $wR^{\mathbb{M}}v$.

We may also note by $(I)^{\boxminus}$ that $w \sqsubseteq w$, so applying our assumptions and $(VII)^{\boxminus}$ again we have that $wR^{\mathbb{M}}w$, which suffices.

$(IX)'$ The last thing to show is "If $w \in P_X^{\ominus}$ and $w \sqsupseteq_X^{\ominus} v$ then $v \in P_X^{\ominus}$." However, this follows immediately by construction from $(VIII)^{\boxminus}$.

QED

With the above, we have what we need to prove the central result regarding these subsystems:

Theorem 3.9.76 (⊟/⊞ EvIL Soundness and Completenesses)**.**
Assume $\Gamma \cup \{\varphi\} \subseteq \mathcal{L}^{\boxminus}(\Phi, \mathcal{A})$. *The following are equivalent:*

$$\begin{aligned} \Gamma \vdash_{\text{EvIL}^{\boxminus}} \varphi &\iff \Gamma \Vdash_{\boxminus\text{EvIL}} \varphi \\ &\iff \Gamma \Vdash_{p\text{EvIL}} \varphi \\ &\iff \Gamma \Vdash_{\text{EvIL}} \varphi \\ &\iff \Gamma \vdash_{\text{EvIL}} \varphi \end{aligned}$$

Assume $\Gamma \cup \{\varphi\} \subseteq \mathcal{L}^{\boxplus}(\Phi, \mathcal{A})$. *The following are equivalent:*

$$\begin{aligned} \Gamma \vdash_{\text{EvIL}^{\boxplus}} \varphi &\iff \Gamma \Vdash_{\boxplus\text{EvIL}} \varphi \\ &\iff \Gamma \Vdash_{p\text{EvIL}} \varphi \\ &\iff \Gamma \Vdash_{\text{EvIL}} \varphi \\ &\iff \Gamma \vdash_{\text{EvIL}} \varphi \end{aligned}$$

Proof. As before, we only show the result for $\text{EvIL}^{\boxminus}$.

The only nontrivial equality is $\Gamma \vdash_{\text{EvIL}^{\boxminus}} \varphi \iff \Gamma \Vdash_{p\text{EvIL}} \varphi$. It is straightforward to verify that $\Gamma \vdash_{\text{EvIL}^{\boxminus}} \varphi \Longrightarrow \Gamma \Vdash_{p\text{EvIL}} \varphi$, since this is just a soundness result. To show the converse, we shall establish the contrapositive. Assume that $\Gamma \nvdash_{\text{EvIL}^{\boxminus}} \varphi$. By Theorem 3.9.68, there is some ⊟EvIL Kripke structure $\mathbb{M}$ with a world w such that $\mathbb{M}, w \Vdash \Gamma$ and $\mathbb{M}, w \nVdash \varphi$. Since $\mathbb{M} \leftrightarrows^{\boxminus} \ominus^{\mathbb{M}}$, then by our grammar restriction assumptions, Proposition 3.9.73 and Lemma 3.9.74 we have $\ominus^{\mathbb{M}}, (w, w) \Vdash \Gamma$ and $\ominus^{\mathbb{M}}, (w, w) \nVdash \varphi$. By 3.9.75 we know that $\ominus$ is partly EvIL. which suffices our claim.

Hence we have:

$$\begin{aligned} \Gamma \vdash_{\text{EvIL}^{\boxminus}} \varphi &\iff \Gamma \Vdash_{\boxminus\text{EvIL}} \varphi && \text{by Theorem 3.9.68} \\ &\iff \Gamma \Vdash_{p\text{EvIL}} \varphi && \text{by our reasoning above} \\ &\iff \Gamma \vdash_{\text{EvIL}} \varphi && \text{by Theorem 3.2.35 from §3.2} \\ &\iff \Gamma \Vdash_{\text{EvIL}} \varphi && \text{by Theorem 3.3.44 from §3.3} \end{aligned}$$

QED

With the above abstract completeness result, by the reasoning we provided in §3.8, we can employ our abstract semantics to avoid dealing directly with the EvIL concrete semantics, and easily arrive at couple lemmas:

Lemma 3.9.77 (⊟/⊞ Weak Soundness and Completeness)**.**
If $\Gamma \subseteq_{\omega} \mathcal{L}^{\boxminus}(\Phi, \mathcal{A})$ *is finite, then*

$$\Gamma \vdash_{\text{EvIL}^{\boxminus}} \varphi \iff \Gamma \Vdash \varphi$$

Likewise, if $\Gamma \subseteq_{\omega} \mathcal{L}^{\boxplus}(\Phi, \mathcal{A})$ *is finite, then*

$$\Gamma \vdash_{\text{EvIL}^{\boxplus}} \varphi \iff \Gamma \Vdash \varphi$$

Proof. This follows by Theorem 3.9.76 above and Lemma 3.8.64 in §taking-stockII. QED

Theorem 3.9.78 (⊟/⊞ Small Model Property)**.** *If* φ *is* ⊟EvIL *or* ⊞EvIL *satisfiable, then it is finitely satisfiable in a* EvIL *Kripke structure with* $O(EXP2(|\varphi|))$ *worlds.*

Proof. This follows by Theorem 3.9.76 above and our previous small model property we established in Theorem 3.5.56 in §3.5. QED

Theorem 3.9.79 (⊟/⊞ Decidability)**.** *The decision problems for* $\vdash_{\text{EvIL}^{\boxminus}} \varphi$ *and* $\vdash_{\text{EvIL}^{\boxplus}} \varphi$ *have time complexity bounded above by* $O(EXP3(|\varphi|))$*.*

Proof. As before, this is an application of Theorem 3.9.76 and Theorem 3.5.57. QED

As a final note, we should mention why these results do not generalize to more than one agent. The central issue is precisely that the observation that downsets and upsets can play the role of islands, as we asserted in Lemma 3.9.71, the Downset/Upset Lemma, do not generalize to multiple agents.

The central issue is that intuitions about how $\circlearrowleft$ functions in the single agent case do not extend to multiple agents. Specifically, consider the plausible extension of EvIL$^{\boxminus}$ and EvIL$^{\boxplus}$ with the following two axioms from ordinary multi-agent EvIL:

$$\vdash \Box_X \varphi \rightarrow \Box_X \boxminus_Y \varphi$$
$$\&$$
$$\vdash \Box_X \varphi \rightarrow \Box_X \boxplus_Y \varphi$$

We may then plausibly extend our notion of downset and upset

$$\downarrow^* w := \left\{ v \in W \;\middle|\; w \left(\bigcup_{X \in \mathcal{A}} \sqsupseteq_X \right)^* v \right\}$$
$$\uparrow^* w := \left\{ v \in W \;\middle|\; w \left(\bigcup_{X \in \mathcal{A}} \sqsubseteq_X \right)^* v \right\}$$

With this, we would indeed get that "if wR_Xv then $wR_X \downarrow^* v$" for ⊟EvIL Kripke structures and the dual for ⊞EvIL Kripke structures.

We stumble when we try to enforce "if $w \in \downarrow^* v$ and $w \in P_X$ then $wR_X \downarrow^* v$." Somehow, we would need to enforce that if there is any path from v to w, and w makes true $\circlearrowleft_X$, then w can see v. One way to do this would be to introduce special shorthand for *vectors* of agents, where

$$\vec{v} = \langle Y_1, Y_2, \ldots Y_n \rangle$$

For $\{Y_1, Y_2, \ldots, Y_n\} \subseteq \mathcal{A}$. Along with this would could employ a shorthand for vector modalities:

$$\boxminus_{\vec{v}} \varphi := \boxminus_{Y_1} \boxminus_{Y_2} \cdots \boxminus_{Y_n} \varphi$$
$$\boxplus_{\vec{v}} \varphi := \boxplus_{Y_1} \boxplus_{Y_2} \cdots \boxplus_{Y_n} \varphi$$

In each system we could then postulate *path* axioms, corresponding to every vector of agents:

$$\vdash \varphi \rightarrow \boxminus_{\vec{v}} (\circlearrowleft_X \rightarrow \Diamond_X \varphi)$$
$$\&$$
$$\vdash \Box_X \varphi \rightarrow \boxminus_{\vec{v}} (\circlearrowleft_X \rightarrow \Diamond_X \varphi)$$

We state without proof that the details sketched here could be formalized to give multi-agent fragments of EvиL. On the other hand, we observe that while single agent intuition generalizes to multiple agents in a straightforward manner for the main EvиL calculus, this is not so for the fragments of the calculus. It is for lack of logical parsimony that we have decided to restrict ourselves to single agent EvиL for the fragments presented here.

3.10 Universal Modality

In this section, we show how EvиL may be extended with a universal modality U, as presented in [vB10, chapter 7, pg 79]. Just as the previous section illustrated that looking at fragments of EvиL added complexity to the completeness theorem, so too do natural extensions to the calculus. We shall mention the relationship of the universal modality to traditional epistemic logic, and discuss an analogue of the Theorem Theorem that the universal modality obeys.

In this section, we shall provide sketches rather than the more extensive proofs as we have so far provided. This is because we really intend for the results in this section to be minor modifications of our previous results. Our intention is to indicate what modifications are to be made to accommodate our proposed extension.

The following gives the extended grammar of EvиL with an added universal modality: $\mathcal{L}^U(\Phi, \mathcal{A})$ is the fragment:

$$\varphi ::= p \in \Phi \mid \varphi \to \psi \mid \bot \mid \Box_X \varphi \mid \boxminus_X \varphi \ \boxplus_X \varphi \mid U\varphi \mid \circlearrowleft$$

It is important to note that our other fragments may be similarly extended.

Universal modality has the following semantics for Kripke structures:

$$\mathbb{M}, w \Vdash U\varphi \iff \mathbb{M}, v \Vdash \varphi \text{ for all } v \in W$$

Likewise, it has corresponding semantics for EvиL models:

$$\mathfrak{M}, w \Vdash U\varphi \iff \mathfrak{M}, v \Vdash \varphi \text{ for all } v \in \mathfrak{M}$$

Universal modality has its own associated analogue of the Theorem Theorem, which while rather trivial, nonetheless allows us to understand the connection of semantics of $U\varphi$ to the logic of EvиL:

Proposition 3.10.80.

$$\mathfrak{M}, (a, A) \Vdash U\varphi \iff Th(\mathfrak{M}) \vdash \varphi$$

Note that since $Th(\mathfrak{M})$ is necessarily closed under deduction, then the above means that $\varphi \in Th(\mathfrak{M})$.

Compare this with the original Theorem Theorem:

$$\mathfrak{M}, (a, A) \Vdash \Box_X \varphi \iff Th(\mathfrak{M}) \cup A_X \vdash \varphi$$

Previously, the background knowledge $Th(\mathfrak{M})$ was implicit in the Theorem Theorem reading of the semantics of $\Box_X \varphi$; we can see now that U makes these semantics explicit.

(U1)	$\vdash U(\varphi \to \psi) \to U\varphi \to U\psi$
(U2)	$\vdash U\varphi \to \varphi$
(U3)	$\vdash U\varphi \to UU\varphi$
(U4)	$\vdash \neg U\varphi \to U\neg U\varphi$
(U5)	$\vdash U\varphi \to \Box_X \varphi$
(U6)	$\vdash U\varphi \to \boxminus_X \varphi$
(U7)	$\vdash U\varphi \to \boxplus_X \varphi$
(I)	$\dfrac{\vdash \varphi}{\vdash U\varphi}$

Table 3.3: Additional axiom and rules for the universal modality

As mentioned in [vB10, Chapter 7.4], universal modality is closely related to the modal logic $S5$. Below, we state the axioms for the univesal modality in EvIL, which are recognizably the axioms for $S5$, along with other axioms asserting that the other relations are subrelations of U.

We can think of the universal modality axioms (appropriately restricted) as extending any of the three systems we have looked at so far; EvIL extends to UEvIL, EvIL$^{\boxminus}$ extends to UEvIL$^{\boxminus}$, and EvIL$^{\boxplus}$ extends to UEvIL$^{\boxplus}$. Abstract completeness for all three systems is achieved in a similar manner.

Theorem 3.10.81 (Universal EvIL Soundness and Completeness)**.**

$$\Gamma \vdash_{U\text{EvIL}} \varphi \iff \Gamma \Vdash_{\text{EvIL}} \varphi$$
$$\Gamma \vdash_{U\text{EvIL}^{\boxminus}} \varphi \iff \Gamma \Vdash_{\text{EvIL}} \varphi$$
$$\Gamma \vdash_{U\text{EvIL}^{\boxplus}} \varphi \iff \Gamma \Vdash_{\text{EvIL}} \varphi$$

Proof. Soundness in all cases is straightforward.

For completeness, in each case we carry out the canonical model construction. Note that axioms 3.10–3.10 enforce that the accessibility relation associated with U forms a partition on the canonical model, and at every point within a given partition. Axioms 5–7 ensure the other relations are a subrelation of the candidate universal relation. In each case the canonical model construction will provide a world w which witnesses Γ but does not witness φ. To complete the construction, one need only take a *point generated submodel* around w; see [BRV01, chapter 2, pg. 210] for a discussion on how "A point generated submodel suffices." This construction preserves the truth of all formulae at w but establishes U as a universal modality.

From there, it is straightforward to verify that all of the bisimilar model completions we have investigated preserve the truth of $U\varphi$ and $\neg U\varphi$ at every world. In each case, these bisimilar completions may be used just as before to establish the abstract completeness theorem desired. QED

Just as the above proof illustrates our previous abstract completeness theorems may be adapted, our finitary approaches may be modified to accommodate universal modality as well.

Theorem 3.10.82 (Small Model Property for Universal EvIL)**.**
For any universal EvIL *formula φ, if it satisfiable then it is satisfiable in an* EvIL *model with $O(EXP2(|\varphi|))$ many worlds.*

Proof. By assumptions and soundness, we know that $\nvdash \neg\varphi$, so we can make a finite model by using the finite canonical model construction ☺ we previously saw in §3.5. As with the abstract canonical model construction, we shall ultimately want to take a point-generated submodel around the world we constructed which witnesses φ.

Just as in the original definition of ☺, our finite model construction needs a special definition for the universal modality. Define the relation associated with the universal modality as follows:

$$wUv \iff (U\varphi \in w \iff U\varphi \in v)$$

Where $\mathbb{M}, w \Vdash U\varphi \iff$ for all $v \in W$.

Along with the axioms 3.10–3.10, these will enforce that U forms an $S5$ modality; see [Boo95, chapter 5, pgs. 81–82] for details.

To ensure that the other relations are subrelations of U, one needs to ensure that Σ is extended so the definition includes the following:

$$\begin{aligned}\Sigma(\Delta, U\varphi) :=& \{U\varphi, \neg U\varphi\} \cup \\ & \{\Box_X\varphi, \neg\Box_X\varphi, \\ & \boxminus_X\varphi, \neg\boxminus_X\varphi, \\ & \boxplus_X\varphi, \neg\boxplus_X\varphi \mid X \in \Delta\}\end{aligned}$$

Given Σ constructed in this fashion, one may readily verify that the Universal EvIL axioms enforce the that other relations are subrelations of our candidate Universal relation. One may then use the ☋ bisimulation to complete ☺ from partly EvIL Kripke structure to a fully EvIL Kripke structure.

Furthermore, we may note that the complexity of Σ is unchanged by this modification, so our previous bound of $O(EXP2(|\varphi|))$ we gave on the number of worlds in ☺ and ☋$^{☺}$ do not change from what we provided in Theorem 3.5.56. QED

The above small model property has two consequences:

Theorem 3.10.83 (UEvIL Decidability)**.** *UEvIL, UEvIL$^{\boxminus}$, and UEvIL$^{\boxplus}$ are decidable and the time complexity of their decision problems is bounded above by $O(EXP3(|\varphi|))$*

Proof. The proof proceeds the same as the proof of Theorem 3.5.57, the EvIL decidability theorem from §3.5. QED

Theorem 3.10.84. *Assuming that Γ is finite and the set of letters Φ is universal:*

$$\begin{aligned}\Gamma \vdash_{U\text{EvIL}} \varphi &\iff \Gamma \Vdash \varphi \\ \Gamma \vdash_{U\text{EvIL}^{\boxminus}} \varphi &\iff \Gamma \Vdash \varphi \\ \Gamma \vdash_{U\text{EvIL}^{\boxplus}} \varphi &\iff \Gamma \Vdash \varphi\end{aligned}$$

Proof. Since by Theorem 3.10.81, we know that $U\text{EvIL}^{\boxminus}$ and $U\text{EvIL}^{\boxplus}$ are subsystems of $U\text{EvIL}$, we need only prove that

$$\Gamma \vdash_{U\text{EvIL}} \varphi \iff \Gamma \Vdash \varphi$$

As usual, we only prove completeness. Assume that $\Gamma \vdash_{U\text{EvIL}} \varphi$, then we know there is some finite EvIL Kripke structure $\mathbb{M}$ with a world w such that $\mathbb{M}, w \nVdash \bigwedge \Gamma \to \varphi$. Next, we may employ induction to extend Lemma 3.7.62, the EvIL translation lemma from §3.7, to include among other equivalences

$$\mathbb{M}, w \Vdash U\psi \iff \text{✡}, \vartheta(w) \Vdash U\psi.$$

Here $U\psi$ is assumed to be a subformula of φ. This establishes that $\text{✡}, \vartheta(w) \nVdash \bigwedge \Gamma \to \varphi$, giving the desired completeness result. QED

As in the previous section, we admit that we have made certain design choices here for the sake of simplicity. Universal modality hints, however, at richer semantics one might choose to develop.

For instance, we might imagine an our original concrete EvIL models might have, associated with them, a family of indexed accessibility relations R_X representing traditional epistemic logic accessibility relations. The semantics for $\Box_X \varphi$ would then be characterized as:

$$\mathfrak{M}, (a, A) \Vdash \Box_X \varphi \iff \forall a R_X b. \mathfrak{M}, (b, B) \Vdash A_X \text{ implies } \mathfrak{M}, (b, B) \Vdash \varphi$$

It is straightforward to see that EvIL is sound and complete for these semantics, given our previous results. We could then extend EvIL with traditional epistemic modalities corresponding to the accessibility relations postulated. Our original Theorem Theorem would have to be relativized in the following manner:

$$\mathfrak{M}, (a, A) \Vdash \Box_X \varphi \iff Th(R[a]) \cup A \vdash \varphi$$

Moreover, we could safely extend the grammar of the belief sets to include formulae containing old-fashioned epistemic modalities, governed by the accessibility relation. One could alternately investigate other extended modalities as well, such as the *difference* modality presented in [vB10, chapter 7.4]. Universal modality suggests that there are many modifications that could potentially be made to EvIL,

However, a system where every agent is equipped with an accessibility relation is more complicated than the simple, universal modality semantics we have developed for EvIL, As in §3.9, we did not choose to modify EvIL in some of the more exotic ways one might imagine, precisely because we wanted EvIL to conform to our original intuitions we developed in §1.

3.11 Lattice of Logics & Complexity

In this section, we discuss the relationship of the various EvIL logics previously developed. Specifically, we shall illustrate how the logics developed here provide a network of conservative extensions.

Hence, we shall be able to complement our previous upper complexity bounds with lower complexity bounds.

Before proceeding, we will review all of the logics we have developed:

Definition 3.11.85. *K – Basic modal logic, with a single modality. We first mentioned this logic in §1.3. It is defined in [BRV01, chapter 4, pg. 194].*

EviL – *The logic axiomatized in Table 3.1 in §3.1. It has multiple agents*

EviL$^{\boxminus}$ – *The $\boxminus$ fragment of single agent* EvиL*. This logic was axiomatized in Table 3.2 in §3.9.*

EviL$^{\boxplus}$ – *The $\boxplus$ fragment of single agent* EvиL*. This logic was axiomatized in Table 3.2 in §3.9.*

UML – Basic modal logic, with a single modality extended with a universal modality. This means that, in addition to the axioms and rules of the basic modal logic K, it also possesses the axioms provided in Table 3.3 from §3.10. Note that the universal modality axioms of UML are restricted to the basic modal language. This system is described in [vB10, chapter 7.4].

*U***EviL$^{\boxminus}$** – *The system* EvиL$^{\boxminus}$ *extended with the axioms provided in Table 3.3 from §3.10.*

*U***EviL$^{\boxplus}$** – *The single agent system* EvиL$^{\boxplus}$ *extended with the axioms provided in Table 3.3 from §3.10.*

*U***EviL** – EvиL *with multiple agents extended with the universal modality axioms in Table 3.3 from §3.10.*

To see how all of these logics inter-relate, we first establish the following lemma:

Lemma 3.11.86 (K/UML EvиL Soundness and Completeness)**.** *For all Γ:*

$$\Gamma \vdash_K \varphi \iff \Gamma \Vdash_{\text{EvиL}} \varphi$$
$$\Gamma \vdash_{UML} \varphi \iff \Gamma \Vdash_{\text{EvиL}} \varphi$$

For any finite Γ, given Φ is infinite:

$$\Gamma \vdash_K \varphi \iff \Gamma \Vdash \varphi$$
$$\Gamma \vdash_{UML} \varphi \iff \Gamma \Vdash \varphi$$

Proof. In all cases, soundness is trivial so we only need to prove completeness. Furthermore, we may restrict ourselves to the abstract Kripke semantics, since result for the concrete semantics follow from the abstract results and Theorem 3.10.84 from §3.9. Finally, we shall only show the result for UML, since the construction for K is similar.

So assume that $\Gamma \nvdash_{UML} \varphi$, then there is some model $\mathbb{M} = \langle W, R, V\rangle$ with a world w such that $\mathbb{M}, w \Vdash \Gamma$ and $\mathbb{M}, w \nVdash \varphi$. We need extend this structure to an EvиL Kripke structure.

It suffices to define $\sqsupseteq_Y$, $\sqsubseteq_Y$, and P_Y in the following manner:

$$\sqsupseteq_Y := \sqsubseteq_Y := id_W = \{(w, v) \in W \times W \mid w = v\}$$
$$P := \{w \in W \mid wRw\}$$

This construction effectively makes every world an island, and ignores additional agents. It is straightforward to check that $\langle W, R, \sqsubseteq, \sqsupseteq, P, V\rangle$ is EvidL, and that it preserves the truth of all UML formulae. QED

We may use the various soundness and completeness theorems to see that the logics defined in Definition 3.11.85 give rise to a network of conservative extensions. Specifically, the results employed are:

- Theorem 3.3.44 from §3.3
- Theorem 3.9.76 from §3.9
- Theorem 3.10.81 from §3.10
- Lemma 3.11.86 above
- The (informal) observation that the extension of any single agent logic to a multi-agent logic is conservative

This network is summarized in Fig. 3.6. The network is a Boolean lattice, where each node corresponds to a set of language features we have axiomatized.

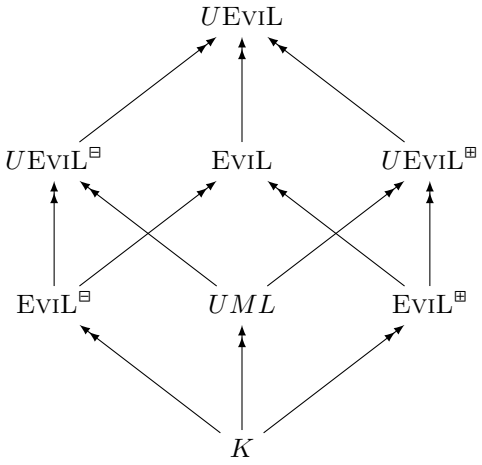


Figure 3.6: EvidL conservative extensions of K

With the understanding of how these languages inter-relate in terms of expressivity, we may give complexity bounds on all of the logics discussed:

Lemma 3.11.87. *For any of the logics defined in Definition 3.11.85, we know their decision problems are $PSPACE$-hard, and can always be decided in $3EXPTIME$.*

The complexity of the decision problem for all logics extending UML is $EXPTIME$ hard.

Proof. We know that all of the logics can be decided in $3EXPTIME$ by Theorem 3.10.83 from §3.10, since all of the systems investigated here are subsystems of UEviL.

We know all the logics are $PSPACE$ hard, since they all extend basic modal logic, which is $PSPACE$ complete [vB10, chapter 6.3].

Similarly, we know all of the logics extending UML are $EXPTIME$ hard, since UML is $EXPTIME$ complete [vB10, chapter 7.4]. QED

The network of EviL logics sheds new light on old logics. An EviL reading of the minimal modal logic K illustrates that it is the logic of a single agent justifying their beliefs with arguments, as we first discussed in §1.3. UML is a logic where one has background knowledge, the traditional knowledge studied in epistemic logic. We may then see that the other logics correspond to additional epistemic mechanisms one may wish to investigate. This is the EviL perspective on modal logic: one begins with concrete intuitions about what is required to know something, and employs Kripke semantics as a powerful abstraction on these intuitions.

This commences our investigations into EviL completeness theory.

Chapter 4

Applications

In this section, we will look at three applications of EvIL to wider philosophical topics. This section is organized roughly as follows:

§4.1 In this section, we investigate collapse issues related to both doxastic and epistemic introspection

§4.2 In this section, we present variations on the Gödel-Tarski-McKinsey translation of intuitionistic logic into EvIL. This sheds a novel light on the relationship of intuitionistic logic and the view on epistemic logic we have extended here.

4.1 Collapse

In this section, we investigate how postulating various sorts of *introspection* may (or may not) lead to collapse results for EvIL.

Negative Doxastic Introspection

We first recall the statement of negative doxastic introspection in the single agent case:

$$\vdash \neg\Box\varphi \rightarrow \Box\neg\Box\varphi$$

It may be read, informally, as "If the agent does not believe some proposition φ, then they believe that they do not believe it." Given the justificatory nature of EvIL, it is hard to understand what this would be like. What is the reason for believing that you do not believe something? Perhaps negative introspection comes from some kind of internal sensory apparatus. We offer no explanations for what might be this phenomenon, as we are admittedly skeptical that it and find counter-intuitive.

As mentioned in [VMTD05, *Meditations II*], it is plausible that, at any moment, one may try to cast just about everything they know into doubt and reason from a minimal number of assumptions. We

should assume, as Descartes, that these minimal assumptions are *safe*. This philosophical insight is easily expressed in as the following: EvıL:

$$\Diamond\!\!\!\!-\!\!\!\circlearrowleft$$

However, together we can see that thinking about how these concepts interact leads to a collapse of belief into truth for EvıL Kripke structures:

Theorem 4.1.88 (Negative Doxastic Collapse)**.** *Consider any* EvıL *Kripke structure* $\mathbb{M}$ *that makes true doxastic negative introspection. Then for all worlds* w *and all formulae* φ

$$\mathbb{M}, w \Vdash \Diamond\!\!\!\!-\!\!\!\circlearrowleft \rightarrow \Box\varphi \rightarrow \varphi$$

Proof. First assume $\mathbb{M}, w \Vdash \Diamond\!\!\!\!-\!\!\!\circlearrowleft$; if this does not hold then the statement is vacuously true. We must show $\mathbb{M}, w \Vdash \Box\varphi \rightarrow \varphi$; by semantics it suffices to show that

$$\mathbb{M}, w \Vdash \Box\varphi \Longrightarrow \mathbb{M}, w \Vdash \varphi.$$

By our assumption know there must be some v such that $w \sqsupseteq v$ and $\mathbb{M}, v \Vdash\!\!\!\circlearrowleft$. Thus vRv from property (VII), and from property (VI) we may further conclude that vRw. Thus, we have:

$$\mathbb{M}, v \Vdash \Box\varphi \Longrightarrow \mathbb{M}, w \Vdash \varphi$$

Now assume that $\mathbb{M}, v \nVdash \Box\varphi$, then $\mathbb{M}, w \nVdash \Box\varphi$. This follows from negative doxastic introspection, and may be observed in the following way:

$$\begin{aligned}
\mathbb{M}, v \nVdash \Box\varphi &\Longrightarrow \mathbb{M}, v \Vdash \neg\Box\varphi \\
&\Longrightarrow \mathbb{M}, v \Vdash \Box\neg\Box\varphi && \text{by negative doxastic introspection} \\
&\Longrightarrow \mathbb{M}, w \Vdash \neg\Box\varphi && \text{since } vRw \\
&\Longrightarrow \mathbb{M}, w \nVdash \Box\varphi
\end{aligned}$$

Contrapositively, this means that

$$\mathbb{M}, w \Vdash \Box\varphi \Longrightarrow \mathbb{M}, v \Vdash \Box\varphi$$

Whence:

$$\begin{aligned}
\mathbb{M}, w \Vdash \Box\varphi &\Longrightarrow \mathbb{M}, v \Vdash \Box\varphi \\
&\Longrightarrow \mathbb{M}, w \Vdash \varphi
\end{aligned}$$

The above suffices to show the theorem. QED

Informally, we may read Theorem 4.1.88 as asserting, given doxastic introspection, "If the agent knows anything, everything she believes is true." In the concrete semantics, doxastic introspection has an even stronger consequence:

Theorem 4.1.89 (Concrete Collapse)**.** *Let $\mathfrak{M}$ be an* EvIL *model making true doxastic introspection. Then for all worlds $(a, A) \in \mathfrak{M}$:*

$$\mathfrak{M}, (a, A) \Vdash \Diamond\!\!\!\!\circlearrowleft \rightarrow \circlearrowleft$$

Proof. It suffices to assume $\mathfrak{M}, (a, A) \Vdash \Diamond\!\!\!\!\circlearrowleft$ and show $a \models A$. So fix some $\psi \in A$.

Since $\mathfrak{M}$ can be understood as an EvIL Kripke model via (as we saw in Proposition 2.2.27 from §), we know from Theorem 4.1.88 we have:

$$\mathfrak{M}, (a, A) \Vdash \Diamond\!\!\!\!\circlearrowleft \rightarrow \Box\varphi \rightarrow \varphi$$

Whence, by our assumption:

$$\mathfrak{M}, (a, A) \Vdash \Box\varphi \rightarrow \varphi$$

Since $\psi \in A$ we know that $\mathfrak{M}, (a, A) \Vdash \Box\psi$, hence $\mathfrak{M}, (a, A) \Vdash \psi$. Hence, by from Lemma 2.1.10, the Truthiness Lemma from §2.1, we have that $a \models \psi$.

Since ψ was arbitrary, we know that $a \models A$, as desired. QED

While abstractly, we know that doxastic introspection implies a collapse of belief into truth, concretely accessibility amounts to truth. We may read 4.1.89 as asserting "If the agent has a sound argument, all of her arguments are sound."

Positive Doxastic Introspection

While negative doxastic introspection leads to a serious collapse, given the existence of a sound subset of the agents beliefs, no such result obtains when postulating positive doxastic introspection. A rather simple example may be given using the concrete semantics:

Proposition 4.1.90. *A* EvIL *model with positive introspection does not necessarily have the collapses seen in Theorem 4.1.88 and Theorem 4.1.89.*

Proof. To show the proposition, we must exhibit a suitable counter-example.

Let $\mathfrak{M} := \{(\varnothing, \varnothing), (\varnothing, \{\bot\})\}$. This model is depicted in Fig. 4.1. It is simple to check that positive introspection holds, as the $R^{\mathfrak{M}}$ accessibility relation is transitive. Moreover, we can see that

$$\mathfrak{M}, (\varnothing, \{\bot\}) \nVdash \Diamond\!\!\!\!\circlearrowleft \rightarrow \Box\bot \rightarrow \bot$$

In a similar vain, we know that:

$$\mathfrak{M}, (\varnothing, \{\bot\}) \nVdash \Diamond\!\!\!\!\circlearrowleft \rightarrow \circlearrowleft$$

This means that it is not possible to prove either of the collapse theorems we previously saw. QED

Hence, with the above observation, we tentatively extend the supposition that positive doxastic introspection may be construed as a relatively safe addition to EvIL, contrary to negative doxastic introspection.

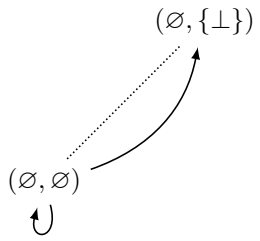


Figure 4.1: A model $\mathfrak{M}$ where $\mathfrak{M}, (\varnothing, \{\bot\}) \not\Vdash \Diamond\!\!\!\!-\!\!\!\!\circlearrowleft\!\!\rightarrow \Box\bot \rightarrow \bot$

Epistemic Introspection

In this section we consider a positive and negative form of epistemic introspection. Unlike the case for doxastic introspection, we shall reveal that neither of these assumptions are safe additions to EvIL.

Recall the definition of knowledge proposed in §1.4; this amounted to defining:

$$K\varphi := \Diamond\!\!\!\!-\!\!\!\!(\circlearrowleft \wedge \Box\varphi)$$

This meant "The agent has a sound argument." We might think of what would happen if we assumed negative epistemic introspection using this definition. This would assert:

$$\neg K\varphi \rightarrow K\neg K\varphi$$

Positive epistemic introspection can similarly be postulated as follows:

$$K\varphi \rightarrow KK\varphi$$

Both forms of introspection have the following consequence:

Theorem 4.1.91. *Assume that* $\mathbb{M}$ *is an* EvIL *model which makes true either* ***negative*** *introspection in the manner asserted above. Then for all formulae* φ *and all worlds* w*, we have:*

$$\mathbb{M}, w \Vdash \Diamond\!\!\!\!-\!\!\!\!\circlearrowleft\!\!\rightarrow \neg K\varphi \rightarrow \boxplus\neg K\varphi$$

Proof. Assume that $\mathbb{M}, w \Vdash \Diamond\!\!\!\!-\!\!\!\!\circlearrowleft$, $\mathbb{M}, w \Vdash \neg K\varphi$, and assume $w \sqsubseteq v$. We must show that $\mathbb{M}, v \Vdash \neg K\varphi$.

We know the following:

- From $\mathbb{M}, w \Vdash \Diamond\!\!\!\!-\!\!\!\!\circlearrowleft$, there is some $u \in P$ such that $u \sqsubseteq w \sqsubseteq v$.
- We know that uRu from property (VII), and from property (VI) we have that uRw and uRv
- We know that $\mathbb{M}, u \Vdash \neg K\varphi$. For suppose towards a contradiction that $\mathbb{M}, u \Vdash K\varphi$, then there would be some $t \sqsubseteq u$ such that $\mathbb{M}, t \Vdash\!\!\circlearrowleft \wedge\Box\varphi$. However, by transitivity (property (II)) we would have that $t \sqsubseteq w$, whence $\mathbb{M}, u \Vdash \neg K\varphi$, contrary to our hypothesis. $\lightning$

- From the above, we may gather:

$$\begin{aligned} \mathbb{M}, u \Vdash \neg K\varphi &\Longrightarrow \mathbb{M}, u \Vdash K\neg K\varphi && \text{by negative doxastic introspection} \\ &\Longrightarrow \mathbb{M}, u \Vdash \Box\neg K\varphi && \text{since } \vdash K\varphi \rightarrow \Box\varphi \\ &\Longrightarrow \mathbb{M}, v \Vdash \neg K\varphi && \text{since } uRv \end{aligned}$$

QED

Theorem 4.1.92. *Assume that* $\mathbb{M}$ *is an* EvIL *model which makes true either* ***positive*** *epistemic introspection. Then for all formulae* φ *and all worlds* w*, we have:*

$$\mathbb{M}, w \Vdash \neg K\varphi \rightarrow \boxplus\neg K\varphi$$

Proof. As before, assume that $\mathbb{M}, w \Vdash \neg K\varphi$, and assume $w \sqsubseteq v$. We must show that $\mathbb{M}, v \Vdash \neg K\varphi$.

In this case we shall instead illustrate the result by exhibiting the contrapositive, namely that if $\mathbb{M}, v \Vdash K\varphi$ then $\mathbb{M}, w \Vdash K\varphi$. Assuming $\mathbb{M}, v \Vdash K\varphi$, then there is some $u \sqsubseteq v$ such that $\mathbb{M}, u \Vdash\!\circlearrowleft \wedge\Box\varphi$.

Evidently, by the EvIL properties (VI) and (VII) we have that uRw. Moreover, we know that

$$\begin{aligned} \mathbb{M}, u \Vdash\!\circlearrowleft \wedge\Box\varphi &\Longrightarrow \mathbb{M}, u \Vdash \Diamond\!\!\!\!-(\circlearrowleft \wedge\Box\varphi) && \text{since } \sqsupseteq \text{ is reflexive} \\ &\Longrightarrow \mathbb{M}, u \Vdash K\varphi \\ &\Longrightarrow \mathbb{M}, u \Vdash KK\varphi && \text{by positive introspection} \\ &\Longrightarrow \mathbb{M}, u \Vdash \Box K\varphi && \text{since } \vdash K\varphi \rightarrow \Box\varphi \\ &\Longrightarrow \mathbb{M}, w \Vdash K\varphi && \text{since } uRw \end{aligned}$$

With the above, we have established what we set out to show. QED

Before proceeding, we offer a way to read the above two theorems.

Negative epistemic introspection, for the proposed semantics for knowledge, entails that "If the agent knows anything at all, then if she does not know something she never will." No matter what more evidence she embraces, none of it will lead to sound arguments for anything she does not already have sound arguments for, if she has any knowledge at all. In a way, negative doxastic introspection models agents with "enlightenment" moments, where as they become more and more aware of all of their evidence around them, they suddenly achieve knowledge of some set of propositions, and then learn all that they ever will.

The case for positive epistemic introspection is stronger than negative introspection. Given positive introspection, an agent can never compose a sound argument by remembering certain details. Knowledge, for any particular island, is completely static (recall the definition of *island* from §3.6). Every world in an island has exactly the same set of propositions for which the agent has a sound argument. Positive epistemic introspection, even more than negative epistemic introspection, postulates that at any possible world, the agent cannot learn anything by appealing to their own evidence.

Hence, we should conclude that neither positive nor negative epistemic introspection are particularly useful or intuitive axioms to enforce on EvIL models and Kripke structures.

4.2 Intuitionistic Logic

Introduction

In this section, we remark on the connection that single agent EvIL bears to *intuitionistic logic*.

Informally, we may find inspiration in the following quote:

> To be sure, intuitionistic logic, too, has its Kripke-style possible world semantics . . . Worlds stand for information states, accessibility encodes possible informational growth, and truth at a world corresponds intuitively to epistemic 'forcing' by the available evidence there. [vB91]

This key observation is analogous to our perspective on agents we took up in §1.6. Recall that we thought of agents as *posets*, where higher nodes represented the agent embracing more evidence. This can be regarded as the essentially the perspective as the epistemic reading of intuitionistic logic given above. On the other hand, another EvIL connection to traditional intuitionistic logic may be seen found. This corresponds to a well known proposal for constructive semantics extended by Hilary Putnam, who suggests "To claim a statement is true is to claim it could be justified" [Put81]. Since our proposed Theorem Theorem (Theorem 2.1.12 from §2.1) equates $\Box$-boxed formulae with justification, we may leverage Putnam's philosophical insight into a formal observation in EvIL.

The rest of this section is devoted to illustrating the EvIL connection to intuitionistic logic.

Preliminaries

We first recall the Kripke semantics for intuitionistic logic may be given as follows:

Definition 4.2.93 (Intuitionistic Kripke Structures)**.** *Let $\mathbb{P} = \langle W, \sqsubseteq, V\rangle$ be a Kripke structure. We say that $\mathbb{P}$ is **intuitionistic** if $\sqsubseteq$ is transitive and reflexive and V is **monotone**, that is if $w \in V(p)$ and $v \sqsupseteq w$ then $v \in V(p)$ (here $v \sqsupseteq w$ is just shorthand for $w \sqsubseteq v$).*

The of intuitionistic language is similar to the basic grammar $\mathcal{L}_0(\Phi)$

Definition 4.2.94. *Define the grammar $\mathcal{L}_{Int}(\Phi)$ as follows:*

$$\varphi ::= p \in \Phi \mid \bot \mid \varphi \to \psi \mid \varphi \wedge \varphi \mid \varphi \vee \psi$$

Likewise, the intuitionistic truth predicate $\Vdash\!\!\!\!-$ is similar to $\vdash$, and can be seen as a function which takes as input:

- A Kripke structure $\mathbb{P}$

(U1)	$\vdash \varphi \to \psi \to \varphi$
(U2)	$\vdash (\varphi \to \psi \to \chi) \to (\varphi \to \psi) \to \varphi \to \chi$
(U3)	$\vdash \varphi \wedge \psi \to \varphi$
(U4)	$\vdash \varphi \wedge \psi \to \psi$
(U5)	$\vdash (\varphi \to \psi) \to (\varphi \to \chi) \to (\varphi \to \psi \wedge \chi)$
(U6)	$\vdash \varphi \to \varphi \vee \psi$
(U7)	$\vdash \psi \to \varphi \vee \psi$
(U8)	$\vdash (\varphi \to \chi) \to (\psi \to \chi) \to (\varphi \vee \psi) \to \chi$
(U9)	$\vdash \bot \to \varphi$
(I)	$\dfrac{\vdash \varphi \to \psi \qquad \vdash \varphi}{\vdash \psi}$

Table 4.1: Intuitionistic Logic

- A world w
- An $\mathcal{L}_{Int}(\Phi)$ formula φ

And outputs something is bool. This may be written technically as:

$$(\Vdash) \ : \ \mathcal{K}_{\Phi,I} \to I \to \mathcal{L}_0(\Phi) \to \mathsf{bool}$$

It is defined recursively as follows:

Definition 4.2.95. *Let $\mathbb{P} = \langle W, \sqsubseteq, V \rangle$ be a Kripke structure:*

$$\begin{aligned}
\mathbb{P}, w \Vdash p &\iff w \in V(p) \\
\mathbb{P}, w \Vdash \varphi \to \psi &\iff \text{for all } v \sqsupseteq w\text{: } \mathbb{P}, v \Vdash \varphi \textit{ implies } \mathbb{P}, v \Vdash \psi \\
\mathbb{P}, w \Vdash \bot &\iff \textit{False} \\
\mathbb{P}, w \Vdash \varphi \wedge \psi &\iff \mathbb{P}, w \Vdash \varphi \ \& \ \mathbb{P}, w \Vdash \psi \\
\mathbb{P}, w \Vdash \varphi \vee \psi &\iff \mathbb{P}, w \Vdash \varphi \text{ or } \mathbb{P}, w \Vdash \psi
\end{aligned}$$

Next, we present an axiomatization of *Intuitionistic Logic*, which can be found in Table 4.1 below. These axioms are taken from [US06, chapter 5, pgs. 104–107]. Intuitionistic logic can be easily understood as the logic of intuitionistic Kripke structures:

Definition 4.2.96. *We shall write*

$$\Gamma \Vdash_{Int} \varphi$$

to mean that for all intuitionistic Kripke structures $\mathbb{P} = \langle W, \sqsubseteq, V \rangle$, for all worlds $w \in W$ if $\mathbb{P}, w \Vdash \Gamma$ then $\mathbb{M}, w \Vdash \varphi$.

Theorem 4.2.97 (Strong Intuitionistic Soundness and Completeness)**.**

$$\Gamma \vdash_{Int} \varphi \iff \Gamma \Vdash_{Int} \varphi$$

Proof. This is Proposition 5.1.10 from [US06, chapter 5, pg. 107]. QED

(U1)	$\vdash \varphi \to \psi \to \varphi$
(U2)	$\vdash (\varphi \to \psi \to \chi) \to (\varphi \to \psi) \to \varphi \to \chi$
(U3)	$\vdash (\neg\psi \to \neg\varphi) \to \varphi \to \psi$
(U4)	$\vdash \boxplus\varphi \to \varphi$
(U5)	$\vdash \boxplus\varphi \to \boxplus\boxplus\varphi$
(U6)	$\vdash \boxplus(\varphi \to \psi) \to \boxplus\varphi \to \boxplus\psi$
(I)	$\dfrac{\vdash \varphi \to \psi \qquad \vdash \varphi}{\vdash \psi}$
(II)	$\dfrac{\vdash \varphi}{\vdash \boxplus\varphi}$

Table 4.2: The Modal Logic $S4$

The Gödel-Tarski-McKinsey Embedding

In order understand how intutionistic logic connects to EvIL, we shall first review the traditional Gödel-Tarski-McKinsey embedding of intuitionistic logic into the modal logic $S4$. We briefly review the grammar and axiomatics of $S4$ before proceeding:

Definition 4.2.98. *Define the grammar $\mathcal{L}_{S4}(\Phi)$ as follows:*

$$\varphi ::= p \in \Phi \mid \bot \mid \varphi \to \psi \mid \boxplus\varphi$$

The axiom systems $S4$ is listed in Table 4.2. We next review the completeness theorem for $S4$:

Definition 4.2.99. *We shall write*

$$\Gamma \Vdash_{S4} \varphi$$

to mean that for all Kripke structures $\mathbb{P} = \langle W, \sqsubseteq, V\rangle$, where $\sqsubseteq$ is transitive and reflexive, for all worlds $w \in W$, if $\mathbb{P}, w \Vdash \Gamma$ then $\mathbb{P}, w \Vdash \varphi$.

Theorem 4.2.100 ($S4$ Strong Soundness and Completeness)**.**

$$\Gamma \vdash_{S4} \varphi \iff \Gamma \Vdash_{S4} \varphi$$

Proof. This is Theorem 4.29 of [BRV01, chapter 4.3, pg. 205]. QED

With the above, we may now provide the traditional Gödel-Tarski-McKinsey embedding, which establishes that intuitionistic logic is a sublogic of $S4$ (up to translation):

Definition 4.2.101 (The Gödel-Tarski-McKinsey Embedding)**.** *The **Gödel-Tarski-McKinsey embedding** $(\cdot)^\circ : \mathcal{L}_{Int}(\Phi) \to \mathcal{L}_{S4}(\Phi)$ is a recursively defined function that takes formulae in the language of intuitionistic logic to formulae in the language of $S4$. It may given programmatically*

as follows:

$$
\begin{aligned}
p^\circ &:= \boxplus p \\
\bot^\circ &:= \bot \\
(\varphi \to \psi)^\circ &:= \boxplus(\varphi^\circ \to \psi^\circ) \\
(\varphi \wedge \psi)^\circ &:= \varphi^\circ \wedge \psi^\circ \\
(\varphi \vee \psi)^\circ &:= \varphi^\circ \vee \psi^\circ
\end{aligned}
$$

Theorem 4.2.102.

$$\Gamma \vdash_{Int} \varphi \iff \Gamma^\circ \vdash_{S4} \varphi^\circ$$

Proof. This is Theorem 3.83 of [CZ97, chapter 3, pg. 97]. QED

We now turn to providing an variation on the above embedding. This will allow us to observe a theorem similar Theorem 4.2.102, only for EvIL instead of $S4$. Our intuition for behind our novel embedding begins with the following observation, which holds for all EvIL $\mathbb{M}$:

$$\mathbb{M}, w \Vdash \Box\varphi \;\; \& \;\; w \subseteq v \Longrightarrow \mathbb{M}, w \Vdash \Box\varphi$$

This is a consequence of the EvIL property (V). Hence every EvIL Kripke structure can be translated into an intuitionistic structure in the following manner:

Definition 4.2.103. *For every Kripke structure* $\mathbb{M} = \langle W, R, \sqsubseteq, \sqsupseteq, V, P\rangle$*, define:*

$$\rho\mathbb{M} := \langle W, \sqsubseteq, V'\rangle$$

Where $V'(p) := \{w \in W \mid \mathbb{M}, v \Vdash \Box p\}$

Since the Theorem Theorem asserts that we may interpret $\mathfrak{M}, (a, A) \Vdash \Box p$ as "the EvIL agent has can justify p using her evidence A", we may construe the above definition as associating "Truth" with "could be justified", following Hilary Putnam's suggestion in [Put81] we previously discussed. Note that this intuition emanates from the special interpretation we gave to concrete EvIL models. On the other hand, while our intuitions are grounded in our concrete semantics, we are unhindered by them. All of our theorems take place in the abstract semantics, where we may obtain our results in a higher level of generality.

Lemma 4.2.104. *If* $\mathbb{M}$ *is* EvIL *then* $\rho\mathbb{M}$ *is an intuitionistic Kripke structure.*

Proof. As we previously remarked, $\rho\mathbb{M}$ is intuitionistic as a consequence of the EvIL property (V). QED

Thinking about this embedding, may arrive at our translation, and illustrate the connection it bears to our above translation in a lemma:

Definition 4.2.105 (The EvIL Gödel-Tarski-McKinsey Embedding)**.** *The* **Evil EviL *Gödel-Tarski-McKinsey embedding*** $(\cdot)^{☠} : \mathcal{L}_{Int}(\Phi) \to \mathcal{L}_{\text{EvIL}}(\Phi)$ *is a recursively defined function that takes formulae in the language of intuitionistic logic to formulae in the* EvIL *language:*

$$\begin{aligned} p^{☠} &:= \Box p \\ \bot^{☠} &:= \bot \\ (\varphi \to \psi)^{☠} &:= \boxplus(\varphi^{☠} \to \psi^{☠}) \\ (\varphi \wedge \psi)^{☠} &:= \varphi^{☠} \wedge \psi^{☠} \\ (\varphi \vee \psi)^{☠} &:= \varphi^{☠} \vee \psi^{☠} \end{aligned}$$

Lemma 4.2.106. *If* $\mathbb{M}$ *is an* EvIL *Kripke structure, then for all worlds* $w \in W$ *and for all* $\varphi \in \mathcal{L}(\Phi)$*:*

$$\rho\mathbb{M}, w \Vdash\!\!\!\Vdash \varphi \iff \mathbb{M}, w \Vdash \varphi^{☠}$$

Proof. The proof proceeds by a trivial induction on φ. QED

Hence, every EvIL Kripke structure may be coerced into an intuitionistic Kripke structure which faithfully preserves the truth of all intuitionistic formulae up to translation.

We may also observe that every intuitionistic Kripke structure may be coerced into a EvIL Kripke structure in a similar manner:

Definition 4.2.107 (Diagonal Functor)**.** *Define*

$$\Delta(S) := \{(s, s) \in S \times S \mid s \in S\}.$$

This is known as the ***Diagonal Functor*** *in the Category Theory literature. See [Awo06, chapter 9, pg. 181] for a discussion of applications of this functor.*

Definition 4.2.108. *For every Kripke structure* $\mathbb{P} = \langle W, \sqsubseteq, V\rangle$*, and for every set of letters* Φ *define:*

$$\partial\mathbb{P} := \langle W', R, \preccurlyeq, \succcurlyeq, V'\rangle, P$$

Where

$$\begin{aligned} W' &:= W \uplus \wp\Phi \\ R &:= \{(w, \Psi) \in W \times \wp\Phi \mid \forall p \in \Phi.\mathbb{P}, w \Vdash\!\!\!\Vdash p \Longrightarrow p \in \Psi\} \\ \preccurlyeq &:= \sqsubseteq \cup \Delta(\wp\Phi) \\ \succcurlyeq &:= \{(w, v) \in W \times W \mid v \preccurlyeq w\} \\ V'(p) &:= \{\Psi \subseteq \Phi \mid p \in \Psi\} \\ P &:= \varnothing \end{aligned}$$

Before proceeding, the intuition behind the above construction is that the we will leave the accessibility of the original intuitionistic structure $\mathbb{P}$ intact, however we will commute local truth

valuations to non-local truth valuations by adding new worlds. The new worlds represent every possible extension of the truth values for one of the original intuitionistic structure. Moreover, each new world is an island unto itself.

The above coercion is enough to turn any intuitionist Kripke structure into an EvIL one:

Lemma 4.2.109. *If $\mathbb{P}$ is an intuitionistic Kripke structure, then $\partial\mathbb{P}$ is* EvIL

Proof. The EvIL properties (I), (II), (III) and (VII) follow immediately by construction and the fact that $\mathbb{P}$ is assumed to be intuitionistic.

(IV) Assume that $w \preccurlyeq u$. We must show that for all $p \in \Phi$, $w \in V(p) \iff u \in V(p)$.

If $w \in W^{\partial\mathbb{P}}$, then either $w \in W^{\mathbb{P}}$ or $w \subseteq \Phi$.

If $w \in W^{\mathbb{P}}$, then we know by construction that $w \preccurlyeq u \iff w \sqsubseteq^{\mathbb{P}} v$, hence $u \in W^{\mathbb{P}}$. We also know by construction that $\forall p \in \Phi. W^{\mathbb{P}} \cap V(p) = \varnothing$, hence we have the desired result.

On the other hand, if $w \subseteq \Phi$ then by construction we have $w \preccurlyeq u \iff w = v$, which suffices.

(V) We must show $(R\circ \preccurlyeq) \subseteq R$. So assume that $w \preccurlyeq v$ and vRu. By construction it must be that

- $\{w, v\} \subseteq W^{\mathbb{P}}$
- There is some Ψ where $u = \Psi \subseteq \Phi$
- If $\mathbb{P}, v \Vdash p$ then $p \in \Psi$

To show wRu we must show $\mathbb{P}, w \Vdash p$ then $p \in \Psi$. So fix p and assume $\mathbb{P}, w \Vdash p$. We know by hypothesis that $\mathbb{P}$ is intuitionistic, hence if $\mathbb{P}, w \Vdash p$ then $\mathbb{P}, v \Vdash p$, since $w \sqsubseteq v$. Hence $\mathbb{P}, v \Vdash p$. However, we can conclude from above that $p \in \Psi$, which suffices.

(VI) We must show:

$$(\preccurlyeq \circ R) \subseteq R$$
and
$$(\succcurlyeq \circ R) \subseteq R$$

Assume that wRv and $u \preccurlyeq v$ or $v \preccurlyeq u$. In either case we must show wRu. By construction it must be that $w \in W^{\mathbb{P}}$ and $v \subseteq \Phi$. However, we may again reason by construction that in either of the cases $u \sqsubseteq v$ or $v \sqsubseteq u$, we have $u = v$, whence wRu as desired.

QED

Lemma 4.2.110. *If $\mathbb{P}$ is an intuitionistic Kripke structure, then for all $w \in W^{\mathbb{P}}$, we have:*

$$\mathbb{P}, w \Vdash \varphi \iff \partial\mathbb{P}, w \Vdash \varphi^{☠}$$

Proof. This proceeds by routine induction on φ. The only case worth mentioning is the case of $p \in \Phi$.

(U1)	$\vdash \varphi \to \psi \to \varphi$
(U2)	$\vdash (\varphi \to \psi \to \chi) \to (\varphi \to \psi) \to \varphi \to \chi$
(U3)	$\vdash (\neg\psi \to \neg\varphi) \to \varphi \to \psi$
(U4)	$\vdash p \to \boxplus p$
(U5)	$\vdash \boxplus\varphi \to \varphi$
(U6)	$\vdash \boxplus\varphi \to \boxplus\boxplus\varphi$
(U7)	$\vdash \boxplus(\varphi \to \psi) \to \boxplus\varphi \to \boxplus\psi$
(I)	$\dfrac{\vdash \varphi \to \psi \qquad \vdash \varphi}{\vdash \psi}$
(II)	$\dfrac{\vdash \varphi}{\vdash \boxplus\varphi}$

Table 4.3: Van Benthem $S4$

Assume that $\mathbb{P}, w \Vdash p$, then we know that if $wR\Psi$ then $p \in \Psi$, whence by construction $\partial\mathbb{P}, \Psi \Vdash p$. This means that $\partial\mathbb{P}, w \Vdash \Box p$, and since $p^{\skull} = \Box p$, we have the desired result.

Next assume that $\partial\mathbb{P}, w \Vdash \Box p$, and let $\Xi := \{q \in \Phi \mid \mathbb{P}, w \Vdash q\}$. Evidently $wR\Xi$. Moreover, by assumption we have that $\partial\mathbb{P}, \Xi \Vdash p$. By construction this implies that $\mathbb{P}, w \Vdash p$, as desired. QED

With the above established, we have enough to illustrate that intuitionistic logic is a sublogic of EviL, after translation:

Theorem 4.2.111.

$$\Gamma \vdash_{Int} \varphi \iff \Gamma^{\skull} \vdash_{\text{EviL}} \varphi^{\skull}$$

Proof. $\Longrightarrow$: Assume that $\Gamma^{\skull} \nvdash_{\text{EviL}} \varphi^{\skull}$, we must show that $\Gamma \nvdash_{Int} \varphi$. By EviL completeness (Theorem 3.3.44) we know there is some EviL Kripke structure $\mathbb{M}$ with a world w such that $\mathbb{M}, w \Vdash \Gamma^{\skull}$ and $\mathbb{M}, w \nVdash \varphi^{\skull}$. By Lemma 4.2.106, we know that $\rho\mathbb{M}, w \Vdash \Gamma$ and $\rho\mathbb{M}, w \nVdash \varphi$. Since $\rho\mathbb{M}$ is intuitionistic by Lemma 4.2.104, and Int is sound for intuitionistic Kripke structures (Theorem 4.2.97), we have that $\Gamma \nvdash_{Int} \varphi$.

$\Longleftarrow$: The proof proceeds as above, via contraposition. Difference is that here one uses intuitionistic completeness (Theorem 4.2.97), Lemmas 4.2.109 and 4.2.110 to coerce intuitionistic structures to faithfully turn EviL, and finally EviL soundness (Theorem 3.3.44). QED

Van Benthem S4

In this section we study how EviL relates to van Benthem $S4$, which provides a simple abstraction of intuitionistic logic into a modal setting. Van Benthem $S4$ is a non-normal extension to $S4$ discussed in [vB96, vB09, vBE89]. It is axiomatized in Table 4.3. It is essentially the same as $S4$, only predicate letters are specified to be upward monotone by a special, non-normal axiom.

Below, we offer several results that may be obtained for van Benthem $S4$. We do not intend to prove these theorems here; a curious reader may read [vB96, vB09, vBE89] for an in depth discussion of this logic.

Definition 4.2.112. *We say that* $\Gamma \vdash_{vBS4} \varphi$ *if and only if there is some* $\Sigma \subseteq_\omega \Gamma$ *such that* $\bigwedge \Sigma \to \varphi$ *is a theorem of van Benthem* $S4$*.*

Proposition 4.2.113 (Van Benthem's Strong Soundness and Completeness)**.**

$$\Gamma \Vdash_{Int} \varphi \iff \Gamma \vdash_{vBS4} \varphi$$

That is, van Benthem $S4$ *is sound and strongly complete for intuitionistic Kripke structures.*

Note that the language of van Benthem $S4$ is modal, and intuitionistic logic is not, hence the two logics are distinct. Moreover, van Benthem $S4$ is a classical calculus, despite having semantics over Intuitionistic Kripke structures. The following proposition summarizes the relationship of these calculi:

Definition 4.2.114 (Van Benthem's Embedding)**.** ***Van Benthem's embedding*** $(\cdot)^\bullet : \mathcal{L}_{Int}(\Phi) \to \mathcal{L}_{S4}(\Phi)$ *is a recursively defined function that takes formulae in the language of intuitionistic logic to formulae in the language of* $S4$*:*

$$\begin{aligned} p^\bullet &:= p \\ \bot^\bullet &:= \bot \\ (\varphi \to \psi)^\bullet &:= \boxplus(\varphi^\bullet \to \psi^\bullet) \\ (\varphi \wedge \psi)^\bullet &:= \varphi^\bullet \wedge \psi^\bullet \\ (\varphi \vee \psi)^\bullet &:= \varphi^\bullet \vee \psi^\bullet \end{aligned}$$

Van Benthem's Embedding is essentially the same as the Gödel-Tarski-McKinsey embedding, save that predicate letters are not boxed. We may observe the following theorem:

Proposition 4.2.115 (van Benthem's Embedding Theorem)**.**

$$\Gamma \vdash_{Int} \varphi \iff \Gamma^\bullet \vdash_{vBS4} \varphi^\bullet$$

We now turn to illustrating how van Benthem's $S4$ is a proper abstraction to think about EvIL embeddings for intuitionistic logic. We may understand the embedding ☠ in terms of an embedding of intuitionistic logic into van Benthem $S4$:

Definition 4.2.116 (Van Benthem's EvIL Embedding)**.** ***Van Benthem's* EviL *embedding*** *has the following type:*

$$(\cdot)^\dagger : \mathcal{L}_{S4}(\Phi) \to \mathcal{L}_{\text{EvIL}}(\Phi)$$

It is a recursively defined function that takes formulae in the language of $S4$ *modal logic logic to formulae in the* EvIL *language:*

$$\begin{aligned} p^\dagger &:= \Box p \\ \bot^\dagger &:= \bot \\ (\varphi \to \psi)^\dagger &:= \varphi^\dagger \to \psi^\dagger \\ (\boxplus\varphi)^\dagger &:= \boxplus\varphi^\dagger \end{aligned}$$

With the above we have two results:

Proposition 4.2.117.

$$\varphi^{☠} = (\varphi^{\bullet})^{\dagger}$$

Lemma 4.2.118. *For any* EvIL *Kripke structure* $\mathbb{M}$*:*

$$\rho\mathbb{M}, w \Vdash \varphi \iff \mathbb{M}, w \Vdash \varphi^{\dagger}$$

For any intuitionistic Kripke structure $\mathbb{P}$*:*

$$\mathbb{P}, w \Vdash \varphi \iff \partial\mathbb{P}, w \Vdash \varphi^{\dagger}$$

Proof. In each case the proof follows from structural induction; they essentially follow the proofs of Lemmas 4.2.106 and 4.2.110. QED

Hence we may illustrate that van Benthem $S4$ is indeed a sublogic of EvIL, after translation:

Theorem 4.2.119.

$$\Gamma \vdash_{vBS4} \varphi \iff \Gamma^{\dagger} \vdash_{\text{EvIL}} \varphi^{\dagger}$$

Proof. The proof follows the same structure as Theorem 4.2.111. Instead of appealing to Theorem 4.2.97, the strong intuitionistic soundness and completeness, one uses Proposition 4.2.113, van Benthem's strong soundness and completeness theorem. Also, instead of using Lemmas 4.2.106 and 4.2.110, one may simply appeal to Lemma 4.2.118. QED

The above theorems mean that EvIL does not just embed intuitionistic logic, but rather it may be elaborated to embed van Benthem $S4$. Hence, if one exhibits an embedding of van Benthem $S4$ into another system, one may automatically provide an embedding of intuitionistic logic, since intuitionistic logic is essentially a subcalculus of van Benthem $S4$. In the subsequent sections, rather than exhibiting how variations on the Gödel-Tarski-McKinsey embedding to illustrate the connection of intuitionistic logic to EvIL, we will instead focus on embedding van Benthem $S4$. In every case, we can observe that simply composing the some embedding of van Benthem $S4$ into EvIL after an embedding of intuitionistic logic into van Benthem $S4$ yields a novel embedding of intuitionistic logic. We are of the opinion that van Benthem $S4$ provides a good abstract domain to study the relationship between intuitionistic structures and EvIL, because of this phenomenon.

Knowledge

In this section, we present an alternative to the previous embedding of van Benthem $S4$ (and hence intuitionstic logic) into EvIL. This time, instead of associating truth conditions in intuitionistic logic with justifiability, we shall illustrate that they can be identified with knowledge. This illustrates that the remarks made of intuitionistic logic taken from [vB91] that we quoted in §4.2 exhibits a clear interpretation in the EvIL semantics we have set forth.

Definition 4.2.120 (The EviL Knowledge Embedding)**.** *The* **EviL *knowledge embedding*** $(\cdot)^{\text{☣}} : \mathcal{L}_{S4}(\Phi) \to \mathcal{L}_{\text{EviL}}(\Phi)$ *is a recursively defined function that takes formulae in the language of* $S4$ *modal logic logic to formulae in the* EviL *language:*

$$\begin{aligned} p^{\text{☣}} &:= Kp \\ \bot^{\text{☣}} &:= \bot \\ (\varphi \to \psi)^{\text{☣}} &:= \varphi^{\text{☣}} \to \psi^{\text{☣}} \\ (\boxplus\varphi)^{\text{☣}} &:= \boxplus\varphi^{\text{☣}} \end{aligned}$$

where $Kp := \Diamond\!\!\!\!-(\circlearrowleft \wedge \Box p)$*, as we first suggested in §1.5*

The idea in the above embedding is to associate truth the notion of knowledge presented previously in §1.4, namely that the agent has a sound argument. We may obtain the following result regarding this embedding:

Definition 4.2.121. *For every Kripke structure* $\mathbb{M} = \langle W, R, \sqsubseteq, \sqsupseteq, V, P\rangle$*, define:*

$$\eta\mathbb{M} := \langle W, \sqsubseteq, V'\rangle$$

Where $V'(p) := \{w \in W \mid \mathbb{M}, v \Vdash Kp\}$

Proposition 4.2.122. *If* $\mathbb{M}$ *is* EviL *then* $\eta\mathbb{M}$ *is an intuitionistic Kripke structure.*

Intuitively, the above lemma follows from the observation that, in the concrete semantics, as the agent has more propositions by which she may compose arguments, she has more subsets of that basis to compose *sound* arguments. Since the interpretation of knowledge for our definition of K is precisely the existence of a sound subsets of one's basic beliefs, this observation is codified in the following validity:

$$\vdash_{\text{EviL}} K\varphi \to \boxplus K\varphi$$

As in the previous embedding, we may also coerce intuitionistic structures to become EviL in a manner appropriate for our embedding ☣, and obtain a correspondence lemma:

Lemma 4.2.123. *For all* EviL *Kripke structure* $\mathbb{M}$*, we have that:*

$$\eta\mathbb{M}, w \Vdash \varphi \iff \mathbb{M}, w \Vdash \varphi^{\text{☣}}$$

Proof. This follows from a routine induction on the complexity of φ. QED

Definition 4.2.124. *For every Kripke structure* $\mathbb{P} = \langle W, \sqsubseteq, V\rangle$*, and for every set of letters* Φ *define:*

$$\mu\mathbb{P} := \langle W', R, \preccurlyeq, \succcurlyeq, V'\rangle, P$$

Where

$$\begin{aligned} W' &:= W \uplus \wp\Phi \\ R &:= W \times W \cup \{(w, \Psi) \in W \times \wp\Phi \mid \forall p \in \Phi. \mathbb{P}, w \Vdash p \Longrightarrow p \in \Psi\} \\ \preccurlyeq &:= \sqsubseteq \cup \Delta(\wp\Phi) \\ \succcurlyeq &:= \{(w, v) \in W \times W \mid v \preccurlyeq w\} \\ V'(p) &:= W \cup \{\Psi \subseteq \Phi \mid p \in \Psi\} \\ P &:= W \end{aligned}$$

Lemma 4.2.125. *If $\mathbb{P}$ is an intuitionistic Kripke structure, then $\mu\mathbb{P}$ is* EvIL

Proof. As in Lemma 4.2.109, the EvIL properties (I), (II), (III) and (VII) follow by construction, and the assumption that $\mathbb{P}$ is intuitionistic.

(IV) Assume that $w \preccurlyeq u$. We must show that for all $p \in \Phi$, $w \in V(p) \iff u \in V(p)$.

If $w \in W^{\partial\mathbb{P}}$, then either $w \in W^{\mathbb{P}}$ or $w \subseteq \Phi$.

If $w \in W^{\mathbb{P}}$, then we know by construction that $u \in W^{\mathbb{P}}$ as well. In this case we have defined our structure such that for all $p \in \Phi$ we have $w \in V(p)$ and $u \in V(p)$, so the result follows immediately.

On the other hand, if $w \subseteq \Phi$ then it must be that $w \preccurlyeq u \iff w = v$, which suffices.

(V) We must show $(R \circ \preccurlyeq) \subseteq R$. So assume that $w \preccurlyeq v$ and vRu. We have by construction:

- $\{w, v\} \subseteq W^{\mathbb{P}}$
- Either (A) $u \in W^{\mathbb{P}}$ or (B) there is some Ψ where $u = \Psi \subseteq \Phi$
- If $\mathbb{P}, v \Vdash\!\Vdash p$ then $p \in \Psi$

To show wRu we have two cases for u. If (A), then we know $u \in W^{mathbbP}$ so we have wRu by construction.

On the other hand, in case (B) the argument is analogous to the proof given in Lemma 4.2.109.

(VI) We must show:

$$(\preccurlyeq \circ R) \subseteq R$$
and
$$(\succcurlyeq \circ R) \subseteq R$$

Assume that wRv and $u \preccurlyeq v$ or $v \preccurlyeq u$. In either case we must show wRu. By construction it must be that $w \in W^{\mathbb{P}}$ and either (A) $vinW^{\mathbb{P}}$ or $v \subseteq \Phi$. In case (A) it must be that $u \in W^{\mathbb{P}}$, hence wRu by construction. In case (B), the argument again proceeds in a fashion analogous to Lemma 4.2.109.

QED

Lemma 4.2.126. *If $\mathbb{P}$ is an intuitionistic Kripke structure, then for all $w \in W^{\mathbb{P}}$, we have:*

$$\mathbb{P}, w \Vdash \varphi \iff \mu\mathbb{P}, w \Vdash \varphi^{☣}$$

Proof. This proceeds by routine induction on φ. As with Lemma 4.2.110, the only case worth mentioning is the case of $p \in \Phi$.

Assume that $\mathbb{P}, w \Vdash p$. By construction we have that $\mu\mathbb{P}, w \Vdash \Box p$. We can also easily see that $\mu\mathbb{P}, w \Vdash ↺$, hence $\mu\mathbb{P}, w \Vdash ↺ \wedge \Box p$. Since $\preccurlyeq$ is reflexive, we have that $\mu\mathbb{P}, w \Vdash \Diamond\!\!\!\!-(↺ \wedge \Box p)$, which is to say that $\mu\mathbb{P}, w \Vdash Kp$.

Now assume that $\mu\mathbb{P}, w \Vdash Kp$, then there is some $v \preccurlyeq u$ such that $\mu\mathbb{P}, v \Vdash \Box p$. As in the proof of Lemma 4.2.110, this implies that $\mathbb{P}, v \Vdash p$. Since $\mathbb{P}$ is intuitionistic, we know that since $w \succcurlyeq v$ we have $\mathbb{P}, w \Vdash p$, as desired. QED

Hence, we may observe that ☣ can be used to embed van Benthem $S4$ into EvиL.

Theorem 4.2.127.

$$\Gamma \vdash_{vBS4} \varphi \iff \Gamma^{☣} \vdash_{\text{EvиL}} \varphi^{☣}$$

Proof. As in the proof of Theorems 4.2.111 and 4.2.119, employs existing completeness results along with Proposition 4.2.122 along with Lemmas 4.2.123, 4.2.125, and 4.2.126. QED

As a corollary, we have an embedding of intuitionistic logic

Corollary 4.2.128.

$$\Gamma \vdash_{Int} \varphi \iff (\Gamma^{\bullet})^{☣} \vdash_{\text{EvиL}} (\varphi^{\bullet})^{☣}$$

Proof. This follows immediately from Proposition 4.2.115 and Theorem 4.2.127. QED

Imagination

In this section we present one final embedding of intuitionistic logic into EvиL. Previously, the inspiration behind the embeddings was the idea that as one ascends in an intuitionistic Kripke-structure, one has more evidence at their disposal to draw conclusions. This in turn meant that they could access fewer worlds, since fewer worlds were compatible with their beliefs.

In this section, we illustrate that the *dual* interpretation also holds. Now was one "ascends" in an intutionistic Kripke structure, the agent holds *fewer* beliefs, and considers more things possible.

This gives rise to the following embedding:

Definition 4.2.129 (The EvиL Imagination Embedding)**.** *The* **EviL *imagination embedding*** $(\cdot)^{☢} : \mathcal{L}_{Int}(\Phi) \to \mathcal{L}_{\text{EvиL}}(\Phi)$ *is a recursively defined function that takes formulae in the language of*

S4 modal logic to formulae in the EvIL *language:*

$$p^{☢} := \Diamond p$$
$$\bot^{☢} := \bot$$
$$(\varphi \to \psi)^{☢} := \varphi^{☢} \to \psi^{☢}$$
$$(\boxplus\varphi)^{☢} := \boxminus\varphi^{☢}$$

Definition 4.2.130. *For every Kripke structure* $\mathbb{M} = \langle W, R, \sqsubseteq, \sqsupseteq, V, P\rangle$*, define:*

$$\zeta\mathbb{M} := \langle W, \sqsupseteq, V'\rangle$$

Where $V'(p) := \{w \in W \mid \mathbb{M}, v \Vdash \Diamond p\}$

As in the previous cases, we may obtain the following results:

Proposition 4.2.131. *If* $\mathbb{M}$ *is* EvIL *then* $\zeta\mathbb{M}$ *is an intuitionistic Kripke structure.*

As in the previous cases, the above proposition can be understood naturally using the concrete semantics. We know that if some situation is compatible with an EvIL agents beliefs, then if that agent were to believe *fewer* things, then that situation would still be compatible with her beliefs. This may be expressed as the following validity:

$$\vdash_{\text{EvIL}} \Diamond\varphi \to \boxminus\Diamond\varphi$$

This previous validity entails the above proposition.

Lemma 4.2.132. *For all* EvIL *Kripke structures* $\mathbb{M}$

$$\eta\mathbb{M}, w \Vdash \varphi \iff \mathbb{M}, w \Vdash \varphi^{☢}$$

Proof. As in the previous embeddings, this follows by induction on the complexity of φ. QED

Definition 4.2.133. *For every Kripke structure* $\mathbb{P} = \langle W, \sqsubseteq, V\rangle$*, and for every set of letters* Φ *define:*

$$\xi\mathbb{P} := \langle W', R, \preccurlyeq, \succcurlyeq, V'\rangle, P$$

Where

$$\begin{aligned} W' &:= W \uplus \wp\Phi \\ R &:= \{(w, \Psi) \in W \times \wp\Phi \mid \forall p \in \Phi. p \in \Psi \Longrightarrow \mathbb{P}, w \Vdash p\} \\ \preccurlyeq &:= \sqsupseteq \cup \Delta(\wp\Phi) \\ \succcurlyeq &:= \{(w, v) \in W \times W \mid v \preccurlyeq w\} \\ V'(p) &:= \{\Psi \subseteq \Phi \mid p \in \Psi\} \\ P &:= \varnothing \end{aligned}$$

It is worth stopping for a moment to discuss the nature of the definition of ξ and its relation to ∂.

In $\partial\mathbb{P}$, we had that $wR\Psi$ if and only if $\mathbb{P}, w \Vdash p$ for all $p \in \Psi$. In this construction, a world w in the original intuitionistic Kripke structure $\mathbb{P}$ can "see" another a set of proposition letters Φ if and only if Ψ extends the valuation at w. Concisely, this is to say:

$$wR\Psi \text{ if and only if } \Psi \text{ is compatible with what is } \textit{believed} \text{ at } w$$

"Future" states from w correspond to possible extensions to the basic beliefs present at w. Evidently, the intuition behind this construction is recognizably the similar as the intuition behind our original concrete semantics.

However, this is evidently not the only way to the atomic truth conditions in intuitionistic logic; ξ suggests an alternative perspective. Instead of thinking of the atomic letters that are true at a particular world w in as assertions the agent believes, we invite the reader to interpret them as *assertions the agent considers possible*. Instead of the future states of w corresponding to informational extensions, the correspond to ways in which an agent might open her mind greater heights of imagination. Under this reading, we will want to enforce

$$wR\Phi \text{ if and only if } \Phi \text{ is compatible with what is imagined possible at } w$$

This means that if $p \in \Psi$ but $\mathbb{P}, w \nVdash p$, then we do not want $wR\Psi$. On the other hand, if we have that for all $p \in \Phi$ that if $\mathbb{P}, w \nVdash p$ then $p \notin \Psi$, then evidently Ψ is not going to be making true any atomic formula that is thought to be *impossible* at w; the contrapositive of this yields the necessary and sufficient conditions for $wR\Psi$ that we have given in ξ.

We may summarize the above as follows. While the traditional reading of intuitionistic Kripke structures is that they model agents acquiring beliefs, or *learning*, the reading behind the ξ construction is that they are can model *forgetting* or Cartesian *doubting*, as we first suggested in §1.5.

Lemma 4.2.134. *If $\mathbb{P}$ is an intuitionistic Kripke structure, then $\xi\mathbb{P}$ is* EvidL

Proof. As in the previous embedding theorems, we immediately have that properties (I), (II), (III) and (VII) hold. The proofs of (IV) and (VI) follow the arguments provided in the proof of Lemma 4.2.109.

Hence, all we have left to show is (V). We must show $(R\circ \preccurlyeq) \subseteq R$. So assume that $w \preccurlyeq v$ and vRu; we must show that wRu. By construction there must be some $\Psi \subseteq \Phi$ such that $u = \Psi$, and moreover we have that $w \sqsupseteq^{\mathbb{P}} v$. To show that $wR\Psi$ we must show that if $p \in \Psi$ then $\mathbb{P}, w \Vdash p$. But we know by construction that $vR\Psi$ implies that that if $p \in \Psi$ then $\mathbb{P}, v \Vdash p$, and since intuitionistic Kripke structures are monotone and $w \sqsupseteq^{\mathbb{P}} v$, we have $\mathbb{P}, w \Vdash p$ as desired. QED

This affords us our final embedding:

Theorem 4.2.135.

$$\Gamma \vdash_{vBS4} \varphi \iff \Gamma^{☢} \vdash_{\text{EvidL}} \varphi^{☢}$$

Proof. As in the case of the previous embedding theorems, the above follows from Proposition 4.2.131 Lemmas 4.2.132, 4.2.134 and our previously established completeness results. QED

Corollary 4.2.136.

$$\Gamma \vdash_{Int} \varphi \iff (\Gamma^{\bullet})^{\bigstar} \vdash_{\text{EvIL}} (\varphi^{\bullet})^{\bigstar}$$

Proof. As in the case of Corollary 4.2.128, this follows from Proposition 4.2.115 and Theorem 4.2.135. QED

This commences our investigations into the connection of EvIL to intuitionistic logic and intuitionistic Kripke structures.

Chapter 5

Epilogue

5.1 Introduction

In this final section, we discuss how EvIL compares to two alternative logics of explicit knowledge, the approach to epistemic logic suggested in [vB91]. The two other approaches are:

- *Velázquez-Quesada Logic* (VQL) from [Vel09]
- *Dynamic Awareness Logic* (DAL) from [vBV09]

Note that VQL and DAL are names we have introduced for the logics discussed above, for ease of reference.

We will conclude with our final thoughts on EvIL.

5.2 Velázquez-Quesada Logic

In this section we introduce Velázquez-Quesada Logic (VQL), the logic introduced in [Vel09]. The author originally called their logic **EI**, which stands for "explicit/implicit." We have chosen the name VQL, as there is more than one system that claims to model explicit and implicit knowledge. The idea behind the semantics of VQL is to equip Kripke structures with special *awareness* relations, as introduced in [FH87]. In [Vel09], the author notes himself how his work compares to other awareness based approaches. Notably, his work is related to Ho Ngoc Duc's logic [Duc95, Duc97, Duc01], Jago's logic for agents with bounded resources [Jag06], and van Benthem's *acts of realization* [vB08].

VQL Kripke structures have two awareness relations. The first relation establishes whether the agent is aware of a particular fact, while the second establishes whether the agent is aware of a rule. By imposing various properties on these enriched Kripke structures, one can model inference via update modalities; this will be illustrated shortly.

We note that early drafts of VQL were presented employing multiple agents, while the version

presented in [Vel09] limited itself to the single agent case.

Formally, we can understand VQL accordingly:

Definition 5.2.137. *A **rule** $\rho \in \wp_\omega \mathcal{L}_0(\Phi) \times \mathcal{L}_0(\Phi)$ is a pair (Γ, γ) where Γ is a finite set of propositional formulae and γ is a single propositional formula.*

We shall abbreviate $\mathcal{R}(\Phi)$ as the set of rules over Φ.

Definition 5.2.138. *The grammar $\mathcal{VQL}(\Phi)$ is defined as:*

$$\varphi ::= p \in \Phi \mid \top \mid \varphi \vee \psi \mid \neg\varphi \mid \Diamond\varphi \mid I\gamma \mid L\rho$$

Here $\gamma \in \mathcal{L}_0(\Phi)$, while $\rho \in \mathcal{R}(\Phi)$. The other connectives are assumed to be defined as usual.

Definition 5.2.139. *$TR : \mathcal{R}(\Phi) \to \mathcal{L}_0(\Phi)$ is a function that takes rules to first order formulae that codify them as follows:*

$$TR(\rho) := \bigwedge \Gamma \to \gamma$$

Where Γ is the set of premises of ρ and γ is the conclusion.

Definition 5.2.140. *A VQL Kripke structure is a model $\mathbb{V} = \langle W, R, V, Y, Z \rangle$ where*

- *W is a set of worlds*
- *$R \subseteq W \times W$ is an accessibility relation*
- *$V : W \to \wp\Phi$ is valuation*
- *$Y : W \to \wp\mathcal{L}_0(\Phi)$ is a propositional awareness function, relating worlds to sets of formulae the agent is aware of at that world*
- *$Z : W \to \mathcal{R}(\Phi)$ is a logical awareness function, relating worlds to sets of rules that the agent is aware of at those worlds*

Likewise, VQL Kripke structures make true the following properties (along with intuitive readings):

(*1*) *R is an equivalence relation*

(*2*) *If $\gamma \in Y(w)$ and wRv then $\gamma \in Y(v)$*

(*3*) *If $\rho \in Z(w)$ and wRv then $\rho \in Z(v)$ – once the agent is aware of some rule then that awareness persists*

(*4*) *If $\gamma \in Y(w)$ then $\mathbb{V}, w \Vdash_{VQL} \gamma$ – the agent is only aware of facts*

(*5*) *If $\rho \in Z(w)$ then $\mathbb{V}, w \Vdash_{VQL} \mathrm{TR}(\rho)$*

We may intuitively read the above properties in the following manner:

(*1*) *Information States* – R partitions the universe into different information states, as one traditionally assumes in epistemic logic

(*2*) *Assertoric Awareness Positive Introspection* – Once the agent is aware of some assertion, they know that they are aware of that rule

(3) *Logical Awareness Positive Introspection* – Once the agent is aware of some rule, they know that they are aware of that rule

(4) *Factiveness* – If the agent is aware of proposition, then it is a fact

(5) *Soundness* – If the agent is aware of some rule, then that rule is sound

The following provides semantics for the grammar in terms of the stipulated structures:

Definition 5.2.141. *VQL's truth predicate* $\Vdash_{VQL}$ *is defined such that:*

$$\begin{aligned}
\mathbb{V}, w \Vdash_{VQL} p &\iff w \in V(p) \\
\mathbb{V}, w \Vdash_{VQL} \top &\text{ always} \\
\mathbb{V}, w \Vdash_{VQL} \varphi \vee \psi &\iff \mathbb{V}, w \Vdash_{VQL} \varphi \text{ or } \mathbb{V}, w \Vdash_{VQL} \psi \\
\mathbb{V}, w \Vdash_{VQL} \neg\varphi &\iff \mathbb{V}, w \nVdash_{VQL} \varphi \\
\mathbb{V}, w \Vdash_{VQL} \Diamond\varphi &\iff \exists v \in W. wRv \ \& \ \mathbb{V}, v \Vdash_{VQL} \varphi \\
\mathbb{V}, w \Vdash_{VQL} I\gamma &\iff \gamma \in Y(w) \\
\mathbb{V}, w \Vdash_{VQL} L\rho &\iff \rho \in Z(w)
\end{aligned}$$

Table 5.1 expresses the axioms of VQL. Certain axioms impose certain structure on the models; soundness and completeness theorems are discussed in [Vel09].

The semantics for VQL also supports modeling inference. The author chose to model inference using model updates, in the style of dynamic epistemic logic as pioneered in [Ger98].

Definition 5.2.142 (Model Deduction Update)**.** *For a VQL Kripke structure* $\mathbb{V} = \langle W, R, V, Y, Z\rangle$*, and let* $\rho \in \mathcal{R}(\Phi)$ *be a rule. Let* Γ *be the premises of* ρ *and let* γ *be the conclusion. Define* $\mathbb{V}_\rho := \langle W, R, V, Y', Z\rangle$ *where:*

$$Y'(w) := \begin{cases} Y(w) \cup \{\gamma\} & \text{if } \Gamma \subseteq Y(w) \text{ and } \rho \in Z(w) \\ Y(w) & \text{o/w} \end{cases}$$

Lemma 5.2.143. *If* $\mathbb{V}$ *is a VQL Kripke structure then so is* $\mathbb{V}_\rho$ *where* $\rho \in \mathcal{R}(\Phi)$

Proof. This is Proposition 1 in [Vel09]. QED

The idea of the above operation is that at every world it adds in the conclusion of ρ if the agent is aware of the premises of ρ. Furthermore, it preserves the various properties that are imposed on VQL structures. We may define syntax that corresponds with the above semantics in the following fashion, by introducing a deduction update modality:

Definition 5.2.144. *Define* $Pre : \mathcal{R}(\Phi) \to \mathcal{L}_0(\Phi)$ *as follows:*

$$Pre(\rho) := \bigwedge_{\psi \in \Gamma} I\psi \wedge L\rho$$

Where Γ *is the set of premises of the rule* ρ*.*

(1)	All propositional tautologies	
(2)	$\vdash \Box(\varphi \to \psi) \to \Box\varphi \to \Box\psi$	*Axiom K*
(3)	$\vdash \Box\varphi \to \varphi$	
(4)	$\vdash \Box\varphi \to \Box\Box\varphi$	*R is an equivalence relation*
(5)	$\vdash \Diamond\varphi \to \Box\Diamond\varphi$	
(6)	$\vdash I\gamma \to \Box I\gamma$	*If* $\gamma \in Y(w)$ *and* wRv *then* $\gamma \in Y(v)$
(7)	$\vdash L\rho \to \Box L\rho$	*If* $\rho \in Z(w)$ *and* wRv *then* $\rho \in Z(v)$
(8)	$\vdash I\gamma \to \gamma$	*If* $\gamma \in Y(w)$ *then* $\mathbb{V}, w \Vdash_{VQL} \gamma$
(9)	$\vdash L\rho \to \mathrm{TR}(\rho)$	*If* $\rho \in Z(w)$ *then* $\mathbb{V}, w \Vdash_{VQL} \mathrm{TR}(\rho)$
(I)	$\dfrac{\vdash \varphi \to \psi \qquad \vdash \varphi}{\vdash \psi}$	*Modus Ponens*
(II)	$\dfrac{\vdash \varphi}{\vdash \Box\varphi}$	*Necessitation*

Table 5.1: Axioms for VQL

(D1)	$\vdash \langle D_\rho \rangle \top \leftrightarrow Pre(\rho)$	
(D2)	$\vdash \langle D_\rho \rangle p \leftrightarrow Pre(\rho) \wedge p$	
(D3)	$\vdash \langle D_\rho \rangle \neg\varphi \leftrightarrow Pre(\rho) \wedge \neg\langle D_\rho \rangle \varphi$	
(D4)	$\vdash \langle D_\rho \rangle (\varphi \vee \psi) \leftrightarrow \langle D_\rho \rangle \varphi \vee \langle D_\rho \rangle \psi$	
(D5)	$\vdash \langle D_\rho \rangle \Diamond\varphi \leftrightarrow Pre(\rho) \wedge \Diamond\langle D_\rho \rangle \varphi$	
(D6)	$\vdash \langle D_\rho \rangle I\gamma \leftrightarrow Pre(\rho)$	if γ is the conclusion of ρ
(D7)	$\vdash \langle D_\rho \rangle I\gamma \leftrightarrow Pre(\rho) \wedge I\gamma$	if γ is **not** the conclusion of ρ
(D8)	$\vdash \langle D_\rho \rangle L\sigma \leftrightarrow Pre(\rho) \wedge L\sigma$	
(DI)	$\dfrac{\vdash \varphi}{\vdash [D_\rho]\varphi}$	Necessitation

Table 5.2: Reduction Axioms for $\langle D_\rho \rangle \varphi$

Definition 5.2.145. *The grammar $\mathcal{VQLD}(\Phi)$ extends $\mathcal{VQL}(\Phi)$ and is given as follows:*

$$\varphi ::= p \in \Phi \mid \top \mid \varphi \vee \psi \mid \neg\varphi \mid \Diamond\varphi \mid I\gamma \mid L\rho \mid \langle D_\rho \rangle \varphi$$

As before, $\gamma \in \mathcal{L}_0(\Phi)$, while $\rho \in \mathcal{R}(\Phi)$. Along with the usual abbreviations, let $[D_\rho]\varphi$ abbreviates $\neg\langle D_\rho \rangle \varphi$.

We have the following semantics of the above syntax:

Definition 5.2.146. *We may extend $\Vdash_{VQL}$ to the full $\mathcal{VQLD}(\Phi)$ grammar by stipulating:*

$$\mathbb{V}, w \Vdash_{VQL} \langle D_\rho \rangle \varphi \iff \mathbb{V}, w \Vdash_{VQL} Pre(\rho) \ \& \ \mathbb{V}_\rho, w \Vdash_{VQL} \varphi$$

Every formula in $\mathcal{VQLD}(\Phi)$ is equivalent to a formula in $\mathcal{VQL}(\Phi)$. This can be seen by observing a series of reduction axioms, exhibited in Table 5.2.

Below is a central result present in [Vel09]:

Definition 5.2.147. *Let*

$$\Gamma \vdash_{VQL} \varphi$$

If and only if there is some $\Delta \subseteq_\omega \Gamma$ such that $\bigwedge \Gamma \rightarrow \varphi$ is derivable using the axioms and rules present in Tables 5.1 and 5.2

Theorem 5.2.148 (VQL Strong Soundness and Completeness)**.**

$$\Gamma \vdash_{VQL} \varphi \iff \Gamma \Vdash_{VQL} \varphi$$

Proof. This is Theorem 3 of [Vel09]. QED

In [Vel09] extends the above to discuss logical dynamics, where the agent can become aware of new rules.

To summarize, VQL is a logic where inference is modeled using update modalities, and where one is aware of a proposition only if it is a fact they implicitly know. The only rules that an agent is permitted to be aware of are sound. The agent can infer new things that they already implicitly knew only after an update.

There is at least two points of commonality between VQL and EviL. Perhaps the most notable is that both frameworks assume that agents can only perform *sound* reasoning. Likewise, both present a notion of implicit knowledge: in EviL, implicit or background knowledge can be interpreted as the universal modality U presented in §3.10, which embodies the background theory of the model that the multiverse where we are modeling EviL agents to dwell; the $\Box$ modality plays an analogous rôle in VQL.

On the other hand, there are more ways in which EviL is different from VQL then there are ways the two are similar. In VQL, agents can only be aware of things which they already implicitly knew. In EviL, agents can be fallible in their beliefs, and believe everything they implicitly know (this follows from $\vdash U\varphi \rightarrow \Box\varphi$). Likewise, the notion of knowledge as "the existence of a sound argument" is very separate from *implicit knowledge*. In EviL, implicit knowledge $U\varphi$ does not imply the existence of a sound argument, for perhaps the EviL agent has not one single true assumption to employ as a premise in any argument. Neither is the converse true, since holding a sound argument for φ in a particular situation does not mean that it is true in every conceivable universe.

In EviL, agents are assumed to be logically omniscient, with all of the deductive powers of EviL itself at their disposal. In EviL, we the modelers must surrender some of our control over how EviL agents are reasoning because of this. In the case of VQL, the agent is modeled as making inferences using very explicit syntax. This allows VQL far more fine grained control over the inferences agents make than EviL, which is *laissez faire* in comparison. It is herein that the key difference between the approaches is apparent: in EviL one can reason intuitively about what sort of conclusions agents must arrive at, while in VQL we can only conclude that the agent will arrive at the sound conclusions we allow them to conclude. In VQL, the agent is at the mercy of the modeler, and in EviL it is more like the modeler is at the mercy of the EviL agents themselves.

Finally, VQL and EviL draw from different philosophical foundations. In VQL, one takes a conservative perspective on what it means for an agent to know something in line with traditional epistemic logic. As we first saw in §1.2, the EviL perspective is more closely related to foundationalism.

5.3 Dynamic Awareness Logic

Dynamic Awareness Logic (DAL) is the logic introduced in [vBV09]. As in the case of VQL, DAL is an original name we offer; the authors do not offer a name for their work themselves. It can be understood as modification on VQL as we previously discussed and van Benthem's *acts of realization* [vBV09]. Unlike VQL, which is intended to model implicit and explicit knowledge, DAL geared towards providing a framework with fewer restrictions and easy to describe dynamics.

We will focus on single agent DAL, however multi-agent DAL is also introduced in [vBV09].

Likewise, *DAL* also accommodates *action model* updates as introduced in [BMS98]. Since our purpose in investigating this framework is to understand how it accommodates explicit inference, we have chosen not to pursue these topics here.

We begin by reviewing the grammars employed in *DAL*:

Definition 5.3.149. *Define the static awareness language* $\mathcal{AL}(\Phi)$ *as:*

$$\varphi ::= p \in \Phi \mid \neg\varphi \mid \varphi \wedge \psi \mid \Box\varphi \mid I\varphi$$

The dynamic awareness language $\mathcal{DAL}(\Phi)$ *is defined as follows:*

$$\varphi ::= p \in \Phi \mid \neg\varphi \mid \varphi \wedge \psi \mid \Box\varphi \mid I\varphi \mid [+\chi]\varphi \mid [-\chi]\varphi$$

Note that the syntax introduced in [vBV09] also contains *public announcements*, as introduced in [Ger98]. Public announcements correspond to factive updates a model can undergo. Since this is the textbook example of dynamics for modal logic, discussed at length in [vDvdHK07] for instance, we have decided to eschew discussion of it in favor of the more novel dynamics embodied by the awareness update modalities $[+\chi]$ and $[-\chi]$.

The structures employed in this framework are similar to *VQL* structures, but with less structure:

Definition 5.3.150. *We call a structure* $\mathbb{A} = \langle W, R, V, A\rangle$ *is a* ***Kripke Awareness Structure*** *over a language* $\mathcal{L}$ *and set of proposition letters* Φ *when*

- W *is a set of worlds*
- $R \subseteq W \times W$ *is an accessibility relation*
- $V : \Phi \to \wp W$ *is a propositional valuation function*
- $Y : \mathcal{L} \to \wp W$ *is a propositional awareness relation*

Along with the above semantics, we also have two update operations:

Definition 5.3.151. *Let* $\mathbb{A} = \langle W, R, V, Y\rangle$ *be a Kripke awareness structure.*

Define $\mathbb{A}_{+\chi} := \langle W, R, V, Y'\rangle$ *where*

$$Y'(\varphi) := \begin{cases} W & \text{if } \varphi = \chi \\ Y(\varphi) & \text{o/w} \end{cases}$$

Define $\mathbb{A}_{-\chi} := \langle W, R, V, Y'\rangle$ *where*

$$Y'(\varphi) := \begin{cases} \varnothing & \text{if } \varphi = \chi \\ Y(\varphi) & \text{o/w} \end{cases}$$

These two operations correspond to adding and subtracting the awareness of χ at all worlds. This allows us to present semantics for the language $\mathcal{DAL}(\Phi)$ as follows:

$\vdash [+\chi]p$	$\leftrightarrow$	p	$\vdash [-\chi]p$	$\leftrightarrow$	p
$\vdash [+\chi]I\chi$	$\leftrightarrow$	$\top$	$\vdash [-\chi]I\chi$	$\leftrightarrow$	$\bot$
$\vdash [+\chi]I\varphi$	$\leftrightarrow$	$I\varphi \quad \text{for } \varphi \neq \chi$	$\vdash [-\chi]I\varphi$	$\leftrightarrow$	$I\varphi \quad \text{for } \varphi \neq \chi$
$\vdash [+\chi]\neg\varphi$	$\leftrightarrow$	$\neg[+\chi]\varphi$	$\vdash [-\chi]\neg\varphi$	$\leftrightarrow$	$\neg[-\chi]\varphi$
$\vdash [+\chi](\varphi \wedge \psi)$	$\leftrightarrow$	$[+\chi]\varphi \wedge [+\chi]\psi$	$\vdash [-\chi](\varphi \wedge \psi)$	$\leftrightarrow$	$[-\chi]\varphi \wedge [-\chi]\psi$
$\vdash [+\chi]\Box\varphi$	$\leftrightarrow$	$\Box[+\chi]\varphi$	$\vdash [-\chi]\Box\varphi$	$\leftrightarrow$	$\Box[-\chi]\varphi$
$\vdash \varphi$	$\Longrightarrow$	$\vdash [+\chi]\varphi$	$\vdash \varphi$	$\Longrightarrow$	$\vdash [-\chi]\varphi$

Table 5.3: Reduction Axioms for DAL

Definition 5.3.152. *DAL's truth predicate $\Vdash_{DAL}$ is defined such that:*

$$\begin{aligned}
\mathbb{A}, w \Vdash_{DAL} p &\iff w \in V(p) \\
\mathbb{A}, w \Vdash_{DAL} \neg\varphi &\iff \mathbb{A}, w \nVdash_{DAL} \varphi \\
\mathbb{A}, w \Vdash_{DAL} \varphi \wedge \psi &\iff \mathbb{A}, w \Vdash_{DAL} \varphi \quad \& \quad \mathbb{A}, v \Vdash_{DAL} \varphi \\
\mathbb{A}, w \Vdash_{DAL} \Box\varphi &\iff \forall v \in W. wRv \Longrightarrow \mathbb{A}, v \Vdash_{DAL} \varphi \\
\mathbb{A}, w \Vdash_{DAL} I\varphi &\iff \varphi \in Y(w) \\
\mathbb{A}, w \Vdash_{DAL} [+\chi]\varphi &\iff \mathbb{A}_{+\chi}, w \Vdash_{DAL} \varphi \\
\mathbb{A}, w \Vdash_{DAL} [-\chi]\varphi &\iff \mathbb{A}_{-\chi}, w \Vdash_{DAL} \varphi
\end{aligned}$$

In [vBV09], the authors do not provide a calculus for DAL and exhibit a completeness theorem. Instead, they employ the observation that every formula in $\mathcal{DAL}(\Phi)$ is logically equivalent to a formula in $\mathcal{AL}(\Phi)$ for all DAL Kripke structures, and provide reduction axioms. Hence when reasoning about this system one can always reason about the static language without loss of generality to the dynamic language.

Theorem 5.3.153. *The valid formulas of the dynamic-epistemic awareness language $\mathcal{DAL}(\Phi)$ are the valid formulae of the static base language $\mathcal{AL}(\Phi)$ plus the reduction axioms and modal inference rules listed in Table 5.3.*

Proof. This is Theorem 1 in [vBV09]. QED

The authors in [vBV09] an intuitive motivation of how their framework is intended to model explicit knowledge. Informally, we will in this setting we will want to employ the following readings

- $\Box\varphi$ asserts "the agent *implicitly* knows that φ."
- $K\varphi$ asserts "the agent *explicitly* knows that φ."

For an agent that is not logically omniscient, we want to enforce that the following is **not** valid:

$$K(\varphi \to \psi) \to K\varphi \to K\psi \tag{5.3.1}$$

However, while the above does not hold for DAL, the authors in [vBV09] suggest that the following

two expressions should hold:

$$\vdash K(\varphi \to \psi) \to K\varphi \to \Box\psi$$
$$\vdash K(\varphi \to \psi) \to K\varphi \to [?]K\psi$$

Here [] is an inference *gap*. The idea here is that agents are thought to implicitly know the logical conclusions of all of their explicit knowledge, while to explicitly draw a conclusion they need a nudge.

The authors propose the following definitions, and illustrate that they are sufficient to ensure the above criteria:

$$\begin{aligned} K\varphi &:= \Box(\varphi \wedge I\varphi) \\ [?]K\psi &:= [+\psi]K\psi \end{aligned}$$

As a side remark, the gap [?] is defined to be sensitive to the context in which is it employed.

The proposed definition for explicit knowledge also renders (5.3.1) for VQL. In addition, the proposed EvIL formulation of knowledge $K_{\text{EvIL}}\varphi := \Diamond(\circlearrowleft \wedge \Box\varphi)$ also invalidates (5.3.1), despite the fact that agents are modeled as logically omniscient. This is due to the fact that the EvIL agent may have a sound argument for $\varphi \to \psi$ and another, entirely different sound argument for φ, but she might not be clever enough to join the bases for these two sound arguments together to argue that ψ. It is straightforward to compose a concrete EvIL model that formalizes this intuition. Failure of explicit logical omniscience is a unifying principle behind all formulations of explicit knowledge.

While VQL naturally accommodates DAL's formulation of explicit knowledge, it cannot accommodate DAL's gap [?]. DAL has a richer internal language that is unavailable to VQL. While VQL supports dynamics for adding awareness of formulae, the agent can only be aware of $\mathcal{L}_0(\Phi)$ formulae. This grammar restriction prohibits formulating the dynamics above.

The gap [?] is also incompatible with EvIL. As previously noted, EvIL is restricted to $\mathcal{L}_0$ in a manner similar to VQL. Another issues is that the abstract EvIL semantics makes no mention of anything analogous to explicit awareness, and the concrete semantics, while superficially similar, is sensitive to dynamics in a way that awareness approaches are not. For instance, in the concrete EvIL semantics we evidently have that:

$$(a, \{p\}) \not\sqsubseteq (a, \{q\}).$$

On the other hand, we have that

$$(a, \{p\} \cup \{p\}) \sqsubseteq (a, \{q\} \cup \{p\}).$$

This poses a problem for trying to formulate an EvIL formulation of $[+\chi]$. Updates conventionally do not *add* accessibility, while in the concrete semantics they would have to in this situation. Given this observation, it is straightforward to contrive a scenario in the concrete EvIL semantics which invalidates the following candidate reduction axiom:

$$[+\chi] \boxplus \varphi \leftrightarrow \boxplus[+\chi]\varphi$$

The above reveals a tension between the dynamic approach and the EvIL framework. In EvIL concrete models, two worlds are related if their beliefs and truth conditions make true various logical relationships, while in a dynamic approach two worlds are related if they were initially related, and through various manipulations to the model this relationship has been preserved.

As a final remark concerning DAL, when one writes $\mathbb{A}, w \vdash K\varphi$, it is not obvious what sort of reasons the agent might have for knowing φ in this circumstance. The story in VQL is that the agent used valid inference rules they were aware of, along with various model updates to arrive at φ. As we have already presented, EvIL has its own way of analyzing the reasoning behind knowledge agents might or might not have. However, no such narrative appears to be evident in the case of DAL. However, given the expressive power of DAL, perhaps DAL will be able to embed other logics of explicit inference. The character of DAL is that it is general, and non-specific regarding the way that it models explicit knowledge, so understanding its precise connection to other approaches seems a promising subject for future research.

5.4 Final Thoughts

The purpose of all of the logics we have presented in this section is to model an agent making explicit inferences behind their deductions. Awareness based approaches tackle this by modeling the mental contents of the agent's mind. At a glance, this appears superficially similar to the EvIL approach. However, at their core, these two approaches are essentially orthogonal.

To summarize the EvIL approach, it is to enforce that beliefs correspond to deductions, and then provide modalities for controlling the bases upon which those deductions are made. It accomplishes this by introducing concrete structures with modal syntax, and then abstracting on those concrete structures with abstract Kripke semantics.

The cost of this approach, however, is that the intended semantics for controlling the bases for inference are not obviously compatible with the intended semantics for updates.

However, there are benefits to the EvIL approach that outweigh the costs. EvIL incorporates a clear foundationalist epistemology into its design, and naturally models deliberative actions one may take in a foundationalist setting. This foundationalist perspective on belief and knowledge clear intuitionistic readings, and this reflected in embedding. However, via *duality*, EvIL goes further and introduces its own reading of intuitionistic semantics, where instead imagination takes the place of knowledge as providing truth conditions.

We hope that the future will allow for fruitful communication between the ideas in EvIL and its allied approaches.

Appendix A

Alternate Semantics

In this section, we shall present an alternative work to the framework proposed in §1.2. These semantics are inspired by game semantics for modal logic, such as those in [vB10], chapter 2.

First, recall the basic modal grammar $\mathcal{L}_K(\Phi)$:

$$\varphi ::= p \in \Phi \mid \varphi \to \psi \mid \bot \mid \Box\varphi$$

Next, consider structures of the form $\langle W, V, \beta, \iota\rangle$ consisting of:

- A set of worlds W
- A propositional valuation function $V : \Phi \to \wp W$
- An belief function $\beta : W \to \wp\mathcal{L}_K(\Phi)$
- An imagination function $\iota : W \to \wp W$

We shall call these *belief-imagination models*. One can think of a model $\mathfrak{M}$ sort of like a of tuples like in §2; however in this case evidently it would have to be $\mathfrak{M} \subseteq \wp\Phi \times \wp\mathcal{L}_K(\Phi) \times \wp\mathfrak{M}$, so apparently it would have to be a non-wellfounded set. This is somewhat natural, given a modal logic setting - see for instance [BM96] for an elaboration on these connections.

Definition 1.0.154. *Define by recursion the following two truth relations:*

First relation:

$$\mathfrak{M}, w \Vdash p \iff p \in V(w)$$

$$\mathfrak{M}, w \Vdash \varphi \wedge \psi \iff \textit{both } \mathfrak{M}, w \Vdash \varphi \textit{ and } \mathfrak{M}, w \Vdash \psi$$

$$\mathfrak{M}, w \Vdash \varphi \vee \psi \iff \textit{either } \mathfrak{M}, w \Vdash \varphi \textit{ or } \mathfrak{M}, w \Vdash \psi$$

$$\mathfrak{M}, w \Vdash \neg\varphi \iff \mathfrak{M}, w \nVdash \varphi$$

$$\mathfrak{M}, w \Vdash \Box\varphi \iff \beta(w) \vdash^* \varphi$$

Where $\vdash^*$ is a sequent that is closed under reflection and resolution:

$$\frac{\varphi \in \Gamma}{\Gamma \vdash^* \varphi} \qquad \frac{\Gamma \vdash^* \neg\varphi \vee \psi \quad \Delta \vdash^* \varphi}{\Gamma \cup \Delta \vdash^* \psi}$$

Second relation:

$\mathfrak{M}, w \Vdash p \iff p \notin V(w)$

$\mathfrak{M}, w \Vdash \varphi \wedge \psi \iff$ *either* $\mathfrak{M}, w \Vdash \varphi$ *or* $\mathfrak{M}, w \Vdash \psi$

$\mathfrak{M}, w \Vdash \varphi \vee \psi \iff$ *both* $\mathfrak{M}, w \Vdash \varphi$ *and* $\mathfrak{M}, w \Vdash \psi$

$\mathfrak{M}, w \Vdash \neg\varphi \iff \mathfrak{M}, w \nVdash \varphi$

$\mathfrak{M}, w \Vdash \Box\varphi \iff$ *there is some* $v \in \iota(w)$ *such that* $\mathfrak{M}, v \Vdash \varphi$

It is necessary to motivate the intuition behind these semantics. Informally, we think of these two truth relations correspond to two players, whom we shall call the *logician* and the *philosopher*. The logician wields a set beliefs given by β and tries to compose compelling arguments, and the philosopher employs a corpus of thought experiments given by ι to thwart the logician's arguments. Of course, the logician and the philosopher are really just two aspects of a single epistemic agent we are trying to model; we shall imagine epistemic agents modeled by this system to be embroiled in internal conflict. This sort of dissension between reason and imagination rages on within us all – it is fundamental to human nature.

These semantics are not naturally bivalent; that is it does not hold that either $\mathfrak{M}, w \nVdash \varphi$ or $\mathfrak{M}, w \Vdash \varphi$, exclusively. To see this consider a model where $\beta(w) = \iota(w) = \varnothing$; then evidently $\mathfrak{M}, w \nVdash \Box p$ and $\mathfrak{M}, w \Vdash \Box p$.

However, bivalence has a convenient semantic characterization:

Proposition 1.0.155. *Let $\mathbb{M}^{\mathfrak{M}} = \langle W^{\mathfrak{M}}, V^{\mathfrak{M}}, R^{\mathfrak{M}} \rangle$ be a model for basic modal logic model based on a belief/imagination model $\mathfrak{M}$, where $wR^{\mathfrak{M}}v := v \in \iota(w)$, and let $\Vdash_{\Box}$ be the modal truth predicate. We have that $\Vdash$ and $\Vdash$ are bivalent if and only if $\mathfrak{M}, w \Vdash \varphi \iff \mathbb{M}^{\mathfrak{M}}, w \Vdash_{\Box} \varphi$.*

Proof. ($\Longrightarrow$) Assume that $\Vdash$ and $\Vdash$ are bivalent and consider any $\varphi \in \mathcal{L}_K(\Phi)$. The proof that $\mathfrak{M}, w \Vdash \varphi$ is equivalent to $\mathbb{M}^{\mathfrak{M}}, w \Vdash_{\Box} \varphi$ proceeds by induction. The case for proposition letters, conjunction and disjunction are straightforward, so we shall only consider negation and modality.

Negation: We have the following chain of equivalences:

$$\begin{aligned}
\mathbb{M}^{\mathfrak{M}}, w \Vdash_{\Box} \neg\varphi &\iff \mathbb{M}^{\mathfrak{M}}, w \nVdash_{\Box} \varphi \\
&\iff \mathfrak{M}, w \nVdash \varphi \qquad \text{(inductive step)} \\
&\iff \mathfrak{M}, w \Vdash \varphi \qquad \text{(bivalence)} \\
&\iff \mathfrak{M}, w \Vdash \neg\varphi
\end{aligned}$$

Modality: We have another chain of equivalences:

$$\begin{aligned}
\mathfrak{M}, w \Vdash \Box\varphi &\iff \mathfrak{M}, w \nVdash \Box\varphi && \text{(bivalence)}\\
&\iff \forall v \in \iota(w).\mathfrak{M}, w \nVdash \varphi && \text{(definition)}\\
&\iff \forall v \in \iota(w).\mathfrak{M}, w \Vdash \varphi && \text{(bivalence)}\\
&\iff \forall v \in \iota(w).\mathbb{M}^{\mathfrak{M}}, w \Vdash_{\Box} \varphi && \text{(inductive step)}\\
&\iff \forall v. wR^{\mathfrak{M}}v \Longrightarrow \mathbb{M}^{\mathfrak{M}}, w \Vdash_{\Box} \varphi && \text{(definition)}\\
&\iff \mathbb{M}^{\mathfrak{M}}, w \Vdash_{\Box} \Box\varphi
\end{aligned}$$

This completes the induction.

($\Longleftarrow$) Assume that $\mathfrak{M}, w \Vdash \varphi$ and $\mathbb{M}^{\mathfrak{M}}, w \Vdash_{\Box} \varphi$ are always equivalent. We have:

$$\begin{aligned}
\mathfrak{M}, w \Vdash \varphi &\iff \mathbb{M}^{\mathfrak{M}}, w \Vdash_{\Box} \varphi\\
&\iff \mathfrak{M}, w \nVdash_{\Box} \neg\varphi\\
&\iff \mathfrak{M}, w \nVdash \neg\varphi && \text{(hypothesis)}\\
&\iff \mathfrak{M}, w \nVdash \varphi
\end{aligned}$$

QED

Corollary 1.0.156. *If $\Vdash$ and $\nVdash$ are bivalent, then $\beta(w) \vdash^* \varphi$ for all $\varphi \in \mathrm{Th}_{\Vdash}(\mathfrak{M})$ for all $w \in W^{\mathfrak{M}}$, where $\mathrm{Th}_{\Vdash}(\mathfrak{M}) = \{\varphi \in \mathcal{L}_K(\Phi) \mid \mathfrak{M}, w \Vdash \varphi \text{ for all } w \in W^{\mathfrak{M}}\}$.*

Evidently bivalence of $\Vdash$ and $\nVdash$ gives rise to semantics where the agent has a proof for every proposition they believe. Furthermore, we can take any modal logic model $\mathbb{M} := \langle W^{\mathbb{M}}, V^{\mathbb{M}}, R^{\mathbb{M}}\rangle$ and define an equivalent belief/imagination model $\mathfrak{M}^{\mathbb{M}} := \langle W^{\mathbb{M}}, V^{\mathbb{M}}, \beta^{\mathbb{M}}, \iota^{\mathbb{M}}\rangle$ where:

$$\begin{aligned}
\beta^{\mathbb{M}}(w) &:= \{\varphi \in \mathcal{L}_K(\Phi) \mid \mathbb{M}, w \Vdash_{\Box} \Box\varphi\}\\
\iota^{\mathbb{M}}(w) &:= \{v \in W^{\mathbb{M}} \mid wR^{\mathbb{M}}v\}
\end{aligned}$$

We can immediately leverage this to give the a characterization of these semantics:

Proposition 1.0.157. *The basic modal logic K is sound and strongly complete for bivalent belief/imagination models.*

Proof. Soundness is trivial given the previous lemma, strong completeness follows by considering the canonical model $\mathbb{K}$ and looking at $\mathfrak{M}^{\mathbb{K}}$. QED

However, this is evidently not entirely necessary. Call a belief/imagination model *reasonable* if the following two constraints are satisfied:

- $\beta(w) \vdash^* \varphi$ for all $\varphi \in \mathrm{Th}_{\Vdash}(\mathfrak{M})$ for all $w \in W^{\mathfrak{M}}$, where $\mathrm{Th}_{\Vdash}(\mathfrak{M}) = \{\varphi \in \mathcal{L}_K(\Phi) \mid \mathfrak{M}, w \Vdash \varphi \text{ for all } w \in W^{\mathfrak{M}}\}$
- $\mathrm{Mod}^{\mathfrak{M}}_{\nVdash}(\beta(w)) \subseteq \iota(w)$, where $\mathrm{Mod}^{\mathfrak{M}}_{\nVdash}(\beta(w)) = \{v \in W^{\mathfrak{M}} \mid \mathfrak{M}, v \nVdash \varphi \text{ for all } \varphi \in \beta(w)\}$
- $\beta(w) \setminus \mathrm{Th}_{\Vdash}(\mathfrak{M})$ is finite

Evidently, forcing these requirements suffices to force bivalence:

Proposition 1.0.158. *Let $\mathbb{M}^{\mathfrak{M}}$ be defined as in Prop. 1.0.155. For any reasonable model $\mathfrak{M}$ and any $w \in W^{\mathfrak{M}}$, we have:*

(i) If $\mathbb{M}^{\mathfrak{M}}, w \Vdash_{\Box} \varphi$ then $\mathfrak{M}, w \Vdash \varphi$

(ii) If $\mathbb{M}^{\mathfrak{M}}, w \nVdash_{\Box} \varphi$ then $\mathfrak{M}, w \Vdash\!\!\!\Vdash \varphi$

Hence we have $\Vdash$ and $\Vdash\!\!\!\Vdash$ are bivalent.

Proof. The propositional, disjunctive and conjunctive cases are all straightforward; we shall focus on negation and modality.

Negation: In the case of (i), we know that

$$\begin{aligned} \mathbb{M}^{\mathfrak{M}}, w \Vdash_{\Box} \neg\varphi &\iff \mathbb{M}^{\mathfrak{M}}, w \nVdash_{\Box} \varphi \\ &\implies \mathfrak{M}, w \Vdash\!\!\!\Vdash \varphi \quad \text{(by the inductive step)} \\ &\iff \mathfrak{M}, w \Vdash \neg\varphi \end{aligned}$$

The proof for (ii) is similar.

Modality: In the case of (i), assume that $\mathbb{M}^{\mathfrak{M}}, w \Vdash_{\Box} \Box\varphi$. Using the definition of reasonableness and the inductive step we know for all $v \in W^{\mathfrak{M}}$ that if $\mathfrak{M}, v \nVdash \psi$ for all $\psi \in \beta(w) \setminus \mathrm{Th}(\mathfrak{M})$ then $\mathfrak{M}, v \Vdash \varphi$.

From this and the fact that $\mathfrak{M}$ is reasonable we can infer that $\bigvee_{\psi \in \beta(w) \setminus \mathrm{Th}(\mathfrak{M})} \neg\psi \vee \varphi \in \mathrm{Th}_{\Vdash}(\mathfrak{M})$. We know further from reasonableness that we have $\mathrm{Th}_{\Vdash}(\mathfrak{M}) \subseteq \beta(w)$. So we can prove by induction that repeatedly applying resolution gets $\beta(w) \vdash^{*} \varphi$, which just means that $\mathfrak{M}, w \Vdash \Box\varphi$, as desired.

The case of (ii) follows trivially by induction. QED

We may continue to obtain weak completeness for these semantics:

Proposition 1.0.159. *$\vdash_K \varphi$ if and only if $\mathfrak{M}, w \Vdash \varphi$ for all reasonable models $\mathfrak{M}$ and $w \in W^{\mathfrak{M}}$*

Proof. Left to right follows straightforwardly, so we just need to prove right to left.

Assume $\nvdash_K \varphi$. As before, let $\mathbb{M} = \langle W^{\mathbb{M}}, V^{\mathbb{M}}, R^{\mathbb{M}} \rangle$ be a finite model and with a world $w \in W^{\mathbb{M}}$ such that $\mathbb{M}, w \nVdash_{\Box} \varphi$. Now consider a slightly modified model $\mathbb{M}' := \langle W^{\mathbb{M}}, V', R^{\mathbb{M}} \rangle$ where

$$V'(p) := \begin{cases} \{v\} & p = \rho(v) \\ V(p) & o/w \end{cases}$$

A proof by induction on subformulae ψ of φ verifies that $\mathbb{M}, w \Vdash_{\Box} \psi$ if and only if $\mathbb{M}', w \Vdash_{\Box} \psi$.

So now consider $\mathfrak{M} := \langle W^{\mathbb{M}'}, V^{\mathbb{M}'}, \tau, \lambda x. R^{\mathbb{M}'}[x] \rangle$ such that

$$\tau(w) := \mathrm{Th}(\mathbb{M}') \cup \left\{ \bigvee_{v \in R^{\mathbb{M}'}[w]} \rho(v) \right\},$$

where $\mathrm{Th}(\mathbb{M}') := \{\psi \in \mathcal{L}_K(\Phi) \mid \mathbb{M}', v \Vdash \psi \text{ for all } v \in W^{\mathbb{M}'}\}$. A proof by induction on ψ shows that $\mathbb{M}', w \Vdash_{\Box} \psi$, $\mathfrak{M}, w \Vdash \psi$ and $\mathfrak{M}, v \nVdash \psi$ are equivalent for all $\psi \in \mathcal{L}_K(\Phi)$. Thus we have that for all $v \in W^{\mathfrak{M}}$ that $\mathfrak{M}, v \nVdash \psi$ for all $\psi \in \mathrm{Th}(\mathbb{M}')$. Moreover, evidently $wR^{\mathbb{M}'}v$ if and only if $\mathbb{M}', v \Vdash_{\Box} \bigvee_{u \in R^{\mathbb{M}'}[w]} \rho(u)$, whence we have that $wR^{\mathbb{M}'}v$ if and only if $\mathfrak{M}, v \nVdash \chi$ for all $\chi \in \tau(w)$. With this we can employ induction and establish that $\mathbb{M}', w \Vdash_{\Box} \psi$ if and only if $\mathfrak{M}, w \models \psi$ for all $\psi \in \mathcal{L}_K(\Phi)$. Since $\mathbb{M}', w \nVdash_{\Box} \varphi$, we have that $\mathfrak{M}, w \not\models \varphi$. Finally, note that in this model we have that $\mathrm{Mod}^{\mathfrak{M}}_{\nVdash}(\beta(w)) = R^{\mathbb{M}'}[w]$. With this and the definition of $\mathfrak{M}$, we can see that $\mathfrak{M}$ is evidently reasonable, and thus we may complete the proof.

QED

Now, while reasonable models attain the goal of modeling agents that have proofs for the things they believe, they should not be considered adequate. These models are only reasonable in the sense that they indeed model agents providing nontrivial proofs for their beliefs. However, they are not reasonable in the sense that they are simple to reckon with. So while the semantics provided in §2 requires a grammar restriction, it should be preferred over the formulation given above, precisely because it is more manageable.

Bibliography

[AM94] Michael Francis Atiyah and Ian Grant Macdonald. *Introduction to commutative algebra*. Westview Press, February 1994.

[AN05] S. Artemov and E. Nogina. Introducing justification into epistemic logic. *Journal of Logic and Computation*, 15(6):1059, 2005.

[Art07] S. N Artemov. Justification logic. *CUNY Graduate Center, New York*, 2007.

[Awo06] Steve Awodey. *Category theory*. Oxford University Press, 2006.

[Bel86] Eric Temple Bell. *Men of mathematics*. Simon and Schuster, 1986.

[BM96] Jon Barwise and Lawrence Stuart Moss. *Vicious circles*. CSLI Publications, 1996.

[BMS98] Alexandru Baltag, Larry Moss, and S. Solecki. The logic of public announcements, common knowledge, and private suspicions. In *Proceedings of the 7th conference on Theoretical aspects of rationality and knowledge*, page 56, 1998.

[Boo95] George Boolos. *The logic of provability*. Cambridge University Press, 1995.

[BRV01] P. Blackburn, M. De Rijke, and Y. Venema. *Modal logic*. Cambridge Univ Pr, 2001.

[CF04] Earl Conee and Richard Feldman. *Evidentialism*. Oxford University Press, USA, June 2004.

[CZ97] Alexander Chagrov and Michael Zakharyaschev. *Modal logic*. Oxford University Press, 1997.

[DeP01] Michael Raymond DePaul. *Resurrecting old-fashioned foundationalism*. Rowman & Littlefield, 2001.

[DP02] B. A. Davey and Hilary A. Priestley. *Introduction to lattices and order*. Cambridge University Press, 2002.

[Duc95] H. N. Duc. Logical omniscience vs. logical ignorance on a dilemma of epistemic logic. *Progress in Artificial Intelligence*, page 237–248, 1995.

[Duc97] H. N. Duc. Reasoning about rational, but not logically omniscient, agents. *Journal of Logic and Computation*, 7(5):633, 1997.

[Duc01] H. N. Duc. Resource-bounded reasoning about knowledge. *Unpublished doctoral dissertation, Faculty of Mathematics and Informatics, University of Leipzig*, 2001.

[FH87] R. Fagin and J. Y Halpern. Belief, awareness, and limited reasoning* 1. *Artificial Intelligence*, 34(1):39–76, 1987.

[Fit04] M. Fitting. A logic of explicit knowledge. *Logica Yearbook*, page 11–22, 2004.

[Fit05] M. Fitting. The logic of proofs, semantically. *Annals of Pure and Applied Logic*, 132(1):1–25, 2005.

[Fon08] Gaëlle Fontaine. Continuous fragment of the mu-Calculus. In *Computer Science Logic*, pages 139–153. 2008.

[Fri06] J. Friedl. *Mastering regular expressions*. O'Reilly Media, Inc., 2006.

[Ger98] J. Gerbrandy. Bisimulations on planet kripke; ILLC dissertation series. *Institute for Logic, Language and Computation, Universiteit van Amsterdam*, 1998.

[Hin69] Jaakko K. Hintikka. *Knowledge and Belief*. Cornell Univ. Pr., 1969.

[Jag06] M. Jago. *Logics for resource-bounded agents*. PhD thesis, Dissertation, Department of Philosophy, University of Nottingham, 2006.

[Lev84] H. J Levesque. A logic of implicit and explicit belief. In *Proceedings of the National Conference on Artificial Intelligence*, pages 198–202, 1984.

[MvdH95] John-Jules Ch Meyer and Wiebe van der Hoek. *Epistemic Logic for AI and Computer Science*. Cambridge University Press, 1995.

[Pla98] Plato. *The Republic*. October 1998. LoC Class PA: Language and Literatures: Classical Languages and Literature.

[Put81] H. Putnam. A problem about reference. *Reason, Truth and History*, page 22–48, 1981.

[SR99] Jonathan D. H. Smith and Anna B. Romanowska. *Post-modern algebra*. Wiley-IEEE, 1999.

[US06] Pawel Urzyczyn and Morten Heine S?rensen. *Lectures on the Curry-Howard Isomorphism*. Elsevier,, Burlington :, 2006.

[vB91] Johan van Benthem. Reflections on epistemic logic. *Logique Anal., Nouv. Sér.*, 34(133-134):5–14, 1991.

[vB96] Johan van Benthem. *Exploring logical dynamics*. CSLI Publications, 1996.

[vB08] Johan van Benthem. Merging observation and access in dynamic epistemic logic. *Studies in Logic*, 1:1–1, 2008.

[vB09] Johan van Benthem. The information in intuitionistic logic. *Synthese*, 167(2):251–270, 2009.

[vB10] Johan van Benthem. *Modal Logic for Open Minds*. Center for the Study of Language and Information, February 2010.

[vBE89] Johan van Benthem and J. Fernandez-Prida H.-D. Ebbinghaus. Semantic parallels in natural language and computation. In *Logic Colloquium'87, Proceedings of the Colloquium held in Granada*, volume Volume 129, pages 331–375. Elsevier, 1989.

[vBV09] Johan van Benthem and F. R Velázquez-Quesada. Inference, promotion, and the dynamics of awareness. *ILLC Amsterdam. To appear in Knowledge, Rationality and Action*, 2009.

[vDvdHK07] Hans van Ditmarsch, Wiebe van der Hoek, and Barteld P. Kooi. *Dynamic Epistemic Logic*. Springer, 1 edition, November 2007.

[Vel09] F. R Velázquez-Quesada. Inference and update. *Synthese*, 169(2):283–300, 2009.

[VMTD05] John Vietch, David B. Manley, Charles S. Taylor, and René Descartes. Descartes' meditations. http://www.wright.edu/cola/descartes/, July 2005.