Evidentialist Logic

Matthew P. Wampler-Doty

Contents

1 Philosophy

1.1 Forward

The idea of applying modal logic to the study of knowledge more or less began with [Hin69], *Knowledge and Belief*, by. In this it is suggested that one can use the possible world framework of modal logic to model ideal logical agents, and reason about concepts like knowledge and belief as modal boxes. In Hintikka's text, some philosophical emphasis is put on the ideas of *introspection*, which have two formulations:

- Positive: $\Box\varphi \rightarrow \Box\Box\varphi$ - "If the agent knows a fact, then she knows that she knows this."
- Negative: $\Diamond\varphi \rightarrow \Box\Diamond\varphi$ - "If the agent does not know a fact, she knows that she does not know this."

Intuitively, the second idea seems like something one ought to reject outright. Many will recall the somewhat famous piece of sophistry put forward by former US secretary of defense Donald Rumsfeld [Rum02]:

> Reports that say that something hasn't happened are always interesting to me, because as we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns – the ones we don't know we don't know. And if one looks throughout the history of our country and other free countries, it is the latter category that tend to be the difficult ones.

This quote was ultimately part of a rather hideous justification for criminal military action. Still, it is undeniably at variance with negative introspection; and despite its malicious intent, it is compelling. Like Rumsfeld, Hintikka also rejects negative introspection. Furthermore, it would seem from the above quote that Rumsfeld does not reject positive introspection. Hintikka does not either, and goes one step further to endorse it explicitly.

Despite philosophical objections, the received view in modern epistemic logic embraces both negative and positive introspection. In addition, the following axiom is also embraced:

- Reflection: $\Box\varphi \rightarrow \varphi$ - "If the agent knows a some statement, that statement is true"

These three axioms together, along with the axioms and rules of elementary modal logic, form C.I. Lewis' system $S5$ [LL51]. Under correspondence theory, these axioms express that the underlying modal accessibility relation is an equivalence relation. That is, they express that the ideal agent under investigation has partitioned her state space into *information states*. It is well known that game theory shares an equivalent notion of information states (see, for instance, [Hal99] and [Rub98], chapter 3).

Although this view of knowledge has been the focus of exhaustive research, even finding industrial application such as in [AHV02] and [HMdV05, HMV04], it is natural for a practitioner to hold lingering philosophical concerns. The purpose of this thesis is to present a framework which tries to address some of these issues. It shall conform to the following structure:

§1 First, we shall elaborate the philosophical issues that will be addressed, and sketch an epistemic logic framework which tackles them

§2 Next, we give formal details of the system we will develop. This section climaxes in the exposition of a completeness theorem.

§3 Third, we shall look at philosophical applications of the framework developed. It shall be revealed that *intuitionistic logic* is profoundly linked to the perspective on epistemic logic we shall investigate here.

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

That being said, we shall now turn to investigating the philosophical issues with epistemic logic, and suggest a manner of addressing them.

1.2 Thermometers

Imagine a 1 m^3 box with a thermometer sealed hermetically inside, as in Fig. 1. Further, pretend that the thermometer reads 290 Kelvin. How many moles of gas are in the chamber?

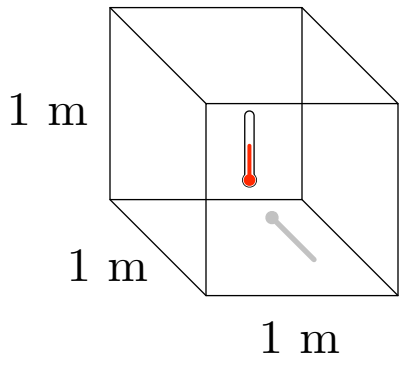


Figure 1: A thermometer in a box

The answer is indeterminate. Recall the *ideal gas law*, originally discovered by Émile Clapeyon [Cla34], which in modern parlance it reads:

$$PV = nRT$$

Where:

- P is the pressure in Pascals
- V is the volume in cubic meters
- n is the number of moles of gas
- T is the temperature in Kelvins

- R is the *ideal gas constant*, ≈ 8.3 J·K^{-1}·mol^{-1}

Applying the ideal gas law, one can observe that the thermometer is effectively in something analogous to an epistemic space. To be explicit, consider a basic modal language with the following grammar $\mathcal{L}_{\text{therm}}$:

$$\varphi ::= x \text{ Pascals} \mid y \text{ moles} \mid z \text{ Kelvin} \mid \varphi \to \psi \mid \bot \mid \Box\varphi$$

Interpreting this language appropriately, the thermometer is evidently an epistemic agent in an $S5$ model. One model for the thermometer-in-a-box is the triple $\langle W, V, \sim \rangle$, where:

- W is pairs (P, n) where P is some positive pressure in Pascals and n is some positive number of moles.
- V is defined as follows:
  - $(P, n) \in V(x \text{ Pascals})$ if $P = x$
  - $(P, n) \in V(y \text{ moles})$ if $n = y$
  - $(P, n) \in V(z \text{ Kelvin})$ if $z = \frac{P}{n \cdot R}$
- Finally, $(P, n) \sim (P', n')$ holds if and only if $P \cdot n' = P' \cdot n$

2 provides a visual representation of the information states in the above model, which form rays emanating from the origin.

The view in philosophy of mind that thermometers are epistemic agents originates Daniel Dennet's *The Intentional Stance* [Den98], with the proviso that Dennett's original discussion originates around thermostats rather than thermometers as we have argued. Dennett writes:

> *It is not that we attribute (or should attribute) beliefs and desires only to things in which we find internal representations, but rather that when we discover some object for which the intentional strategy works, we endeavor to interpret some of its internal states or processes as internal representations. What makes some internal feature of a thing a representation could only be its role in regulating the behavior of an intentional system.*
>
> Now the reason for stressing our kinship with the thermostat should be clear. There is no magic moment in the transition from a simple thermostat to a system that really has an internal representation of the world around it. The thermostat has a minimally demanding representation of the world, fancier thermostats have more demanding representations of the world, fancier robots for helping around the house would have still more demanding representations of the world. Finally you reach us.

The aim of epistemic logic is to model agents modeling the world; and in doing so its development mirrors the increasing levels of complexity that Dennett illustrates. To give an exemplar of the modern level of sophistication achieved by epistemic logic, consider probablistic dynamic epistemic logic as developed in [vBGK09].

Moreover, just as thinking about thermostats serves as a vehicle for philosophy of mind for Daniel Dennett, thinking about thermometers can elucidate intuitions behind basic epistemic logic, and

how it might be extended. Imagine we were to go up to one of the agents modeled in basic epistemic logic, and ask her why she knows some proposition φ. What would she possibly say? She would say she feels φ with every fiber of her being, that it is true in every world she can conceive. The reason that φ occurs to the agent is because it is what her sensory instruments (modeled as her accessibility relations) tell her. In this respect, the analogy of the thermometer seems pretty apt.

Some natural philosophical features of knowledge cannot be captured by this basic approach. If we were to ask a person on the street why she believes a proposition φ, an appeal that she cannot imagine or conceive of the contrary as possible would not slake our curiousity. We typically want some kind of *explanation*, especially if φ were a piece of mathematics, for instance. It would certainly make the enterprise of mathematics far simpler if proving theorems amounted to exhibiting that their negation is not imaginable. This gives rise to the following philosophical observation:

> **Anti-Thermometer Principle**: *Traditional epistemic agents, like thermometers, don't always have knowledge, since one must sometimes have reasons for the things they believe.*

How might epistemic logic be saved by the objection that the above principle proposes? To find out, we turn to diagnosing the underpinning of the above principle.

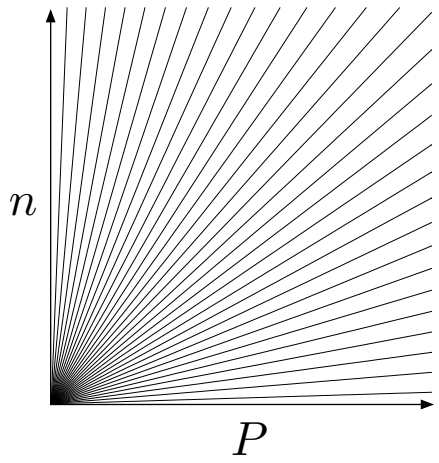


Figure 2: Thermometer information states

1.3 Explicit Justification

I hold that the *Anti-Thermometer Principle* is an expression of the following more fundamental idea:

> **Justification Principle**: *In order to know something, one must sometimes demonstrate inferential justification.*

Demanding the *Justification Principle* is tantamount to demanding a sort of *explicit justification* for beliefs in epistemic logic. This particular desiderata been suggested previously. 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].

In this text we shall investigate a novel framework in this line of research, in the hopes of offering further response that practitioners of epistemic logic may exhort in answer to the problems raised by the aforementioned *justification principle*. To model beliefs with justifications, we shall modify the semantics of modal logic to incorporate certain *basic beliefs*, which we should interpret as non-inferentially justified. These basic beliefs then inferentially generate the rest of what the agent believes.

This perspective amounts to what is called *classical 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]

The view presented would not deny that thermometers or traditional epistemic agents have knowledge, no more than it would deny that one has knowledge of that her right hand has five fingers attached to the end. Rather, our aim is to try to modify the semantics of epistemic logic, more specifically *doxastic logic*, with the ingredients for a foundationalist analysis of knowledge. As will be demonstrated, this can be done without modifying the basic modal logic syntax.

1.4 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 §2.1.1. 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:

[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.

Definition 1.4.1.

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

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

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

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

Since the semantics like the above shall be the principle objects of study, we will give how to read them philosophically. In these semantics, 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 we shall refer to as propositions which she ascents to. We shall refer to these interchangeably as *premises*, *assumptions*, *basic beliefs*, *experiences*, or *evidence*. These sets of propositions also represent, in a way, the agent's *frame of mind*; we shall return to developing this perspective in §1.9. For now we will stick to the former readings.

To understand , we think of the agent as producing derivations in a logical calculus on the basis of her evidence. So we alternately read as *the agent believes phi*, *the agent can deduce* φ or *can compose an argument for* φ. We shall make further reference to this reading explicitly again in §.

Like 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. We shall prefer to think of this particular idealization as thinking about what an agent might conclude *eventually*.

By focusing on basic items of evidence as forming the foundation for belief, we consider this approach to be roughly in line with the *evidentialist* view on epistemology, which may be described 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

However, while our sympathies are with this perspective on epistemology, they differ foundationally - while evidentialism develops intuitions using analytical philosophy, our approach shall be founded in formal semantics like the one above.

As alluded to in §1.2, we shall read $\Diamond\varphi$ as *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*.

That is, if a proper foundation can be provided at all. Admittedly, the above formulation of truth immediately runs into a *paradox* - for instance, 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. So let $\mathcal{L}_0(\Phi)$ be the

propositional fragment of $\mathcal{L}_K$, with the following grammar:

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

We shall restrict the truth value to $\mathfrak{M} \subseteq \Phi \times \wp\mathcal{L}_0(\Phi)$. This suffices to make every truth value of this logic determinate.

We may observe that the logic of these semantics is familiar:

Proposition 1.4.2 (Translation)**.** *Assuming that the set of proposition letters Φ is infinite*

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

Here K is basic modal logic.

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

Intuitively, the idea 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 far more elaborate construction, based on the similar principles, shall be presented in §2.3.7.

Armed with this, we can see that these semantics are adequate for modeling agents according to our declared intentions. Recall the following definitions from basic logic and modal model theory[5]:

Definition 1.4.3. *(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*** *$\mathfrak{M}$.*

[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].

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

(*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:

Proposition 1.4.4. *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 B is finite):

$$A \cup B \vdash_K \varphi \iff A \vdash_K \bigwedge B \to \varphi \tag{1.4.1}$$

The above follows from Definition 1.4.3 part (*3*), and is one of the standard results in modal logic.

The next observation is also rather basic:

$$Th(\mathfrak{M}) \vdash_K \varphi \iff \varphi \in Th(\mathfrak{M}) \tag{1.4.2}$$

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:

$$\begin{aligned}
Th(\mathfrak{M}) \cup A \vdash_K \varphi &\iff Th(\mathfrak{M}) \vdash \bigwedge A \to \varphi && \text{by (1.4.1)} \\
&\iff \bigwedge A \to \varphi \in Th(\mathfrak{M}) && \text{by (1.4.2)} \\
&\iff \mathfrak{M}, (b, B) \Vdash \bigwedge A \to \varphi \text{ for all } (b, B) \in Th(\mathfrak{M}) && \text{by Def. 1.4.3 part (1)} \\
&\iff \mathfrak{M}, (a, A) \Vdash \Box\varphi \text{ for any } a \text{ where } (a, A) \in \mathfrak{M} && \text{by Def. 1.4.1}
\end{aligned}$$

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 basis of our future investigations.

1.5 The Human Condition

To supplement to this basic framework, it shall be illustrated how further inspiration can be drawn from a philosophical perspective. This is in stark contrast to the received view in epistemic logic [Len78, pg. 34]:

> The search for the correct analysis of knowledge, while certainly of extreme importance and interest to epistemology, seems not significantly to affect the object of epistemic logic, the question of the validity of certain epistemic-logical principles.

Quite to the contrary, we urge that epistemic logic should not turn its back on philosophy. Philosophy critically provides guidance for the intuitions behind how knowledge should be correctly modeled. It also provides a solid grounding in a proper treatment of knowledge. However, engaging with philosophy is evidently not the thrust of mainstream epistemic logic.

Most mainstream epistemic logic, the object of study is really the nature of information, not human knowledge. It applies equally well to robots, *homo economicus*, or thermometers as suggested in 1.2. Its inspiration is not really in what it is like to be a living person; it is more naturally based in artificial intelligence, information theory, automata theory, algebra, topology, and other abstract disciplines.

In contrast, we shall propose the following principle:

> **The Human Condition**: *The analysis of knowledge should strive for a basis in human experience*

The above principle indeed underpins the justification principle provided in §1.3. This is because we feel that the belief in a proposition can be thought of human only if the agent has a reason associated with it. Otherwise, it seems that in the absence of reason, no account can be given for how the belief came about other than through instrumentation, which is the thermometer view.

By embracing the human condition, we shall now turn to the development of the logical perspective on knowledge extended here from its philosophical origins.

1.6 Soundness

So to give a shallow example of a basic application of a philosophical idea, 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*. This can be done 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:

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

But 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 $\mathfrak{M}, (c, C) \not\Vdash \circlearrowleft$, so this previous definition of knowledge fails. But should it? This is doubtful; 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.

1.7 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's a little suspicious, she might be able to compose an argument with a sound basis that Alex is alive. Taking Descartes as inspiration, we might think of 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 might reasonably express a more adequate notion of knowledge:

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

1.8 Contradictions

There's hidden virtue in the previous analysis. To see what it is, inspiration can be found in the 19th century philosopher Ralph Waldo Emerson, who writes in his essay *Self-Reliance* [Eme08]:

> Why drag about this corpse of your memory, lest you contradict somewhat you have stated in this or that public place? Suppose you should contradict yourself; what then? It seems to be a rule of wisdom never to rely on your memory alone, scarcely even in acts of pure memory, but to bring the past for judgment into the thousand-eyed present, and live ever in a new day. ...
>
> A foolish consistency is the hobgoblin of little minds, adored by little statesmen and philosophers and divines. With consistency a great soul has simply nothing to do. He may as well concern himself with his shadow on the wall. Speak what you think now in hard words and to-morrow speak what to-morrow thinks in hard words again, though it contradict every thing you said to-day. – 'Ah, so you shall be sure to be misunderstood.' – Is it so bad then to be misunderstood?

A healthy lack of consistency is just part of what makes up the day to day life of any living, sane person. Isn't error-prone reasoning a hallmark of human thought? And if a love sick epistemic agent $\exists$ is getting mixed signals from another epistemic agent $\forall$, why can't she draw inconsistent conclusions about $\forall$'s feelings on the one hand, but still have basic knowledge that $734 \times 12 = 8808$ and other such irrelevant facts? There is no compelling reason why not. Under these considerations, the following is compelling:

> **Emerson's Principle**: *One can be inconsistent and still have knowledge*

We may observe how the framework so far developed accommodates this. We may draw further inspiration by a friend and contemporary of Emerson's, the poet Walt Whitman. In *Leaves of Grass* [Whi08], he writes:

> Do I contradict myself?
> Very well then I contradict myself,
> (I am large, I contain multitudes.)

So consider the model $\mathfrak{M}$ in Fig. 3; this is intended to be a toy model of how one might interpret Walt Whitman in the above stanza. This figure 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 $\mathfrak{M}, (b, B) \Vdash A$

Observe that $\mathfrak{M}, (\{p\}, \{p, \neg p\}) \Vdash \Box\bot$; it's obvious that in this state Walt is being inconsistent since he clearly believes contradictory things. Simultaneously, we have that $\mathfrak{M}, (\{p\}, \{p, \neg p\}) \Vdash \Diamond(\circlearrowleft$

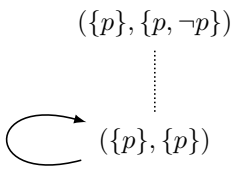


Figure 3: Inconsistent, yet still has knowledge

$\wedge\Box p$); so we figure that Walt has a sound argument that p. Walt might be inconsistent, but it would appear that *at least one* of his arguments makes sense. And this is naturally because Walt contains a multiplicity of inner selves, just like he says, which the the $\boxminus$ modality gives access to.

1.9 Irrationality

Embracing contradiction runs contrary to the received view on epistemic logic. For instance, [HS06] write:

> Epistemic logic does carry epistemological significance but in an inevitably idealized sort of way: One restricts attention to a class of rational agents where rationality is defined by certain postulates. Thus, agents have to satisfy at least some minimal conditions to simply qualify as rational. This is by and large what Lemmon originally suggests [LH59].

Furthermore, it is conventional to think that rational agents do not hold contradictions.[6] For instance, in [KL86], $\neg\Box\bot$ is taken as an axiom (it is A9 in their numbering).

This is similar to the thermometer concept of knowledge we provided in §1.3, since like the thermometer view, it is incompatible with a human perspective. Hence we shall extend the following:

> **Irrationality Principle**: *Since humans are not rational, views on epistemic logic that postulate this should be rejected*

I should mention that while this perspective is not typically embraced in epistemic logic[7], it finds sympathy in other logical traditions, namely in *relevance logic* and *paraconsistent logic*, as already noted [see GG02, chapters 1 & 4].

Apart from inconsistency, we do not really accommodate very much irrationality; frameworks like [Ran82] and and [Lev84] employing *impossible world* semantics are far more accommodating to

[6] It should be remarked that [Pri06] explicitly rejects this perspective on rationality. Priest points out that in times of scientific revolution, rational people naturally hold contradictory views. He suggests that a paraconsistent logic framework could account for a rational agent holding contradictory beliefs. While our hearts are indeed sympathetic to Priest's perspective, we are nonetheless confident that this does not represent the received view which we are attacking.

[7] Noted exceptions to this are [Ran82] and [Lev84].

irrationality than the semantics we are investigating. However, these frameworks do not provide compelling explanation for the mechanism of irrationality, contrary to the perspective presented here. Regardless, allowing for an agents beliefs to be generated from inconsistent premises is already orthogonal to the assumption that agents are rational.

1.10 Quine

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. Thus the choice to visually represent them as *Hasse diagrams*, as we have done in Fig. 3, 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.

A similar perspective has been presented before, however in a different setting, in *Two Dogmas of Empiricism*:

> Certain statements, though about physical objects and not sense experience, seem peculiarly germane to sense experience – and in a selective way: some statements to some experiences, others to others. Such statements, especially germane to particular experiences, I picture as near the periphery. But in this relation of "germaneness" I envisage nothing more than a loose association reflecting the relative likelihood, in practice, of our choosing one statement rather than another for revision in the event of recalcitrant experience. For example, we can imagine recalcitrant experiences to which we would surely be inclined to accommodate our system by re-evaluating just the statement that there are brick houses on Elm Street, together with related statements on the same topic. We can imagine other recalcitrant experiences to which we would be inclined to accommodate our system by re-evaluating just the statement that there are no centaurs, along with kindred statements. A recalcitrant experience can, I have already urged, be accommodated by any of various alternative re-evaluations in various alternative quarters of the total system; but, in the cases which we are now imagining, our natural tendency to disturb the total system as little as possible would lead us to focus our revisions upon these specific statements concerning brick houses or centaurs. These statements are felt, therefore, to have a sharper empirical reference than highly theoretical statements of physics or logic or ontology. *The latter statements may be thought of as relatively centrally located within the total network, meaning merely that little preferential connection with any particular sense data obtrudes itself.* [Qui51]

The emphasis on the last sentence is our addition. The above paragraph importantly anticipates ideas in belief revision theory (such as in [AGM85] and subsequent studies), as well as recent trends in probabilistic dynamic epistemic logic [such as in vB03, vBGK09, BS08, Koo03, etc.]. However, in the framework has been developing so far, what Quine refers to as the "periphery" of his web of belief corresponds to a higher node in a belief poset, while what Quine referes to as the "center" reflects something like a lower node. This is visually depicted in Fig. 4. Beliefs that are members of lower nodes, and the ideas that follow from them, can be thought of as belonging to the agent's world-view.

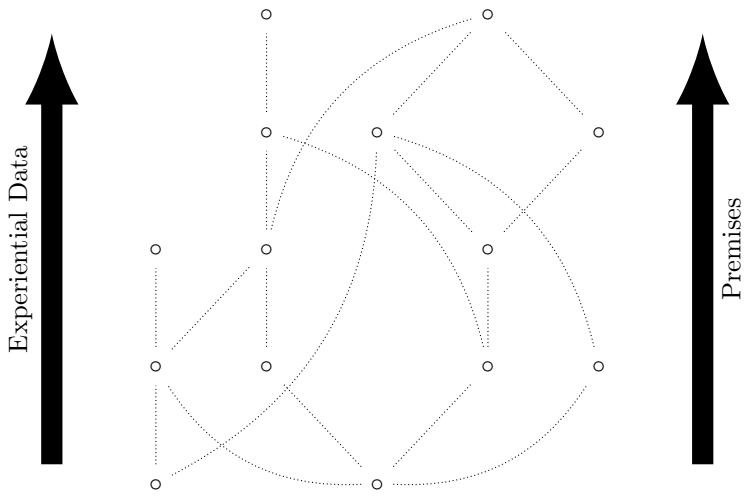


Figure 4: A network of beliefs

The above observation informs a corresponding perspective on epistemology. If an agent's world view largely rested legends about the Norse gods, we should be reluctant to say she knows various facts about nature, such as why lightning strikes. This is because all of her explanations would inevitably be based upon myths in one way or another, which would all occupy lower nodes in her belief network. This dictates that *sanity* plays a role in how much knowledge an agent can have - it is permissible to grant that an inconsistent agent has knowledge provided that the inconsistency follows only shallowly from her experiential data, and it is something she would readily give up. However, if a contradiction is intrinsic to the agent's psychology, and thus follows from a lower node in her belief poset, this suggests she does not really have knowledge. So while we should believe that irrational agents can possess knowledge, as we have argued in §1.9, we should rightfully not contend that they *always* posses knowledge. Moreover, the sort of irrationality that we are considering here need not be superficial - both mundane as well as deeply demented characters can be modeled.

Note that the above essentially presents a subjective interpretation of Quine's web of belief, which might be contentious. On the other hand, we should feel both the quote from Quine and the quote from Whitman in §1.8 suggest the following principle without too much controversy:

> **Quine/Whitman Principle**: *Epistemic agents are compound entities, which invite compositional analysis.*

The above presents a final philosophical principle that shall be extended extend. Apart from this, from the previous discussion may extract an additional thing: Figure 4 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 suggested in §1.7. This suggests the introduction of a new operator $\boxplus$. The semantics for $\boxplus$ are given as follows:

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

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

This concludes the presentation of novelties we propose for the practice of modelling knowledge.

1.11 Closing Remarks

The various principles extended in the previous sections are not independent - some of them are more basic than others. Their relationship is summarized in Fig. 5 - here the lower a principle is depicted, the more basic. Dotted lines indicate the philosophical justification for the higher principle supervenes on the justification of the lower principle. In addition, in further development

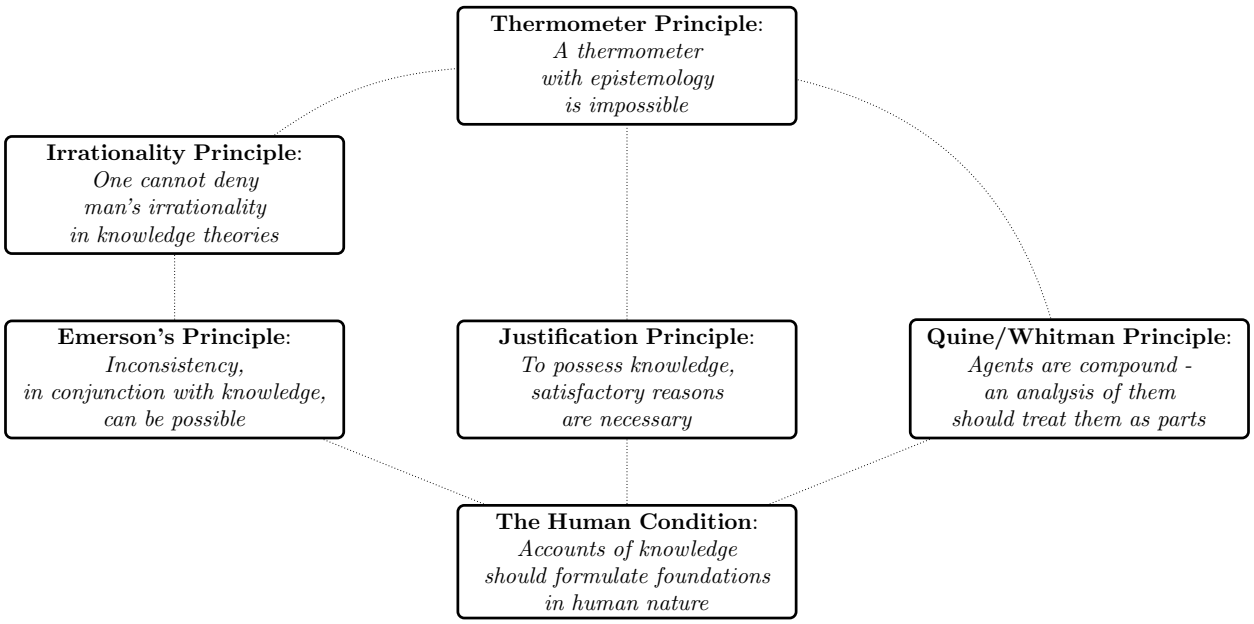


Figure 5: A visualization of the relationship of the principles presented here

of the framework sketched in §1.3, we shall want the following criteria, based on the ideas given in relevant sections:

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

- To avoid paradoxes, correct foundations must be provided. Ideally, we would like our semantics to correspond to a provably terminating computation, granting certain non-deterministic operations such as a *choice operator* ε, as described in [Hil22].

For a set of beliefs A:

§10 It should be expressible whether everything in A is sound

§1.7 Certain subsets $B \subseteq A$ should be accessible

§1.10 Certain extensions $B \supseteq A$ could also be accessed

In line with evidentialist epistemology, as mentioned in §1.4, we have decided to call the logic presented here *Evidentialist Logic*, or EviL for brevity.

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.

§2.3 In this final section, completeness for EviL is investigated.

2.1 Elementary EviL

2.1.1 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 §2.2.2, the semantic framework offered here is extended to incorporate multiple agents. In Appendix B, yet another framework is offered employing gamelike semantics, which avoids the grammar restriction suggested in §1.4.

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 §1.4, in favor of demonstrably well-defined semantics. This shall be achieved in two steps.

Definition 2.1.1. *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 2.1.2. *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 2.1.3. *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 2.1.4. *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 2.1.5. *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 2.1.6 (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 2.1.7.

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

Theorem 2.1.8 (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. The proof proceeds the exactly as the proof of Proposition 1.4.4 from §1.4. QED

I shall present $\vdash_{\text{EvIL}}$, the logical consequence turnstile for EvIL, in §2.3.1.

I chose 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 2.1.8 establishes one of the central desiderata outlined in §1.11 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 §1.4. 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.

This sort of self referential circularity is a celebrated theme in mathematics. It is similar, in a way, to the old alchemical conception of mathematics, exemplified by the following quote due to Sir Thomas Browne:

> All things began in order, so shall they end, and so shall they begin again; according to the ordainer of order and mystical Mathematicks of the City of Heaven.[Bro36, chapter 5]

Another, related notion of self reference was championed by Douglas Hofstadter in his book *Gödel, Escher, Bach: An Eternal Golden Braid*, in what he calls "Strange Loops":

> The flexibility of intelligence comes from the enormous number of different rules, and levels of rules. The reason that so many rules on so many different levels must exist

> is that in life, a creature is faced with millions of situations of completely different types. In some situations, there are stereotyped responses which require "just plain" rules. Some situations are mixtures of stereotyped situations-thus they require rules for deciding which of the "just plain" rules to apply. Some situations cannot be classified-thus there must exist rules for inventing new rules ... and on and on. Without doubt, Strange Loops involving rules that change themselves, directly or indirectly, are at the core of intelligence. Sometimes the complexity of our minds seems so overwhelming that one feels that there can be no solution to the problem of understanding intelligence-that it is wrong to think that rules of any sort govern a creature's behavior, even if one takes "rule" in the multilevel sense described above. [Hof79, pg. 24]

In the setting of EvIL, the strange loop is as follows: the rules of the logic affect, indirectly, the truths of the semantics, which in turn affect the rules of the logic. It is perhaps dangerous to engage in this sort of modeling, since it invites paradox. However, the central goal of epistemic logic is to provide a logic which models agents with knowledge. This is because we are ourselves agents with knowledge; epistemic logic's ultimate purpose is to say ourselves by studying the subject. EvIL was designed with to make true the Theorem Theorem, precisely with this final reason in mind.

It cannot be stressed enough, Theorem 2.1.8 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.

2.1.2 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 2.1.8, 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 §1.7 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

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

then it seems natural that she might imagine that

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

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

$$\boxminus(\Box\neg g \to \Diamond p). \tag{2.1.1}$$

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:

$$\neg\boxminus(\Box\neg f \to \Diamond r) \tag{2.1.2}$$

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 (2.1.2), which is expressed as follows:

$$\Diamond\!\!\!-(\Diamond\neg f \wedge \Box b \wedge \neg\Diamond r) \tag{2.1.3}$$

Further, the intuition for how to read $\oplus\varphi$ 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:

$$\Diamond s \wedge \boxplus(\Box l \to \Box\neg s) \tag{2.1.4}$$

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 §1.10, we would extend a position asserting that remembering factive data is the same as embracing as much of one's evidence as possible.

In terms of the semantics outlined, $\boxminus$ corresponds to a subsetset relation while $\boxplus$ corresponds to a superset relation. Because of this, we shall sometimes read $\boxminus\varphi$ closer to the formal semantics, as saying something like "for all subsets of the agent's basic beliefs or premises, φ holds" and dually for $\boxplus\varphi$. This is admittedly even less natural than the reading of remembering as the opposite of casting into doubt. So be it; we shall have to be comfortable with EviL agents being at best twisted cartoon versions of actual people. Is this so unnatural? Logic, in general, only affords at best cartoon versions of whatever we are trying to model with it. Consider, for instance, Peano arithmetic, which is in general not powerful enough to prove a basic number theoretic fact like Goodstein's theorem [KP82]. Or consider the stronger system of second order arithmetic, which is in turn not strong enough to prove semantics stipulated in §2.1.1, EviL agents apparently have sets for brains, which makes an EviL agent a strange effigy for a person indeed – with the possible exception of set theorists, whose brains are typically constructed entirely of sets or urelements.

Furthermore, it is under the set theoretical reading that $\circlearrowleft$ makes the most sense. We should read it as asserting something like "the basis for the EviL agent's beliefs is sound" or "the EviL agent's arguments only use true premises." It further means that the actual state of affairs is compatible with what the agent believes - reality has not been ruled out by something that the agent is taking as evidence. Moreover, sound premises intuitively exhibit the following property - any subset of them is also sound, since soundness isn't a phenomenon that is subject to synchronicity or other failures of compositionality. A set of premises is sound if and only if all of its subsets are also sound.

2.1.3 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

$$\Vdash\circlearrowleft \leftrightarrow \boxminus \circlearrowleft \tag{2.1.5}$$

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:

$$\nVdash\circlearrowleft\rightarrow \boxplus \circlearrowleft \tag{2.1.6}$$

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 (2.1.5), 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 2.1.8). This is expressed as

$$\Vdash\circlearrowleft\rightarrow \Box\varphi \rightarrow \varphi \tag{2.1.7}$$

The formula (2.1.5) 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:

$$\Vdash \Diamond\varphi \rightarrow \boxminus\Diamond\varphi \tag{2.1.8}$$

One may read (2.1.8) 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

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

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 2.1.8, 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 §2.2.1.

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:

$$\Vdash \diamondminus^{+}\varphi \rightarrow \diamondminus\varphi \tag{2.1.10}$$

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:

$$\Vdash \varphi \rightarrow \diamondminus\varphi \tag{2.1.11}$$

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:

$$\Vdash \Box\varphi \rightarrow \Box\boxminus\varphi \tag{2.1.12}$$

$$\Vdash \Box\varphi \rightarrow \Box\boxplus\varphi \tag{2.1.13}$$

We can observe that this generalizes to multiple agents, as specified in §2.2.2.

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:

$$\Vdash \varphi \rightarrow \boxminus(\circlearrowleft\rightarrow \Diamond\varphi) \tag{2.1.14}$$

$$\Vdash \varphi \rightarrow \boxplus(\circlearrowleft\rightarrow \Diamond\varphi) \tag{2.1.15}$$

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:

$$\Vdash p \rightarrow \boxminus p \tag{2.1.16}$$

$$\Vdash p \rightarrow \boxplus p \tag{2.1.17}$$

$$\Vdash \neg p \rightarrow \boxminus\neg p \tag{2.1.18}$$

$$\Vdash \neg p \rightarrow \boxplus\neg p \tag{2.1.19}$$

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 (2.1.5) 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 Basic EviL

2.2.1 Elimination

In section §2.1.3, 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 summerizes the structural validities will be studied in the subsequent discussion:

Lemma 2.2.1. *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 2.2.1, first observe that EvIL makes true the usual substitution rule:

Lemma 2.2.2. *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 2.2.3. *Define the following fragments:*[8]

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

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

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

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

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

Definition 2.2.4. *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:*

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

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

Theorem 2.2.5 (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* $\Vdash \neg(\psi^A) \leftrightarrow \psi$*.*

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 2.2.1:

Definition 2.2.6. *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 2.2.7 (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 2.2.1.

- 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 2.2.1, 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 2.2.1, namely $\Vdash \boxminus\Diamond\varphi \leftrightarrow \Diamond\varphi$ and $\Vdash \oplus\Diamond\varphi \leftrightarrow \Diamond\varphi$

- Cases $\boxminus, \oplus$: The final step follows from one more application of Lemma 2.2.1, namely by employing the following four validities

$$\begin{array}{ll} \Vdash \boxminus\oplus\varphi \leftrightarrow \oplus\varphi & \Vdash \oplus\oplus\varphi \leftrightarrow \oplus\varphi \\ \Vdash \boxminus\boxminus\varphi \leftrightarrow \boxminus\varphi & \Vdash \oplus\boxminus\varphi \leftrightarrow \boxminus\varphi \end{array}$$

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 $\oplus$. 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:

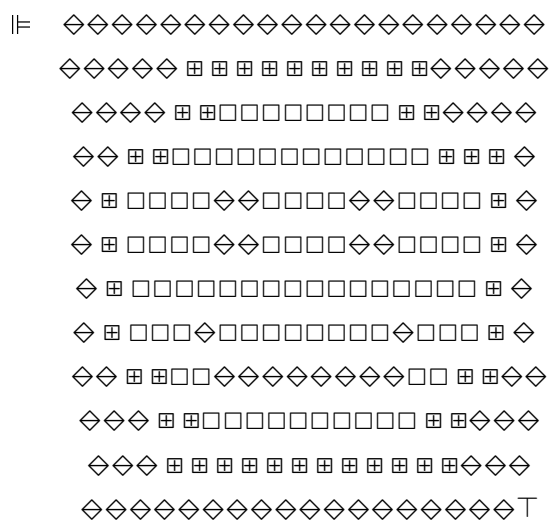
$$\begin{aligned} \Vdash \psi &\leftrightarrow \neg(\psi^A) \\ &\leftrightarrow \neg((\psi^A)^*) \\ &\leftrightarrow \neg((\psi^*)^A) \\ &\leftrightarrow \neg(\neg(((\psi^*)^A)^B)) \\ &\leftrightarrow \neg\neg\psi^* \\ &\leftrightarrow \psi^* \end{aligned}$$

QED

Example 2.2.8. *The following validities of* EvidL *are consequences of Theorem 2.2.7:*

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

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



One way to read Theorem 2.2.7 is that $\boxminus$ and $\diamondplus$ are empty modalities on $\mathcal{L}_A(\Phi)$, and dually for $\mathcal{L}_B(\Phi)$ with $\diamondminus$ 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 $\diamondplus$ and $\diamondminus$ 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 2.2.7 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.

2.2.2 Multiple Agents

In this section we extend the semantics for EvIL from a single agent, as presented in §2.1.1, 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 2.2.9. *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 §2.1.1, 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 $(\Vdash)$: [9]

Definition 2.2.10.

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

Just as in §2.1.1, Lemma 2.1.6 and Theorem 2.1.8 can be seen to obtain for the new generalized semantics. Furthermore, all of the validities mentioned in §2.1.3 and §2.2.1 hold, along with Theorem 2.2.7, 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:

$$\Vdash \Box_X \varphi \to \Box_X \boxminus_Y \varphi$$
$$\Vdash \Box_X \varphi \to \Box_X \boxplus_Y \varphi$$

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 (2.1.12) and (2.1.13) in §2.1.3. 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 §2.3 can now be given in higher generality.

2.2.3 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

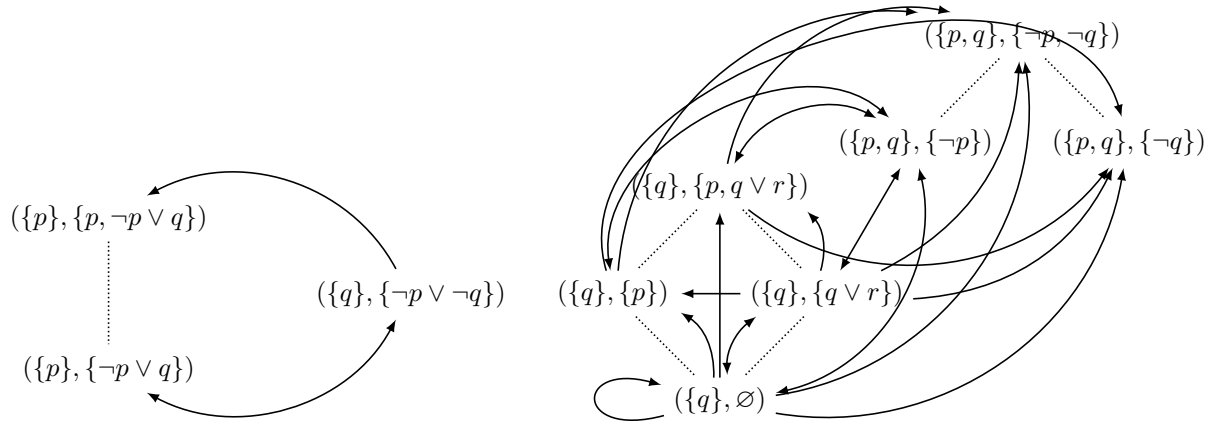
[9] Where $X \in \mathcal{A}$, we shall use A_X to denote $A(X)$ provided that $A : \mathcal{A} \to \wp\mathcal{L}_0(\Phi)$

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 §1.8. This involved thinking of the EvIL models as *posets* with arrows, as we first presented in Fig. 3 in §1.8. We also saw additional examples of this visualization technique below in Figs. 6(a) and 6(b). Recall that 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.



(a) A fairly simple example

(b) A more complex example

Figure 6: EvIL model visualizations

Following the modified semantics provided in §2.2.2, the developments this section will assume multiple agents.

Definition 2.2.11. *Let Φ be a set of letters and let $\mathcal{A}$ be a set of agents. A **Kripke structure** is 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*[10]*:*

- *W is a set of worlds*
- *$R : \mathcal{A} \to \wp(W \times W)$, $\sqsubseteq$: $\mathcal{A} \to \wp(W \times W)$, and $\sqsupseteq$: $\mathcal{A} \to \wp(W \times W)$ are $\mathcal{A}$-indexed sets of relations*[11]

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

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

- $V : \Phi \to \wp(W)$ *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* $W \subseteq I$*.*

Definition 2.2.12. *By a minor abuse of notation, we shall write* $w \sqsubseteq v$ *to mean that* $w \sqsubseteq_X v$ *for all agents* X*.*

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

Definition 2.2.13. *Let* $\mathbb{M}$ *be a Kripke structure:*

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

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

Definition 2.2.14 ($\mathfrak{U}^{\mathfrak{M}}$ Translation)**.** *Let* $\mathfrak{M}$ *be an* EvиL *model. Define* $\mathfrak{U}^{\mathfrak{M}} := \langle \mathfrak{M}, R^{\mathfrak{M}}, \sqsubseteq^{\mathfrak{M}}, \sqsupseteq^{\mathfrak{M}}, V^{\mathfrak{M}}, P^{\mathfrak{M}} \rangle$*, where*

- $(a, A) R^{\mathfrak{M}}_X (b, B) \iff \forall \psi \in A_X . b \models \psi$
- $(a, A) \sqsubseteq^{\mathfrak{M}}_X (b, B) \iff a = b$ *and* $A_X \subseteq B_X$
- $(a, A) \sqsupseteq^{\mathfrak{M}}_X (b, B) \iff a = b$ *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 2.2.15. *For all* $\mathfrak{M}$ *and all* $(a, A) \in \mathfrak{M}$*,*

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

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

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

Proposition 2.2.16. *For any* EvиL *model* $\mathfrak{M}$*,* $\mathfrak{U}^{\mathfrak{M}}$ *has the following properties, for all agents* $\{X, Y\} \subseteq \mathcal{A}$*: :*

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

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

(III) $\sqsubseteq^{\mathfrak{M}}$ *is a partial order*

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

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

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

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

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

The situation in (VI) *can be visualized in a commutative diagram depicted in 7(a), while* (VII) *can be split into Figs. 7(b) and 7(c).*

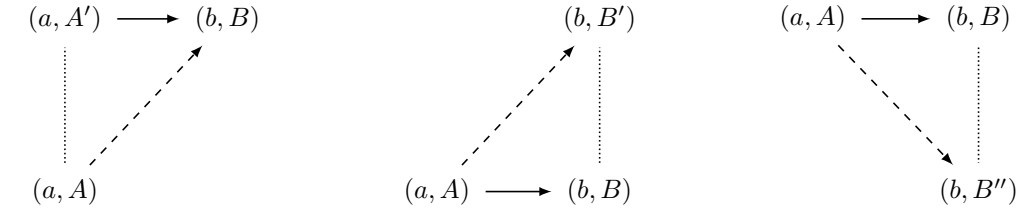


Figure 7: Visualizations of the relationships in Proposition 2.2.16

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

(VI) 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.

(VII) 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 $(a, A)R^{\mathfrak{M}}_X(c, C)$, which was what was to be shown.

QED

Definition 2.2.17. *A Kripke structure is called* EvIL *if it makes true the above properties* (I) *through* $(VIII)$*.*

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

- If $xR^{\mathfrak{M}}_X y$, then at world x the agent X can imagine y is true, since y is compatible with what the agent believes
- If $x \sqsubseteq^{\mathfrak{M}}_X y$, 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 (VI) 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. 2.2.16 presents itself as a sort of representation lemma, the relationship between EvIL semantics and Kripke semantics is not reciprocal. Proposition 2.2.19 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 2.2.18. *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 2.1.2 in §2.1.1 are the usual semantics for classical propositional logic. Remembering this, the above is an elementary result in basic logic. QED

Proposition 2.2.19 (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. 8. We shall show that $\mathbb{M}$ is not represented by any EvIL model.

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

$$\mathbb{M}, w \Vdash \Diamond\top \tag{2.2.1}$$
$$\mathbb{M}, w \Vdash \Box\neg p \text{ for all } p \in \Phi \tag{2.2.2}$$
$$\mathbb{M}, w \Vdash \neg p \text{ for all } p \in \Phi \tag{2.2.3}$$
$$\mathbb{M}, w \Vdash \neg\Diamond\Diamond\top \tag{2.2.4}$$

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

(*2*) From (2.2.2), we know that for the b mentioned above it must be that $b = \varnothing$. This is a direct consequence of Lemma 2.1.6, the Truthiness Lemma.

(*3*) From (2.2.3), evidently $a = \varnothing$

(*4*) From (2.2.4), it must be that $a \not\models A$. Otherwise by the semantics of EvIL as defined in §2.1.1 we would have $\mathfrak{M}, (a, A) \Vdash \Diamond\Diamond\top$

Since $a = b = \varnothing$ and $b \models A$ then by Lemma 2.2.18 it must be that $a \models A$. But this clearly is absurd! $\lightning$ QED

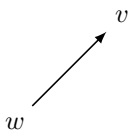


Figure 8: A Kripke structure $\mathbb{M}$ with no EvIL representation

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 *most* of the benefits of basic modal logic. Indeed, we shall see in §2.3.3 that EvIL is strongly complete for EvIL Kripke models.

Perhaps the most important formal feature that EvIL semantics lacks in comparison to abstract Kripke semantics is that, as a consequence of the observations made in Proposition 2.2.19, EvIL is not compact.

Theorem 2.2.20 (Failure of Compactness)**.** *If the set of proposition letters Φ is infinite, then* EvIL *is not compact for* EvIL *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 EviL semantics, by the arguments presented in the proof of Proposition 2.2.19.

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. 9. It is straightforward to verify that $\mathfrak{M}, (\{\rho\}, \{\neg\rho\}) \Vdash \tau[\Psi]$, so $\mathfrak{M}$ is a suitable witness. QED

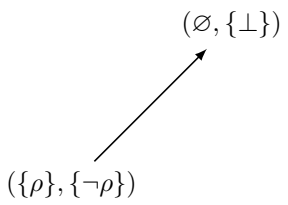


Figure 9: A model $\mathfrak{M}$ where $\mathfrak{M}, (\{\rho\}, \{\neg\rho\}) \Vdash \tau[\Psi]$ for $\Psi \subseteq_\omega \Phi$ and $\rho \notin \Psi$

A consequence of the failure of compactness, while strong completeness can be obtained for EviL using Kripke semantics, to achieve completeness for EviL semantics a finitary proof must be carried out. Recall that this was exactly the strategy used in our original sketch of EviL that we gave in Proposition 1.4.2 in §1.4.

The next section is devoted to studying completeness for EviL.

2.3 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:

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

§2.3.2 Next, we give a definition of the class of *partly* EviL Kripke models.

We then reveal that EviL is sound and strongly complete for the class of partly EviL models. 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.

§2.3.3 In this section we recall the definition of an EviL *Kripke Model*, as we gave in Definition 2.2.17 from §2.2.3, and show that every *partly* EviL Kripke model may be "completed" by constructing a bisimilar EviL model.

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

§2.3.4 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 EvиL.

§2.3.5 In this section we show that EvиL has a small model property for *partly* EvиL Kripke models. This is accomplished by constructing a Kripke model 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].

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

§2.3.7 Following the proof of Proposition 1.4.2 in §1.4, we shall show that every finite EvиL Kripke structure is *(almost)-homomorphic* to another, EvиL model 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 EvиL is weakly complete for EvиL models.

§2.3.8 In this section, we once again pause to take stock of the results we have established in previous sections. We discuss how our results give rise to complexity observations for EvиL, and discuss the relationship between the abstract completeness we have previously established and our concrete completeness.

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

§2.3.10 We provide an extension to EvиL 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.

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

2.3.1 Axioms of EviL

In this section, we shall present the axiomatics for multi-agent EvиL. The axioms of EvиL are presented in Table 1, which comprises a Hilbert systems for $\vdash_{\text{EvиL}}$[12]. In addition to giving each axiom, we provide a philosophical reading of what each axiom says. As remarked in §2.1.3, EvиL is not *normal*, that is it is not closed under variable substitution.

[12]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 1: A Hilbert style axiom system for EvIL

Needless to say, these axioms shall be the focus of study for all of our future investigations in §2.3.

2.3.2 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 2.2.17 from §2.2.3, only property $(VIII)$ has been weakened to $(VIII)'$ and $(IX)'$.

Definition 2.3.1 (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*

$(III)'$ $\sqsubseteq$ *is a partial order*

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

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

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

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

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

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

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 1 in §2.3.1. We can see this by observing the following theorem:

Definition 2.3.2. *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 $\mathbb{M} \Vdash \varphi$.*

Theorem 2.3.3 (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][13]. Let $\mathcal{E}$ be the set of maximally consistent sets of formulae for EvиL. 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 EvиL, by establishing that it satisfies the properties given in Definition 2.3.1.

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

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

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

(*V*)′ 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 $(IV)'$).

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

(*III*)′ We have deferred the proof that $\sqsubseteq$ is a partial order, since it depends on the above results. We know that it is reflexive and transitive by $(I)'$ and $(II)'$. All that is left is to show that it is anti-symmetric. Assume that $w \sqsubseteq v$ and $v \sqsubseteq w$. By the Truth Lemma it suffices to show that $\mathscr{E}, w \Vdash \varphi$ if and only if $\mathscr{E}, v \Vdash \varphi$, since then $\varphi \in w$ if and only if $\varphi \in v$. We shall show this by induction on φ.

[13] 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 EvиL logic under consideration.

$p \in \Phi$ – We have this step from the assumption that $w \sqsubseteq v$ and $(V)'$.

$\bot$ – This case is trivially true.

$\boxplus_X \varphi$ **and** $\boxminus_X \varphi$ – These steps follows from $(II)'$, $(IV)'$ and the assumption.

$\Box_X \varphi$ – This case follows from $(VI)'$ and the assumption.

$\circlearrowleft_X$ – This step follows from $(IX)'$ and the assumption.

$\varphi \to \psi$ – The final step follows trivially from the inductive hypothesis.

(VII)′ Given $(IV)'$, 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).

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

QED

2.3.3 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 $(VIII)$, 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 2.3.4. *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 2.3.5 (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 2.3.6.

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

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

Definition 2.3.7 (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.

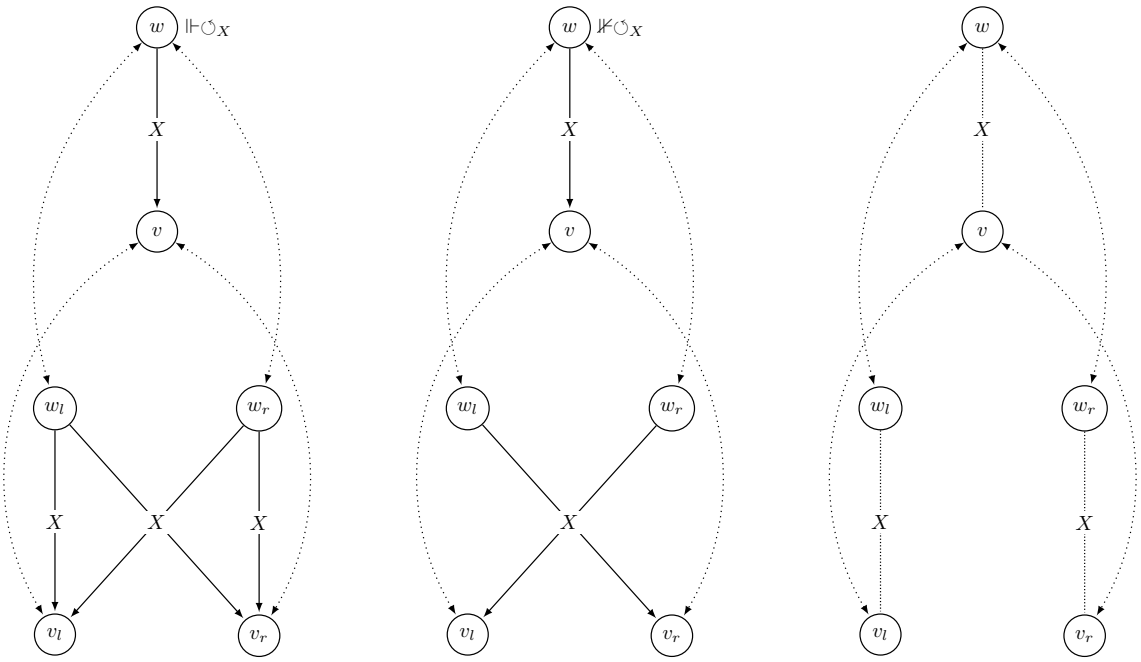
(*2*) If $\mathbb{M}, w \Vdash \circlearrowleft_X$, then the copies w_l and w_r of w can see both v_l and v_r provided that $wR^{\mathbb{M}}v$ to begin with.

(*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. 10(a) and 10(b). The third rule is depicted in 10(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 $\blacktriangleleft\cdots\blacktriangleright$, 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 $\overset{X}{\longrightarrow}$, then wR_Xv, taking care to note which model we are reasoning in.



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

Figure 10: Visualizations of ♋'s operation

- If one point w is connected and *above* to another point v by a densely line with no arrow and a label X $\overset{\cdots X \cdots}{}$, then $w \sqsupseteq_X v$.

We summarize the mechanics of ♋ in the following proposition:

Proposition 2.3.8. *Let $\{w, v\} \subseteq W^{♋}$, 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

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

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

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

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

Lemma 2.3.9. *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 $Z : \mathbb{M} \leftrightarrows ♋^{\mathbb{M}}$:*

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

Proof. It follows directly from the definition of $\circlearrowleft$ 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 2.3.10 (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 (V) and (VII) follow by construction, and the fact that since $\mathbb{M}$ is partly EvIL by hypothesis it makes true $(I)'$ through $(V)'$ and $(VII)'$. All that is left to show is (VI) and $(VIII)$.

(*VI*) We must show

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

So assume that $w \sqsubseteq_X^{\circlearrowleft} u R_X^{\circlearrowleft} v$. 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 $w R_X^{\circlearrowleft} v$; 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 $w R_X^{\circlearrowleft} v$.

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 $(IX)'$ 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 $w R_X^{\circlearrowleft} v$ as desired.

(*VIII*) 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 2.3.3 from partly EvIL models to the fully EvIL models originally defined in Definition 2.2.17 from §2.2.3:

Definition 2.3.11. *As in Definition 2.3.35, 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* $\mathbb{M} \Vdash \varphi$*.*

Theorem 2.3.12 (EvIL Strong 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 2.3.3.

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 2.3.3 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 2.3.9 that $\mathscr{E} \leftrightarrows \mathfrak{S}^{\mathscr{E}}$, hence by Theorem 2.3.5, *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 2.3.10 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.

2.3.4 Taking Stock I

In the previous sections, we showed that the logic of EvIL we presented in 2.3.1 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 2.1.4 in §2.1.1. These semantics were carefully crafted to make true the mystical Theorem 2.1.8, the *Theorem Theorem*. This equated $\Box\varphi$ with a proof of φ, in the following manner:

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

In the above, we assume that A_X is finite. This above property of EvIL was enforced to accommodate the *Justification Principle* from §1.3, 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 2.2.19. 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:

$$\Gamma \Vdash_{\text{EvIL}} \varphi \iff \Gamma \Vdash \varphi \tag{2.3.1}$$

... 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 (2.3.1) 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 §2.3.5 through §2.3.7 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 2.2.20, we shall establish a *small model property* for partly EvIL and EvIL Kripke structures in §2.3.5. By modifying the translation system for finite Kripke structures to EvIL modifying we originally gave in the proof of Proposition 1.4.2 from §1.4, we shall show how to translate finite EvIL Kripke structures into EvIL models in §2.3.7. We shall find need to make use of the concept of *islands*, which we shall introduce in §2.3.6.

After the above developments, we shall once again take stock of our observations in §2.3.8. We shall prove the equation (2.3.1), and make use of our previous results to establish complexity bounds on the decision procedure for EvIL.

2.3.5 Small Model Construction

In this section we provide definitions and lemmas related to the subformula construction $\circledcirc^{\varphi}$. 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, pgs. 243]. We shall also introduce *pseudo-boxes*, which are defined as follows:

Definition 2.3.13.

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

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 2.3.14.

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

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

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 2.3.15. *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 $\varphi \in \mathcal{L}(\mathcal{A}, \Phi)$

- *Sigma* 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 2.3.16. *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 2.3.17. $\Sigma(\delta(\varphi), \varphi)$ *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 2.3.18 (Atoms)**.** *Let* $At(\Psi)$ *denote the maximally consistent subsets of* Ψ

We next have the Lindenbaum Lemma:

Lemma 2.3.19 (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, pgs. 79]. QED

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

Definition 2.3.20. *Define*

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

Where:

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

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; we are consciously imitating the completeness techniques given in [Boo95, chapter 5, pgs. 78–84] 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 2.3.21 (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 $\vdash \boxslash_X \psi \leftrightarrow \boxminus_X \psi$, we have that $w \nvdash \boxslash_X \psi$.

Now abbreviate:

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

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 2.3.14 and rule (3) and the axioms we have that[14]

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

So let

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

Since $\boxslash_X \boxslash_X \chi = \boxslash_X \chi$ by Lemma 2.3.14, we have $A' = \boxslash_X A$. Moreover, by Lemma 2.3.14, axiom (12), and classical logic we can see that

$$\vdash \sim \chi \to \boxslash_X \sim \boxtimes_X \chi$$

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

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

However, evidently $A' \cup B' \subseteq w$, hence $w \vdash \boxslash_X \psi$, 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 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 2.3.14 that $\boxtimes_X \boxtimes_X \psi = \boxtimes_X \psi$, which once again implies that v is inconsistent. $\lightning$

[14] Here $\boxslash_X S$ is shorthand for $\{\boxslash_X \chi \mid \chi \in S\}$.

$\boxplus_X\varphi$ – 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 2.3.22 ($\textcircled{\scriptsize ☺}^{\varphi}$ is Partly EviL)**.** *$\textcircled{\scriptsize ☺}^{\varphi}$ is a finite partly* EviL *Kripke structure.*

Proof. The fact that $\textcircled{\scriptsize ☺}^{\varphi}$ is finite follows from the fact that $W^{\textcircled{\scriptsize ☺}} \subseteq \Sigma(\delta(\varphi), \varphi)$ and $\Sigma(\delta(\varphi), \varphi)$ is itself finite, as we established in Lemma 2.3.14.

The remainder of the proof is devoted to establishing the partly EviL properties for $\textcircled{\scriptsize ☺}^{\varphi}$.

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

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 2.3.14, 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 2.3.31, and the fact that $\sqsubseteq_X$ is reflexive for EviL models.

(*II*)′ A quick glance at the definition of $\textcircled{\scriptsize ☺}^{\varphi}$ reveals that "$\sqsubseteq_X$ is transitive" is immediate by construction.

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

(*V*)′ 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 2.3.16 we have that $\boxplus_X p = \boxtimes_X p \in \Sigma(\delta(\varphi), \varphi)$. By axiom (6) and the Truth Lemma (Lemma 2.3.21), we have that $\boxtimes_X p \in w$, whence $p \in v$ by construction of $\textcircled{\scriptsize ☺}^{\varphi}$. The other direction is similar, however it uses axiom (5) instead.

(*VI*)′ To prove the assertion

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

We first note that if $X \notin \delta(\varphi)$, then $R_X \circ \sqsubseteq_X = R_X$ by construction.

So assume $X \in \delta(\varphi)$. We shall show only show that

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

Since the other inclusion follows from $(IV)'$, which we have previously established.

So 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 2.3.16 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.

(*IX*)′ 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 $(V)'$ 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.

(*III*)′ Just as in the proof of 2.3.3 from §2.3.2, we have intentionally deferred the proof that $\sqsubseteq$ is a partial order, since it depends on the above results. As before, that it is reflexive and transitive by $(I)'$ and $(II)'$. Moreover, the exact same inductive proof, along with the Truth Lemma, may be used to establish anti-symmetry.

(*VII*)′ We must show:

$$(\sqsubseteq_Y \circ R_X) = R_X = (\sqsupseteq_Y \circ R_X)$$

Given $(I)'$, we know that

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

So we shall 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 $\sqsubseteq_Y \circ R_X = R_X$ as we want. 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.

QED

We may combine the results above with what we have shown previously in §2.3.3 to give the following series of results:

Theorem 2.3.23 (Abstract Finite EvIL Soundness and Weak 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 2.3.22, so from Theorem 2.3.10 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 2.3.24 (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 $|\delta(\varphi)| \in O(EXP(|\varphi|))$.

Since 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, we may again do a worst-case analysis. 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 2.3.25 (EvIL Decidability)**.** EvIL *is decidable and the time complexity of the decision problem for* EvIL *is bounded above by $O(EXP3)$*

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 2.3.24, 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*.

2.3.6 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 2.3.26. *Let* $\mathbb{M}$ *be a partly* EvIL *Kripke structure. Let*

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

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.

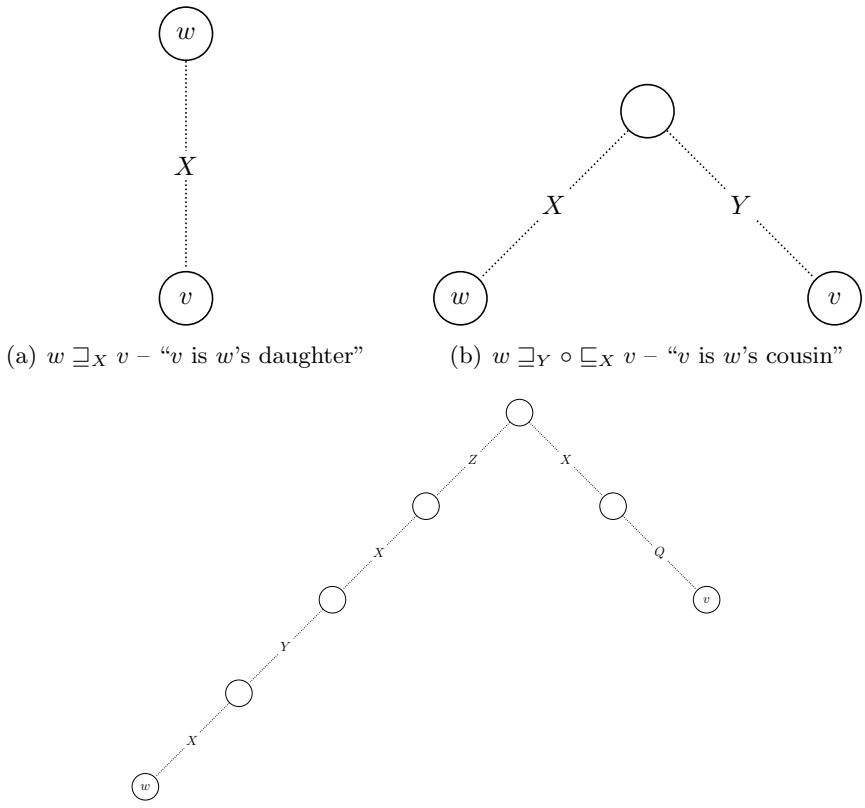
(*i*) One way to understand $\lceil w \rfloor$ is that this is the set of worlds that are graph reachable from w using $\sqsubseteq_X$ and $\sqsupseteq_X$ for any agent X. Since we are considering both $\sqsubseteq_X$ and $\sqsupseteq_X$, then we know that we are thinking about graph reachability on 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 §2.3.7 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 $w \sqsupseteq_Y \circ \subseteq_X v$ means that v is w's cousin. These sorts of relationships are depicted in Figs. 11(a), 11(b), and 11(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 §1.10. This way of thinking about islands makes the most sense in the single agent case. Every island is a poset, which we have been representing with Hasse diagrams in Figs. 3 and 4. Note that as we asserted in §1.10, as one travels down a belief poset, one can imagine more things. EvIL Kripke models are good abstractions on this intuition; indeed, property (*VI*) and axiom 7 reflect exactly this idea Anticipating what we shall reveal in lemma 2.3.27, we have pictured an agent's island in Fig. 12. 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. 13.

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:



(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 11: Family relationships within columns

Lemma 2.3.27 (Island Lemma)**.** *The following hold if* $\mathbb{M}$ *is partly* EVIL*:*

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

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

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

(*4*) wR_Xv *if and only if* $wR_X\lceil v \rfloor$[15]

(*5*) *If* $w \in \lceil v \rfloor$ *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.

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

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

(*5*) Assume that $w \in \lceil\!\lfloor v \rfloor\!\rceil$, and that $w \in V(p)$. Then we may induct on path length, and use $(V)'$ to see that $v \in V(p)$. From (*1*) and (*2*) above, we know that $w \in \lceil\!\lfloor v \rfloor\!\rceil$ implies that $v \in \lceil\!\lfloor w \rfloor\!\rceil$, 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.

2.3.7 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[16]. 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 1.4.2 from §1.4. From this correspondence, we shall obtain a weak completeness theorem for EvIL and its intended semantics.

Recall that in the proof of Proposition 1.4.2 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 $\lceil\!\lfloor w \rfloor\!\rceil$ 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 2.3.28. *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*

$$\Lambda := \{\{w\}, \lceil\!\lfloor w \rfloor\!\rceil \mid w \in W^{\mathbb{M}}\}$$

[16]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.

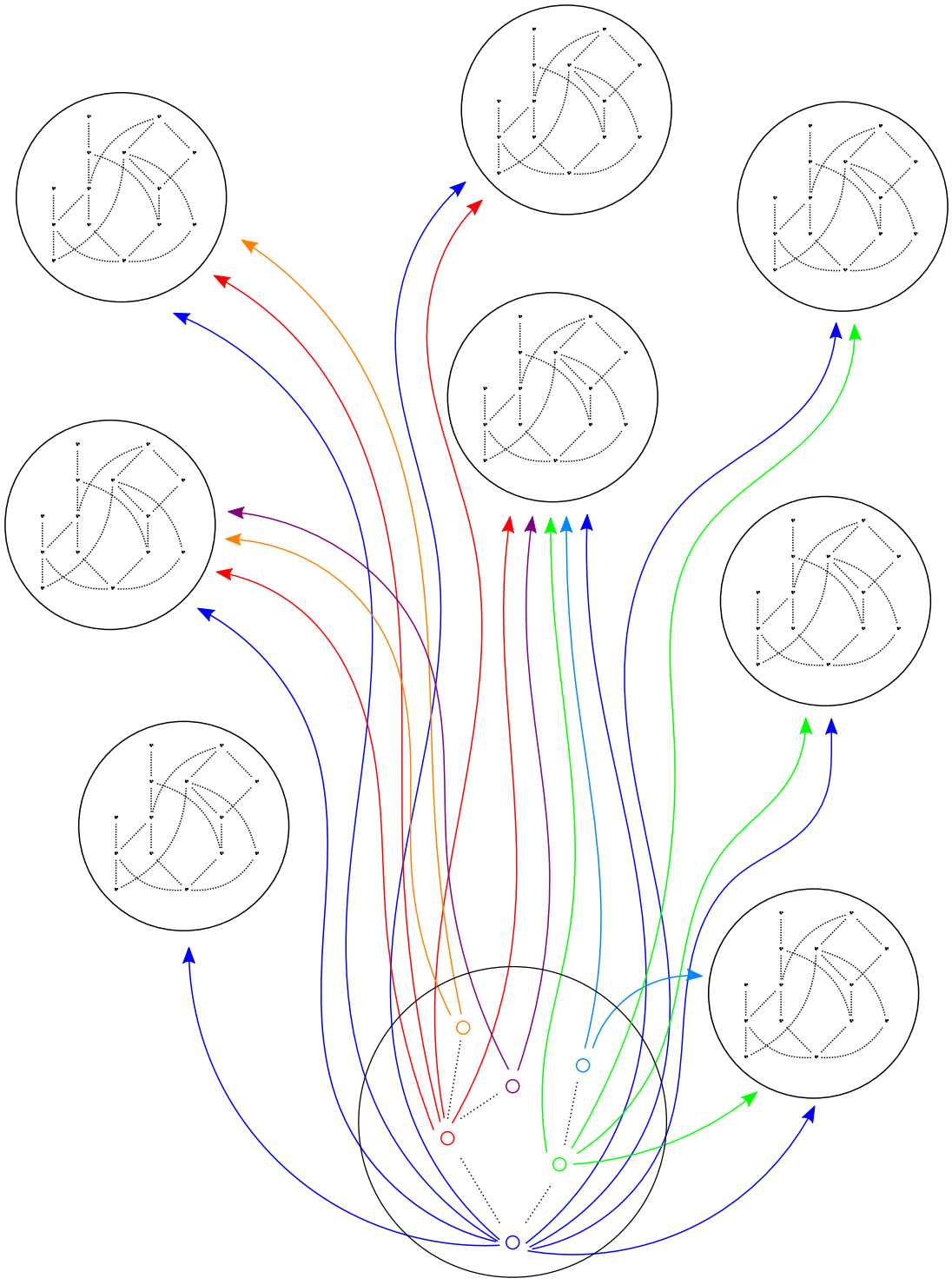


Figure 12: The inner functioning of an island and its relations to other islands

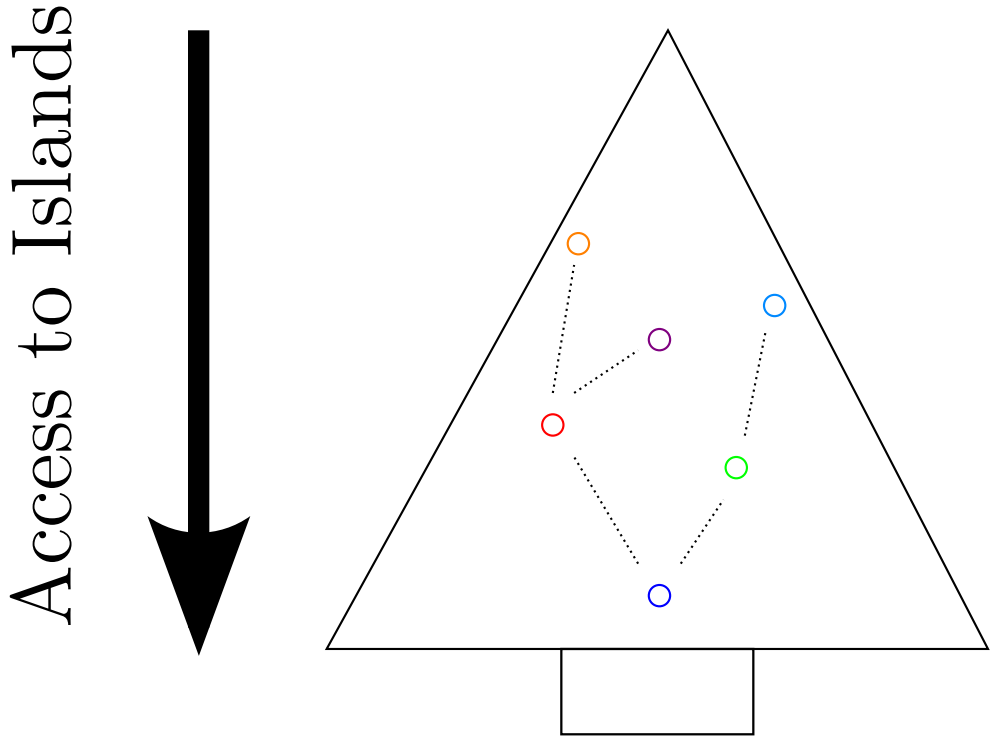


Figure 13: An island is like a *Christmas tree*

That is, Λ is the set of worlds and islands.

- *Let $p : \Lambda \rightarrowtail \Phi \backslash \Psi$ be an injection, assigning names to worlds and surnames to islands*[17].
- *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:

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

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

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

◇ $\vartheta_{2B} : W^{\mathbb{M}} \times \mathcal{A} \to \wp(\mathcal{L}_0(\Phi))$ *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, that constitute their the set of basic beliefs as we originally introduced in §1.4.

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 2.2.14 from §2.2.3. 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:

$$q \in \Psi \Longrightarrow (\mathbb{M}, w \Vdash q \iff \maltese, \vartheta(w) \Vdash q) \tag{2.3.2}$$

$$\mathbb{M}, w \Vdash \circlearrowleft_X \iff \maltese, \vartheta(w) \Vdash \circlearrowleft_X \tag{2.3.3}$$

$$w \sqsubseteq_X^{\mathbb{M}} v \iff \vartheta(w) \sqsubseteq_X^{\maltese} \vartheta(v) \tag{2.3.4}$$

$$w \sqsupseteq_X^{\mathbb{M}} v \iff \vartheta(w) \sqsupseteq_X^{\maltese} \vartheta(v) \tag{2.3.5}$$

$$v \in \ulcorner w \urcorner^{\mathbb{M}} \iff \vartheta(v) \in \ulcorner \vartheta(w) \urcorner^{\maltese} \tag{2.3.6}$$

$$w R_X^{\mathbb{M}} v \iff \vartheta(w) R_X^{\maltese} \vartheta(v) \tag{2.3.7}$$

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 2.3.29.

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 (2.3.4), (2.3.5) and (2.3.6). From definition 2.2.14 from §2.2.3 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 (2.3.6), then we know that we cannot encode any information in $\vartheta_1(w)$ without putting it into $\vartheta_1(v)$ for any $v \in \ulcorner w \urcorner^{\mathbb{M}}$. 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 $\neg p_{\ulcorner v \urcorner}$, 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 (2.3.7). 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 2.3.27(*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 (2.3.4) and (2.3.5), 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 2.3.29. *Provided that $\mathbb{M}$ is* EviL*, our definition of $\maltese$ suffices* (2.3.2) *through* (2.3.7)*.*

Proof.

- (2.3.2)

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

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

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

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

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 $q \in \{p_{\ulcorner w \urcorner}\}$. 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.

- (2.3.3)

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

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 (VII) of EvIL Kripke models it suffices to prove (2.3.7) below.

- (2.3.4)

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

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 (VI) 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.

- (2.3.5)

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

This follows from (2.3.4) and the fact that both $\mathbb{M}$ and ⛤ are EviL, hence $x \sqsubseteq_X y \iff y \sqsupseteq_X x$ for both structures.

- (2.3.6)

$$v \in \ulcorner w \urcorner^{\mathbb{M}} \iff \vartheta(v) \in \ulcorner \vartheta(w) \urcorner^{\text{⛤}}$$

The fact that islands in both structures correspond follows from the correspondences between $\sqsubseteq_X$ and $\sqsupseteq_X$, as we already saw in (2.3.4) and (2.3.5).

- (2.3.7)

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

$\Longrightarrow$ First assume that $wR_X^{\mathbb{M}} v$, we want to show $\vartheta(w) R_X^{\text{⛤}} \vartheta(v)$. 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 $p_{\ulcorner u \urcorner} \in \vartheta_1(u)$ and $\neg p_{\ulcorner u \urcorner} \in \vartheta_{2A}(w)$, which means that $\neg wR_X^{\mathbb{M}} u$. Since $p_{\ulcorner u \urcorner} \notin L(\Phi)$ it must be that $p_{\ulcorner u \urcorner} = p_{\ulcorner v \urcorner}$, hence $\ulcorner u \urcorner = \ulcorner v \urcorner$. Then by the Island Lemma we have $\neg wR_X^{\mathbb{M}} v$ after all. ↯

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

$\Longleftarrow$ Assume $\vartheta(w) R_X^{\text{⛤}} \vartheta(v)$, 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 $\neg p_{\ulcorner v \urcorner} \in \vartheta_{2A}(w)$. However we know that $p_{\ulcorner v \urcorner} \in \vartheta_1(v)$, hence $\vartheta_1 \models p_{\ulcorner v \urcorner}$, which means that $\vartheta_1(v) \not\models (\vartheta(w))_X$, which contradicts our assumption. ↯

QED

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

Lemma 2.3.30 (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*

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

Proof. Using induction, and Lemma 2.3.29, the result follows from the fact that ⛤ 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 2.3.31 (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 2.3.23 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 1.4.2 from §1.4. Since $\varphi \in \mathcal{L}(L(\varphi), \mathcal{A})$, from lemma 2.3.30 we have that $\text{✡}, \vartheta(w) \nVdash \varphi$. Then evidently ✡ is our desired counter model, hence we have the theorem. QED

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

2.3.8 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 §2.3.4, which we expressed in (2.3.1).

Lemma 2.3.32. *For finite* Γ*:*

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

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 common place 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 quote "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 concrete Noetherian ring. Indeed, it is by understanding the integers that the theory of Noetherian rings proceeds. 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 abstracting Birkoff's theorem [DP02, chapters 11 and 5, respectively], and the Yoneda Lemma abstracting 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 §2.3.2 until we were convinced that EviL Kripke structures generalized our concrete structures. The process by which EviL was developed involved finding the results in §2.3.5 first, and letting those properties we deemed necessary to coerce a Kripke structure into a Evil model 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.

Since EviL Kripke structures really are abstract idealizations of concrete EviL models, as we have illustrated, we are granted a particularly comfortable situation. 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.

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 (2.3.1) to transfer these results to EviL models.

2.3.9 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 2.3.33. *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)	$\vdash \boxminus_X \varphi \to \varphi$	(4)	$\vdash \boxplus_X \varphi \to \varphi$
(5)	$\vdash \boxminus_X \varphi \to \boxminus_X \boxminus_X \varphi$	(5)	$\vdash \boxplus_X \varphi \to \boxplus_X \boxplus_X \varphi$
(6)	$\vdash p \to \boxminus_X p$	(6)	$\vdash p \to \boxplus_X p$
(7)	$\vdash \neg p \to \boxminus_X \neg p$	(7)	$\vdash \neg p \to \boxplus_X \neg p$
(8)	$\vdash \Diamond_X \varphi \to \boxminus_X \Diamond_X \varphi$	(8)	$\vdash \Box_X \varphi \to \boxplus_X \Box_X \varphi$
(9)	$\vdash \Box_X \varphi \to \Box_X \boxminus_Y \varphi$	(9)	$\vdash \Box_X \varphi \to \Box_X \boxplus_Y \varphi$
(10)	$\vdash \varphi \to \boxminus_X (\circlearrowleft_X \to \Diamond_X \varphi)$	(10)	$\vdash \varphi \to \boxplus_X (\circlearrowleft_X \to \Diamond_X \varphi)$
(11)	$\vdash \circlearrowleft_X \to \boxminus_X \circlearrowleft_X$	(11)	$\vdash \neg \circlearrowleft_X \to \boxplus_X \neg \circlearrowleft_X$
(12)	$\vdash \Box_X (\varphi \to \psi) \to \Box_X \varphi \to \Box_X \psi$	(12)	$\vdash \Box_X (\varphi \to \psi) \to \Box_X \varphi \to \Box_X \psi$
(13)	$\vdash \boxminus_X (\varphi \to \psi) \to \boxminus_X \varphi \to \boxminus_X \psi$	(13)	$\vdash \boxplus_X (\varphi \to \psi) \to \boxplus_X \varphi \to \boxplus_X \psi$
(I)	$\dfrac{\vdash \varphi \to \psi \qquad \vdash \varphi}{\vdash \psi}$	(I)	$\dfrac{\vdash \varphi \to \psi \qquad \vdash \varphi}{\vdash \psi}$
(II)	$\dfrac{\vdash \varphi}{\vdash \Box_X \varphi}$	(II)	$\dfrac{\vdash \varphi}{\vdash \Box_X \varphi}$
(III)	$\dfrac{\vdash \varphi}{\vdash \boxminus_X \varphi}$	(III)	$\dfrac{\vdash \varphi}{\vdash \boxplus_X \varphi}$

Table 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 §1.7, it is harder to motivate $\mathcal{L}^{\boxplus}(\Phi)$. However, we regard each fragment as worthy of study in its own right.

Table 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$^{\boxplus}$.

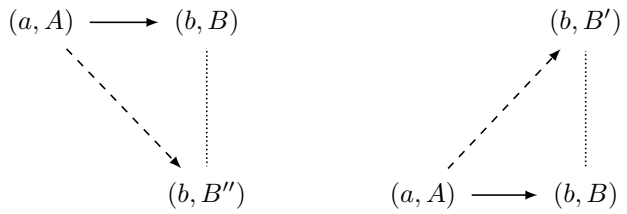
From these axioms, we shall define two sorts of EvIL models the correspond to the properties defined by the above axiom systems.

Definition 2.3.34. *The following properties specify* **⊟EviL** *and* **⊞EviL** *Kripke structures:*

⊟EviL	**⊞EviL**
$(I)^{\boxminus}$ $\sqsupseteq$ *is reflexive*	$(I)^{\boxplus}$ $\sqsubseteq$ *is reflexive*
$(II)^{\boxminus}$ $\sqsupseteq$ *is transitive*	$(II)^{\boxplus}$ $\sqsubseteq$ *is transitive*
$(III)^{\boxminus}$ $\sqsupseteq$ *is anti-symmetric*	$(III)^{\boxplus}$ $\sqsubseteq$ *is anti-symmetric*
$(IV)^{\boxminus}$ $w \sqsupseteq v$ *if and only if* $v \sqsubseteq w$	$(IV)^{\boxplus}$ $w \sqsubseteq v$ *if and only if* $v \sqsupseteq w$
$(V)^{\boxminus}$ *If* $w \sqsupseteq v$ *then* $(w \in V(p)$ *if and only if* $v \in V(p))$	$(V)^{\boxplus}$ *If* $w \sqsubseteq v$ *then* $(w \in V(p)$ *if and only if* $v \in V(p))$
$(VI)^{\boxminus}$ $(R\circ \sqsubseteq) \subseteq R \subseteq (R\circ \sqsupseteq)$	$(VI)^{\boxplus}$ $(R\circ \sqsubseteq) \subseteq R \subseteq (R\circ \sqsupseteq)$
$(VII)^{\boxminus}$ $(\sqsupseteq \circ R) \subseteq R \subseteq (\sqsubseteq \circ R)$	$(VII)^{\boxplus}$ $(\sqsubseteq \circ R) \subseteq R \subseteq (\sqsupseteq \circ R)$
$(VIII)^{\boxminus}$ *If* $w \sqsupseteq v$ *and* $v \in P$ *then* vRw	$(VIII)^{\boxplus}$ *If* $w \sqsubseteq v$ *and* $v \in P$ *then* vRw
$(IX)^{\boxminus}$ *If* $w \in P$ *and* $w \sqsupseteq v$ *then* $v \in P$	$(IX)^{\boxplus}$ *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 §2.2.3, we may naturally visualize certain properties in commutative diagrams:

- Property $(VII)^{\boxminus}$ is depicted as Fig. 14(a), which is the same as Fig. 7(b)
- Property $(VII)^{\boxplus}$ is depicted as Fig. 14(a), which is the same as Fig. 7(c)



(a) – $(VII)^{\boxminus}$ – (b) – $(VII)^{\boxplus}$ –

Figure 14: Visualizations of the relationships in Proposition 2.2.16

We may recall that the commutative diagrams depicted split the original EviL property (VI); 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 $(VIII)^{\boxminus}$ and $(VIII)^{\boxplus}$. Note that in the presence of the other properties of ⊟EviL and ⊞EviL Kripke structures, we may observe that $(VIII)^{\boxminus}$ and $(VIII)^{\boxplus}$ are logically equivalent.

- ⊟EvIL and ⊞EvIL Kripke structures *weaken* $(VII)'$ to $(VII)^{\boxminus}$ and $(VII)^{\boxplus}$, respectively.
- With the exception of $(VII)^{\boxminus}$ and $(VII)^{\boxplus}$, 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 2.3.35. *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* $\mathbb{M} \Vdash \varphi$*.*

Moreover, we shall write

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

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

Theorem 2.3.36 (⊟/⊞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 2.3.3 from §2.3.2. 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)
$(IV)^{\boxminus}$ – Since the canonical model construction in this case does not specify $\sqsubseteq$, we shall define $w \sqsubseteq v$ if and only if $v \sqsupseteq w$	$(IV)^{\boxplus}$ – Since the canonical model construction in this case does not specify $\sqsupseteq$, we shall define $w \sqsupseteq v$ if and only if $v \sqsubseteq w$
$(V)^{\boxminus}$ corresponds to the EviL$^{\boxminus}$ axioms (6) and (7)	$(V)^{\boxplus}$ corresponds to the EviL$^{\boxplus}$ axioms (6) and (7)
$(VI)^{\boxminus}$ corresponds to the EviL$^{\boxminus}$ axiom (8)	$(VI)^{\boxplus}$ corresponds to the EviL$^{\boxplus}$ axiom (8)
$(IX)^{\boxminus}$ corresponds to the EviL$^{\boxminus}$ axiom (11)	$(IX)^{\boxplus}$ corresponds to the EviL$^{\boxplus}$ axiom (11)
$(III)^{\boxminus}$ – Just as in the case of the proof of Theorem 2.3.3, this step follows from the above properties, the Truth Lemma, and an inductive argument	$(III)^{\boxplus}$ – Just as in the case of the proof of Theorem 2.3.3, this step follows from the above properties, the Truth Lemma, and an inductive argument
$(VII)^{\boxminus}$ corresponds to the EviL$^{\boxminus}$ axiom (9)	$(VII)^{\boxplus}$ corresponds to the EviL$^{\boxplus}$ axiom (9)
$(VIII)^{\boxminus}$ corresponds to the EviL$^{\boxminus}$ axiom (10)	$(VIII)^{\boxplus}$ 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 2.3.37 ($\ominus$ and $\oplus$ Bisimulators)**.** *Let $\mathbb{M}$ be a Kripke model, then define:*

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

where:

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

where:

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

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

Definition 2.3.38. *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 2.3.39 (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 $(VII)^{\boxminus}$

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

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

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

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

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

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

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 2.3.39 illustrates that downsets and upsets are similar in many respects to islands, just as we saw in Lemma 2.3.27, the Island Lemma, from §2.3.6. 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 2.3.40. *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 2.3.5, that we gave in 2.3.3.

Proposition 2.3.41. *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 2.3.42. *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 2.3.43. *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 $(IV)^{\boxminus}$.

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

$(III)'$ asserts that $\sqsubseteq^{\ominus}$ is a partial order. To show this, by the above all we have to show is that $\sqsubseteq^{\ominus}$ is anti-symmetric. We shall use that $\mathbb{M}$ makes true $(III)^{\boxminus}$ and $(IV)^{\boxminus}$.

Assume that $(a,b) \sqsubseteq^{\ominus} (c,d)$ and $(b,a) \sqsubseteq^{\ominus} (d,c)$. We want to show that $a = d$ and $b = c$. By construction it must be that $a = d$. We also have by construction that $b \sqsubseteq^{\mathbb{M}} c$, $c \sqsubseteq^{\mathbb{M}} b$. As we noted, $\sqsubseteq^{\mathbb{M}}$ is anti-symmetric by $(III)^{\boxminus}$ and $(IV)^{\boxminus}$, hence $b = c$.

$(IV)'$ 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 $(V)^{\boxminus}$ for $\mathbb{M}$.

$(VI)'$ asserts

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

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

$(VII)'$ asserts

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

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

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

By $(IV)'$, all that is left to show is that

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

So assume that

QED

We note that we cannot generalize the above to multiple agents. This because ...

2.3.10 Universal Modality

2.3.11 Lattice of Logics & Complexity

In this section, we discuss the relationship of the various EvиL logics developed in this section. Using the observed relationships, as well as our previously derived small model properties, we shall present some basic known complexity results for satisfiability in these various logics.

We shall first prove the following lemma:

Lemma 2.3.44. EvиL, EvиL$^{\boxminus}$ *and* EvиL$^{\boxplus}$ *with a single agent are all conservative extensions of the basic modal logic with just axiom K. That is, if* $\nvdash_K \varphi$ *then* $\nvdash_{\text{EvиL}} \varphi$ *and similarly for the fragments* EvиL$^{\boxminus}$ *and* EvиL$^{\boxplus}$.

EvиL *with* $m > n$ *agents is a conservative extension of* EvиL *with* n *agents, and likewise for the fragments* EvиL$^{\boxminus}$ *and* EvиL$^{\boxplus}$

Proof. Assume that $\nvdash_K \varphi$, then we know from modal logic that there's a finite Kripke Structure $\mathbb{M} := \langle W, V, R\rangle$ such and a world $w \in W$ such that $\mathbb{M}, w \nvdash \varphi$. Now extend $\mathbb{M}$ to $\mathbb{M}' := \langle W, V, P, R_\Box, R_\boxminus, R_\boxplus\rangle$ where

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

This model is trivially completely EvIL. Moreover we know that $\mathbb{M}$ is an elementary submodel of $\mathbb{M}'$, so $\mathbb{M}', w \nvdash \varphi$. Hence by the Lemma 2.3.30 we have a model $\mathfrak{M}$ and $(a, A) \in \mathfrak{M}$ such that $\mathfrak{M}, (a, A) \not\models \varphi$; so by soundness for EvIL we have the desired result.

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

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

QED

By similar arguments, EvIL is a conservative extension of EvIL$^\boxminus$ and EvIL$^\boxplus$, and that all three of these are conservative extensions of K. This is summerized in the Fig. 15.

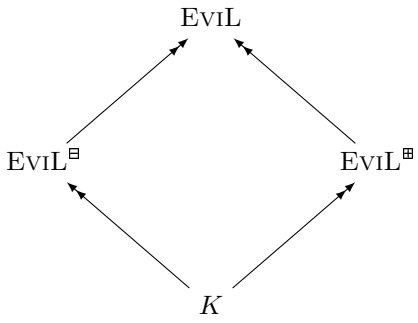


Figure 15: EvIL conservative extensions of K

Lemma 2.3.45. EvIL *is* PSPACE *hard*

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

B Alternate Semantics

In this section, we shall present an alternative work to the framework proposed in §1.3. 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 B.0.1. *Define by recursion the following two truth relations:*

First relation:

$\mathfrak{M}, w \Vdash p \iff p \in V(w)$

$\mathfrak{M}, w \Vdash \varphi \wedge \psi \iff$ *both* $\mathfrak{M}, w \Vdash \varphi$ *and* $\mathfrak{M}, w \Vdash \psi$

$\mathfrak{M}, w \Vdash \varphi \vee \psi \iff$ *either* $\mathfrak{M}, w \Vdash \varphi$ *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:

$\mathfrak{M}, w \nVdash p \iff p \notin V(w)$

$\mathfrak{M}, w \nVdash \varphi \wedge \psi \iff$ *either* $\mathfrak{M}, w \nVdash \varphi$ *or* $\mathfrak{M}, w \nVdash \psi$

$\mathfrak{M}, w \nVdash \varphi \vee \psi \iff$ *both* $\mathfrak{M}, w \nVdash \varphi$ *and* $\mathfrak{M}, w \nVdash \psi$

$\mathfrak{M}, w \Vdash\!\!\!| \neg\varphi \iff \mathfrak{M}, w \nVdash \varphi$

$\mathfrak{M}, w \Vdash\!\!\!| \Box\varphi \iff$ *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 $\mathfrak{M}, w \nVdash\!\!\!| \Box p$.

However, bivalence has a convenient semantic characterization:

Proposition B.0.2. *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:

$$\begin{aligned} \mathfrak{M}, w \Vdash \varphi &\iff \mathbb{M}^{\mathfrak{M}}, w \Vdash_{\Box} \varphi \\ &\iff \mathfrak{M}, w \nVdash_{\Box} \neg\varphi \\ &\iff \mathfrak{M}, w \nVdash \neg\varphi \quad \text{(hypothesis)} \\ &\iff \mathfrak{M}, w \Vdash\!\!\!| \varphi \end{aligned}$$

QED

Corollary B.0.3. *If $\Vdash$ and $\Vdash\!\!\!|$ 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 $\Vdash\!\!\!|$ 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 B.0.4. *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, recalling on the remarks presented in §1.2, it is wrong for agents to be able to have everything they believe in their minds; this is about as bad as the thermometer theory of knowledge. 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 B.0.5. *Let $\mathbb{M}^{\mathfrak{M}}$ be defined as in Prop. B.0.2. 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 $\mathfrak{M}, w \Vdash\!\!\!| \varphi$*

Hence we have $\Vdash$ and $\Vdash\!\!\!|$ 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

$$\begin{aligned} \mathbb{M}^{\mathfrak{M}}, w \Vdash_{\Box} \neg\varphi &\iff \mathbb{M}^{\mathfrak{M}}, w \nVdash_{\Box} \varphi \\ &\implies \mathfrak{M}, w \Vdash\!\!\!\!\!\!\sim\ \varphi \quad \text{(by the inductive step)} \\ &\iff \mathfrak{M}, w \Vdash \neg\varphi \end{aligned}$$

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 B.0.6. *$\vdash_K \varphi$ 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 $w R^{\mathbb{M}'} v$ if and only if $\mathbb{M}', v \Vdash_{\Box} \bigvee_{u \in R^{\mathbb{M}'}[w]} \rho(u)$, whence we have that $w R^{\mathbb{M}'} v$ 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.

C An Application of Pure Model Theory to EviL Semantics

Recall that (VI), presented in Prop. 2.2.16 in §2.2.3 states:

$$(R^{\mathfrak{M}}_X \circ \sqsubseteq^{\mathfrak{M}}_X) \subseteq R^{\mathfrak{M}}_X \qquad (VI)$$

Along with this principle, the following philosophical reading was offered:

"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."

In fact, in light of Theorem 2.1.8, the Theorem Theorem, the interplay expressed in (VI) follows from a general model theoretic relationship.

For any given Kripke structure $\mathbb{M}$, define two operators $Mod^{\mathbb{M}} : \wp\mathcal{L}(\Phi, \mathcal{A}) \to \wp(W^{\mathbb{M}})$ and $Th^{\mathbb{M}} : \wp(W^{\mathbb{M}}) \to \wp\mathcal{L}(\Phi, \mathcal{A})$

$$Mod^{\mathbb{M}}(\Delta) = \{x{\in}W \mid \forall\psi{\in}\Delta.\mathbb{M}, x{\Vdash}\psi\}$$
$$Th^{\mathbb{M}}(\nabla) = \{\psi{\in}\mathcal{L}(\Phi, \mathcal{A}) \mid \forall x{\in}\nabla.\mathbb{M}, x{\Vdash}\psi\}$$

We then have, for any $\Delta \in \wp\mathcal{L}(\Phi, \mathcal{A})$ and $\nabla \in \wp(W^{\mathbb{M}})$:

$$\nabla \subseteq Mod^{\mathbb{M}}(\Delta) \text{ if and only if } \Delta \subseteq Th^{\mathbb{M}}(\nabla)$$

From this, we may observe that these two operations form what is referred as an *antitone Galois connection*, between the lattice $\wp(W^{\mathbb{M}})$ and the lattice $\wp\mathcal{L}(\Phi, \mathcal{A})$. It follows from the theory of Galois connections [Rom08, chapter 3] that we have the following two properties:

$$\text{If } \nabla \supseteq \nabla' \text{ then } Th^{\mathbb{M}}(\nabla) \subseteq Th^{\mathbb{M}}(\nabla') \qquad (C.0.1)$$
$$\text{If } \Delta \supseteq \Delta' \text{ then } Mod^{\mathbb{M}}(\Delta) \subseteq Mod^{\mathbb{M}}(\Delta') \qquad (C.0.2)$$

We can see that (VI) follows from (C.0.2). To see this, assume that $(a, A) \sqsupseteq^{\mathfrak{M}}_X (b, B)$. Then

observe:

$$
\begin{aligned}
(a, A) \sqsupseteq_X^{\mathfrak{M}} (b, B) \quad &\Longrightarrow a = b \text{ and } A_X \supseteq B_X && \text{by the definition of } \sqsupseteq_X^{\mathfrak{M}} \\
&\Longrightarrow A_X \supseteq B_X && \text{weakening} \\
&\Longrightarrow Mod^{\mathfrak{M}}(A_X) \subseteq Mod^{\mathfrak{M}}(B_X) && \text{from (C.0.2)} \\
&\Longrightarrow \text{if } \mathfrak{M}, (c, C) \models A_X \text{ then } \mathfrak{M}, (c, C) \models B_X && \text{by the definition of } Mod^{\mathfrak{M}} \\
&\Longrightarrow \text{if } (a, A) R_X^{\mathfrak{M}} (c, C) \text{ then } (b, B) R_X^{\mathfrak{M}} (c, C) && \text{by the definition of } R_X^{\mathfrak{M}}
\end{aligned}
$$

The above line of reasoning illustrates that structural features of EvIL models are consequences of the decision to set $\mathfrak{M}, (a, A) \Vdash \Box\varphi \iff Th(\mathfrak{M}) \cup A \vdash \varphi$.

References

[AGM85] Carlos E. Alchourron, Peter Gardenfors, and David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *The Journal of Symbolic Logic*, 50(2):510–530, June 1985. ArticleType: primary_article / Full publication date: Jun., 1985 / Copyright © 1985 Association for Symbolic Logic.

[AHV02] N. Agray, W. Van Der Hoek, and E. De Vink. On BAN logics for industrial security protocols. *Lecture notes in computer science*, page 29–36, 2002.

[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.

[Bel86] Eric Temple Bell. *Men of mathematics*. Simon and Schuster, 1986.

[BM96] Jon Barwise and Lawrence Stuart Moss. *Vicious circles*. CSLI Publications, 1996.

[Boo95] George Boolos. *The logic of provability*. Cambridge University Press, 1995.

[Bro36] Sir Thomas Browne. *The garden of Cyrus*. printed in the year, 1736.

[BRV01] P. Blackburn, M. De Rijke, and Y. Venema. *Modal logic*. Cambridge Univ Pr, 2001.

[BS08] Alexandru Baltag and Sonja Smets. Probabilistic dynamic belief revision. *Synthese*, 165(2):179–202, November 2008.

[CF04] Earl Conee and Richard Feldman. *Evidentialism*. Oxford University Press, USA, June 2004.

[Cla34] É Clapeyron. Mémoire sur la puissance motrice de la chaleur. *J. l'ecole polytechnique*, 14:153–190, 1834.

[Den98] Daniel Dennett. *The Intentional Stance*. MIT Press, Cambridge Mass, 7th edition, 1998.

[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.

[Eme08] Ralph Waldo Emerson. *Essays — First Series*. December 2008. LoC Class PS: Language and Literatures: American and Canadian literature.

[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.

[GG02] Dov M. Gabbay and F. Guenthner. *Handbook of Philosophical Logic: Volume 6*. Springer, Dordrecht [u.a.], 2nd edition, May 2002.

[Hal99] Joseph Y. Halpern. Set-theoretic completeness for epistemic and conditional logic. *Annals of Mathematics and Artificial Intelligence*, 26(1-4):1–27, 1999.

[Hil22] David Hilbert. Neubegründung der mathematik. erste mitteilung. *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, 1(1):157–177, December 1922.

[Hin69] Jaakko K. Hintikka. *Knowledge and Belief*. Cornell Univ. Pr., 1969.

[HMdV05] A. Hommersom, J. J Meyer, and E. de Vink. Toward reasoning about security protocols: A semantic approach. *Electronic Notes in Theoretical Computer Science*, 126:53–75, 2005.

[HMV04] A. Hommersom, J. J Meyer, and E. De Vink. Update semantics of security protocols. *Synthese*, 142(2):229–267, 2004.

[Hof79] Douglas Hofstadter. *Go?del, Escher, Bach : an eternal golden braid*. Basic Books, New York, 1979.

[HS06] V. F Hendricks and J. Symons. Where's the bridge? epistemology and epistemic logic. *Philosophical Studies*, 128(1):137–167, 2006.

[KL86] Sarit Kraus and Daniel Lehmann. Knowledge, belief and time. In *Automata, Languages and Programming*, pages 186–195. 1986.

[Koo03] Barteld P. Kooi. Probabilistic dynamic epistemic logic. *Journal of Logic, Language and Information*, 12(4):381–408, 2003.

[KP82] L. Kirby and J. Paris. Accessible independence results for peano arithmetic. *Bulletin of the London Mathematical Society*, 14(4):285, 1982.

[Len78] Wolfgang Lenzen. *Recent work in epistemic logic*. North-Holland, 1978.

[Lev84] H. J Levesque. A logic of implicit and explicit belief. In *Proceedings of the National Conference on Artificial Intelligence*, pages 198–202, 1984.

[LH59] E. J. Lemmon and G. P. Henderson. Symposium: Is there only one correct system of modal logic? *Proceedings of the Aristotelian Society, Supplementary Volumes*, 33:23–56, 1959. ArticleType: primary_article / Full publication date: 1959 / Copyright © 1959 The Aristotelian Society.

[LL51] Clarence Irving Lewis and Cooper Harold Langford. *Symbolic Logic*. Dover Publications, 1951.

[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.

[Pri06] Graham Priest. *Doubt truth to be a liar*. Clarendon Press; Oxford University Press, Oxford ; New York, 2006.

[Qui51] W. V. Quine. Two dogmas of empiricism. *The Philosophical Review*, 60(1):20–43, January 1951. ArticleType: primary_article / Full publication date: Jan., 1951 / Copyright © 1951 Cornell University.

[Ran82] V. Rantala. Impossible worlds semantics and logical omniscience. *Intensional Logic: Theory and Applications*, 1982.

[Rom08] Steven Roman. *Lattices and Ordered Sets*. Springer, 1 edition, September 2008.

[Rub98] Ariel Rubinstein. *Modeling Bounded Rationality*. MIT Press, 1998.

[Rum02] Donald H. Rumsfeld. Defense.gov news transcript: DoD news briefing - secretary rumsfeld and gen. myers. http://www.defense.gov/transcripts/transcript.aspx?transcriptid=2636, February 2002.

[SR99] Jonathan D. H. Smith and Anna B. Romanowska. *Post-modern algebra*. Wiley-IEEE, 1999.

[vB91] J. van Benthem. Reflections on epistemic logic. *Logique Anal., Nouv. Sér.*, 34(133-134):5–14, 1991.

[vB03] Johan van Benthem. Conditional probability meets update logic. *Journal of Logic, Language and Information*, 12(4):409–421, 2003.

[vB10] Johan van Benthem. *Modal Logic for Open Minds*. Center for the Study of Language and Information, February 2010.

[vBGK09] Johan van Benthem, Jelle Gerbrandy, and Barteld Kooi. Dynamic update with probabilities. *Studia Logica*, 93(1):67–96, October 2009.

[vBV09] J. van Benthem and F. R Velázquez-Quesada. Inference, promotion, and the dynamics of awareness. *ILLC Amsterdam. To appear in Knowledge, Rationality and Action*, 2009.

[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.

[Whi08] Walt Whitman. *Leaves of Grass*. August 2008. LoC Class PS: Language and Literatures: American and Canadian literature.