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 (23), *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: $\Box\phi \to \Box\Box\phi$ - "If the agent knows a fact, then she knows that she knows this."
- Negative: $\Diamond\phi \to \Box\Diamond\phi$ - "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 (39):

> 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: $\Box\phi \to \phi$ - "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 $S5$ (31). 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, (20) and (38), chapter 3).

Although this view of knowledge has been the focus of exhaustive research, even finding industrial application such as in (1) and (24, 25), 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, I elaborate the philosophical issues I intend to address, and sketch an epistemic logic framework which tackles them

§2 Next, I give formal details of the system I want to develop. This section climaxes in exposition of weak completeness for my system.

§3 Third, I illustrate philosophical applications of EviL. As a bit of serendipity, it shall be revealed that *intuitionistic logic* is profoundly linked to the thoughts I develop.

§4 Finally, I shall compare my framework to other approaches, and mention lingering philosophical concerns I feel must be addressed.

That being said, I shall now turn to investigating the philosophical issues with epistemic logic, and sketch the solution I shall propose.

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?

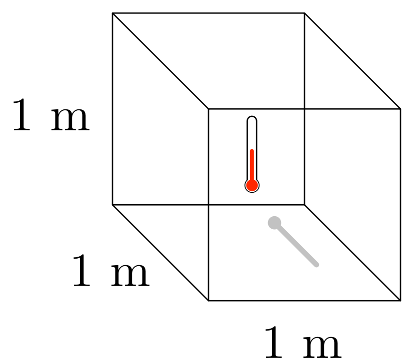


Figure 1: A thermometer in a box

The answer is indeterminate. Recall the *ideal gas law*, originally discovered by Émile Clapeyon (10), 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*, $\approx 8.3\ \mathrm{J{\cdot}K^{-1}{\cdot}mol^{-1}}$

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 $\mathcal{L}_{\text{therm}}$:

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

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* (12), with the proviso that Dennett's original discussion originates around thermostats rather than thermometers as I 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 (43).

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.

I argue that some natural philosophical features of knowledge cannot be captured by this basic approach. If I were to ask a person on the street why she believes a proposition ϕ, I likely would not accept her appeals that she cannot imagine or conceive of the contrary as possible. I would 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, I turn to diagnosing the underpinning of the above principle.

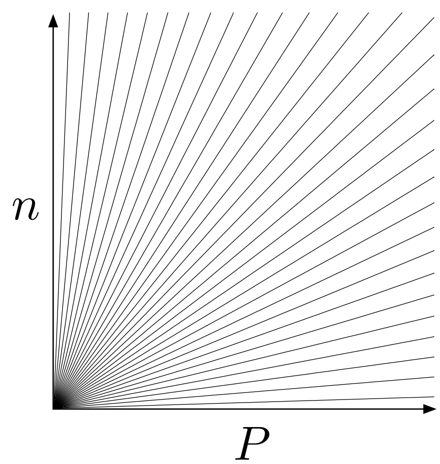


Figure 2: Thermometer information states

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 *explicit justification* for beliefs in epistemic logic. I am not the first person to suggest this. The hunt for logics of explicit justification was initiated in (40)[1]. One framework which has been proposed to achieve this is *Justification Logic* (4, 3, 15, 16). Alternative frameworks for reasoning about implicit/explicit information have also been proposed in (44) and (45).

I offer 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 *Justification Principle*. My idea is to modify the semantics of modal logic to incorporate certain *basic beliefs*, which I take to be 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. In (13, chapter 1), 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.

The view I am presenting would not deny that thermometers or traditional epistemic agents have knowledge, no more than it would deny that I have knowledge of that my right hand has five fingers attached to the end. Rather, my 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 I will demonstrate, this can be done without modifying the basic modal logic syntax.

1.4 Sketch

In this section I intend to give a very informal presentation of the basic elements I intend for 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 $\mathcal{L}_K(\Phi)$:

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

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 (30).

[2] In the proceeding discussion, it should be noted that I refer to *noninferentially justified beliefs* as *basic beliefs* or premises.

Definition 1.4.1.

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

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

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

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

Since the semantics like the above shall be the principle objects of study, I will give how I read them philosophically. In these semantics, instead of thinking of every world individually, I 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 I shall refer to as propositions which she ascents to. I 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*; I 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 I alternately read as *the agent believes phi*, *the agent can deduce ϕ* or *can compose an argument for ϕ.* Like the original formulation of epistemic logic in (23), I assume that agents are *doxastically omniscient* - that is they believe all of the consequences of their beliefs. I 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, I consider this approach to be roughly in line with the *evidentialist* view on epistemology, which (11) describe 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.

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 $\Diamond\phi$ 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, (32)). 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:

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

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. *Assuming that the set of proposition letters Φ is infinite*

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

Here K is basic modal logic.

Proof. Left to right is trivial, so we shall focus on right to left. Assume that $\nvdash_K \phi$, 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 $\mathbb{M}, w \nVdash \phi$ (see (7), chapters 2 & 4 for details of these facts).

Now let $PropVars_\phi$ be the proposition letters that occur as subformulae of ϕ, and let $q : W \hookrightarrow \Phi \backslash PropVars_\phi$ be an injection. In other words q 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]

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

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 $\mathbb{M}, w \nVdash \phi$ then we know that $\Theta, \theta(w) \nVdash \phi$, 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 §2.3.21.

Armed with this, we can see that these semantics are adequate for modeling agents according to my 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 $Th(\mathfrak{M})$:*

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

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

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

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

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

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

[3] In this vein, I shall abbreviate $q(w)$ as q_w. Note that because $PropVars_\phi$ are *finite* and Φ is assumed to be *infinite*, such an inject always exists. This is a consequence of *The Axiom of Choice*.

[4] I am indebted to Johan van Benthem for the invention of this particular function.

[5] This notation consciously imitates the notation employed in (7).

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 $\mathfrak{M}, (a, A) \Vdash \Box\phi$ if and only if $Th(\mathfrak{M}) \cup A \vdash_K \phi$.*

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

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

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:

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

The proof of this follows from the fact that if $\vdash_K \phi$ then $\phi \in Th(\mathfrak{M})$, 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:

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

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* I have 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 my future investigations.

1.5 The Human Condition

To supplement to this basic framework, I shall try to illustrate how further inspiration and desiderata can be drawn from the philosophical literature. It should be remarked that I do this in stark contrast to the received view in epistemic logic (29, 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. It's inspiration is not really in what it's like to be a living person; it's more naturally based in artificial intelligence, automata theory, algebra, topology, and other abstract disciplines.

In contrast, I 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 I 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.

Embracing this principle, I shall turn to the development of my thoughts from their 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*. I suggest this can be done by introducing a new operator $\circlearrowleft$ with the following semantics:

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

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:

$$\circlearrowleft \wedge \Box\, \phi.$$

But is this anything like an adequate analysis of knowledge?

No. To illustrate why I shall resort to a thought experiment to motivate why I think to the contrary. Imagine that Charlotte suspects, correctly, that if John has tried to murder on Alex, then Alex has survived. She further learns, correctly, that John has indeed tried to murder Alex. But later, she "learns" some erroneous information asserting Vietnam is south of Malaysia. If we codify all of this as a set C, and let the real world be denoted c and the universe $\mathfrak{M}$, evidently we have $\mathfrak{M}, (c, C) \nVdash \circlearrowleft$, so this previous definition of knowledge fails. But should it? I don't think so; 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* (33). In answer to the problem faced by the above analysis of knowledge, I think guidance can be found in Descartes' *Meditations* (46). 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.

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, I would suggest a new semantic operation:

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

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:

$$⟠(\circlearrowleft \wedge \Box\, \phi)$$

1.8 Contradictions

There's hidden virtue in the previous analysis. To see what it is, I am inspired by the 19th century philosopher Ralph Waldo Emerson, who writes in his essay *Self-Reliance* (14):

> 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? I don't see why not. Under these considerations, I'd embrace the following:

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

Permit me illustrate how the framework I have given accommodates this. My treatment is further inspired by a friend and contemporary of Emerson's, the poet *Walt Whitman*. In *Leaves of Grass* (47), 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 I 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 $\longrightarrow$, this means that $\mathfrak{M}, (b, B) \Vdash A$

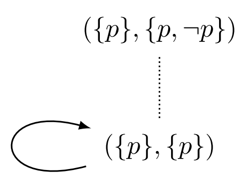


Figure 3: Inconsistent, yet still has knowledge

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 $\mathfrak{M}, (\{p\}, \{p, \neg p\}) \Vdash \diamondminus(\circlearrowleft \wedge \Box\, p)$; 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, (21) 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 (28).

Furthermore, it is conventional to think that rational agents do not hold contradictions.[6] For instance, in (27), $\neg \Box \bot$ 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's incompatible with a human perspective. Hence I 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 19, chapters 1 & 4).

Apart from inconsistency, I do not really accommodate very much irrationality; I will freely admit that frameworks like (36) and and (30) employing *impossible world* semantics are far more accommodating to irrationality than the semantics I am proposing. Regardless, allowing for an agent's beliefs to naturally be inconsistent is already orthogonal to the assumption that agent's are rational.

1.10 Quine

I shall now return to developing my framework. To recap, so far I 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 I have done in Fig. 3, follows the standard practice in lattice theory.

Furthermore, I should point out 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

[6] It should be remarked that (34) 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. I profess sympathy for Priest's perspective; however, I am confident that this does not represent the received view which I am arguing against.

[7] Noted exceptions to this are (36) and (30).

using less pure logic to come to conclusions. I feel this suggests the following - namely, that as we consider levels higher and higher in the poset of an agent's beliefs, this sort of corresponds to embracing an agent's experience and interpretation of their sensory data. But 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* (35):

> 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.*

The emphasis on the last sentence is my addition. The above paragraph importantly anticipates ideas in belief revision theory (such as in (2) and subsequent studies), as well as recent trends in probabilistic dynamic epistemic logic (such as in 41, 43, 5, 26, etc.). However, in the framework that I have so far been developing, 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.

I feel the above observation informs a corresponding perspective on epistemology. If an agent's world view largely rested legends about the Norse gods, I'd 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

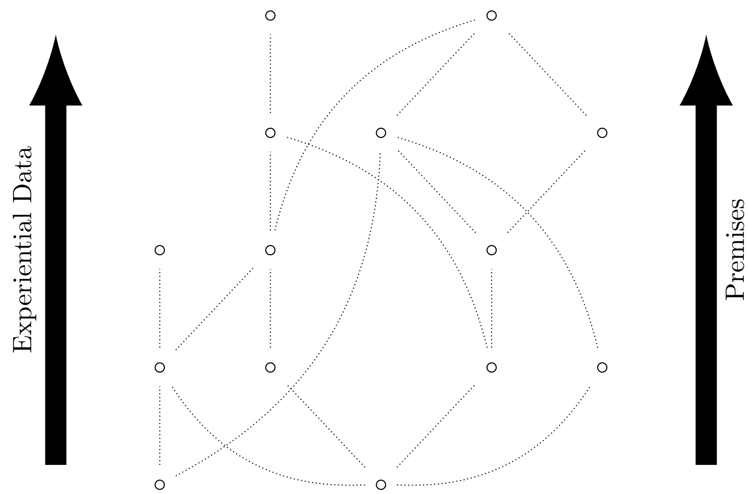


Figure 4: A network of beliefs

in her belief poset, my analysis suggests she doesn't really have knowledge. So while I believe that irrational agents can possess knowledge, as I have argued in §1.9, I do not contend that they *always* posses knowledge. Moreover, I hold that the sort of irrationality that I am considering needn't be superficial - both mundane as well as deeply demented characters can be modeled.

I'll admit that the above essentially presents my own interpretation of Quine's web of belief, which might be contentious. On the other hand, I 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 the final philosophical principle that I intend to extend. Apart from this, I would say from the previous discussion, I would like to 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 I suggested in §1.7. Along these lines, I would suggest the introduction of a new operator $\boxplus$. The semantics for $\boxplus$ are given as follows:

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

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 essentially concludes the sketch of novelties I propose in 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 I feel it is. Dotted lines indicate that I feel the philosophical justification for the higher principle supervenes on the justification of the lower principle. In addition, in further

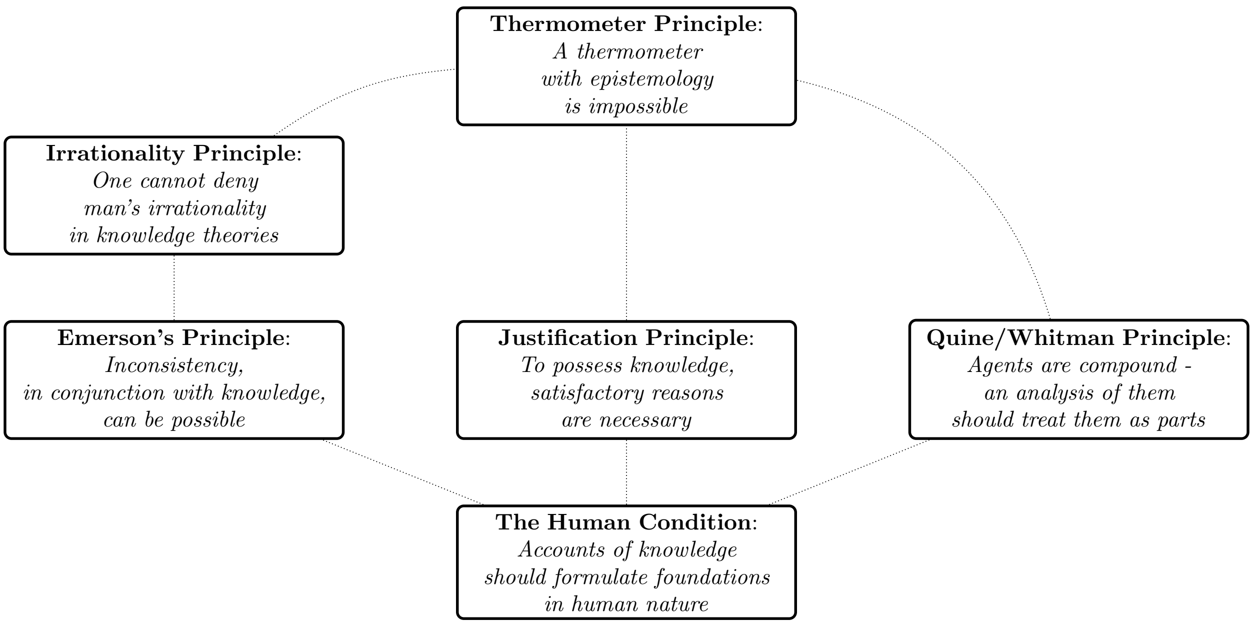


Figure 5: A visualization of the relationship of the principles I have suggested

development of the framework sketched in §1.3, I shall want the following criteria, based on my 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. Idealy, I would like my semantics to correspond to a provably terminating computation, granting certain non-deterministic operations such as a *choice operator* ε, as described in (22).

For a set of beliefs A:

§1.6 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, I have decided to call the logic I shall develop *Evidentialist Logic*, or EviL for brevity.

2 Introduction to EviL

With the philosophical intuitions and scaffolding provided from §1, I shall now turn to giving a precise account of my previously developed ideas. This shall be done in three movements:

§2.1 In the first section I will 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 I develop several topics in the pure theory of EviL which I consider a bit beyond the bare essentials.

§2.3 In this section, completeness and decidability is discussed in relation to EviL and two sublogics.

2.1 Basic EviL

2.1.1 Grammar & Semantics

In this section I turn to developing the formal semantics for EviL with a single agent. I imagine the object of study in EviL is an agent, which I 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, I shall discard the previous definition of $(\Vdash)$ I suggested, 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:*

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

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*

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

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:*

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

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:*

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

Remark 2.1.5. *I will write* $\mathfrak{M} \Vdash \phi$ *to mean* $\mathfrak{M}, (a, A) \Vdash \phi$ *for all* $(a, A) \in \mathfrak{M}$*. Further, I will write* $\Vdash \phi$ *to mean* $\mathfrak{M} \Vdash \phi$ *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* $\phi \in \mathcal{L}_0(\Phi)$*. Then:*

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

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

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

...with this, we have the following, mirroring Prop. 1.4.4:

Definition 2.1.7. *Define the following:*

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

Theorem 2.1.8 (Theorem Theorem). *If A is finite, then $\mathfrak{M}, (a, A) \Vdash \Box\phi$ if and only if $Th(\mathfrak{M}) \cup A \vdash_{\text{EviL}} \phi$.*

I shall present $\vdash_{\text{EviL}}$, the logical consequence turnstile for EviL, in §2.3.2.

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 my previous flirtation with paradox. And 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 circularity brings to mind the alchemical love mathematics, exemplified by the following quote due to (9):

> 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.

We cannot stress enough, Theorem 2.1.8 is what EviL is all about. It is basically essential to the perspective of epistemology this essay is intended to investigate.

2.1.2 Intuitions

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

As per the traditional doxastic reading of $\Box\phi$, I read this as asserting "The EviL agent believes ϕ". Because of Theorem 2.1.8, the Theorem Theorem, I shall freely conflate this with the assertion "The EviL agent has an argument for ϕ," which I take to be a proof.

My intuition for how to read $\diamondminus\phi$ 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, I tend to read $\boxminus\phi$ as saying something like "For all the ways that the EviL agent might use her imagination, ϕ holds." I recognize that these interpretations might seem inconsistent – however, I 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 I set aside for a moment my belief that

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

then it seems natural to imagine that

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

In the symbology of EviL formulae, I would code this intuition as

$$\boxminus(\Box\neg g \to \Diamond p).$$

To give another example, if I pretend that it isn't the case that:

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

I might be able to imagine a scenario where

$$\text{I am swimming comfortably in the Amstel river} \tag{r}$$

But not really. I really can't really swim at ease in the Amstel, not just because it has tons of garbage, but also because

$$\text{I don't own a bathing suit,} \tag{b}$$

Frankly, I am not so bold that I could go skinny dipping in Amstel without that being awkward. Hence I would say in the language of EvIL that:

$$\neg \boxminus (\Box\neg f \to \Diamond r)$$

This is because I can cast into doubt the assumption of the filthiness of the canals of Amsterdam, while still retaining my belief that I don't have a bathingsuit, so swimming in Amstel would still be awkward for me. In symbols, I would write express this sentiment as the following expression:

$$\diamondminus(\Diamond\neg f \wedge \Box b \wedge \neg\Diamond r)$$

Further, my intuition for how to read $\diamondplus\phi$ is "If the EvIL agent were to remember something, then ϕ would hold." For instance, I can think of an instance where I woke up and searched myself for my bike keys. To my horror, they weren't there – in I immediately assumed that I might have left my keys in the lock on my bike, and figured there was a fair likelihood that

$$\text{my bike has been stolen because I left the keys in it.} \tag{s}$$

But once I recalled that

$$\text{I had lent my bike to a friend,} \tag{l}$$

my fear subsided. I would have said that prior to remembering, while I thought it might be possible that my bike was stolen due to my negligence, if I remembered what I had done then I no longer would have entertained that possibility. I would express this observation as:

$$\Diamond s \wedge \boxplus(\Box l \to \Box\neg s)$$

I 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 doesn't 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, I 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, I sometimes read $\boxminus\phi$ closer to the formal semantics, as saying something like "for all subsets of the agent's beliefs, ϕ holds" and dually for $\boxplus\phi$. This is admittedly even less natural than the reading of remembering as the opposite of casting into doubt. So be it;

I am comfortable with EviL agents being at best twisted cartoon versions of actual people, who actually have minds and engage in remembering, imagining, and other similar activities. After all, according to the 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. I 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

$$\Vdash \circlearrowleft \leftrightarrow \boxminus \circlearrowleft \tag{2.1.1}$$

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:

$$\nVdash \circlearrowleft \rightarrow \boxplus \circlearrowleft \tag{2.1.2}$$

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 (2.1.1), 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

$$\Vdash \circlearrowleft \rightarrow \Box\phi \rightarrow \phi \tag{2.1.3}$$

The formula (2.1.1) 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:

$$\Vdash \Diamond\phi \rightarrow \boxminus\Diamond\phi \tag{2.1.4}$$

I read (2.1.4) as saying something like "If the EviL agent can conceive/imagine 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

$$\Vdash \Box\phi \rightarrow \boxplus \Box \phi \tag{2.1.5}$$

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:

$$\Vdash \diamondminus^{+}\phi \to \diamondminus\phi \tag{2.1.6}$$

Here $^{+}$ is taken from the syntax for *regular expressions* commonly used in computer science and UNIX programming to mean "one or more" (18). Similarly, I 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:

$$\Vdash \phi \to \diamondminus\phi \tag{2.1.7}$$

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:

$$\Vdash \Box\phi \to \Box \boxminus \phi \tag{2.1.8}$$

$$\Vdash \Box\phi \to \Box \boxplus \phi \tag{2.1.9}$$

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., $\Diamond\phi$ holds). This is expressed as the following two validities:

$$\Vdash \phi \to \boxminus(\circlearrowleft \to \Diamond\phi) \tag{2.1.10}$$

$$\Vdash \phi \to \boxplus(\circlearrowleft \to \Diamond\phi) \tag{2.1.11}$$

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

$$\Vdash p \to \boxminus p \tag{2.1.12}$$

$$\Vdash p \to \boxplus p \tag{2.1.13}$$

$$\Vdash \neg p \to \boxminus \neg p \tag{2.1.14}$$

$$\Vdash \neg p \to \boxplus \neg p \tag{2.1.15}$$

Hence, EviL is not a *normal* logic. This should admittedly be considered a wart on the semantics, since it appears that it rules out the conventional algebraic duality most modal logics exhibit (see (7), chapter 5).

On the other hand, it is by the same assumption of Cartesian dualism that underlies the non-normality that (2.1.1) 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 has two aspects – it compromises the algebraic elegance of the semantics, while simultaneously giving rise to a philosophically appealing principle.

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, I presented the structural validities of EviL from a philosophical perspective. That being the case, my manner of presentation followed my intuition, which I admit is altogether unorganized. In this section, I shall give the validities of EviL a more systematic presentation. In doing so, I shall showcase an elimination theorem, that I feel sits at the heart of EviL.

To start, the following lemma summerizes the structural validities that I will be studying in the subsequent discussion:

Lemma 2.2.1. *The following validities hold for all* EviL *models:*

$$\begin{array}{ll}
\VDash \boxminus p \leftrightarrow p & \VDash \boxplus p \leftrightarrow p \\
\VDash \boxminus \neg p \leftrightarrow \neg p & \VDash \boxplus \neg p \leftrightarrow \neg p \\
\VDash \boxminus \Diamond \phi \leftrightarrow \Diamond \phi & \VDash \boxplus \Box \phi \leftrightarrow \Box \phi \\
\VDash \diamondplus \Diamond \phi \leftrightarrow \Diamond \phi & \VDash \diamondminus \Box \phi \leftrightarrow \Box \phi \\
\VDash \boxminus \boxminus \phi \leftrightarrow \boxminus \phi & \VDash \boxplus \boxplus \phi \leftrightarrow \boxplus \phi \\
\VDash \boxminus \diamondplus \phi \leftrightarrow \diamondplus \phi & \VDash \boxplus \diamondminus \phi \leftrightarrow \diamondminus \phi \\
\VDash \boxminus \circlearrowleft \leftrightarrow \circlearrowleft & \VDash \boxplus \neg \circlearrowleft \leftrightarrow \neg \circlearrowleft
\end{array}$$

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* $\VDash \phi \leftrightarrow \psi$ *is a validity, then* $\VDash \chi \leftrightarrow \chi[\phi/\psi]$ *is a validity for any* $\chi \in \mathcal{L}(\Phi)$.

Next, I offer two sublanguages of the main language of EviL:

Definition 2.2.3. *Define the following fragments:*[8]

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

$$\phi ::= p \mid \neg p \mid \top \mid \bot \mid \circlearrowleft \mid \phi \wedge \psi \mid \phi \vee \psi \mid \Diamond \phi \mid \boxminus \phi \mid \diamondplus \phi$$

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

$$\phi ::= \neg p \mid p \mid \bot \mid \top \mid \neg \circlearrowleft \mid \phi \vee \psi \mid \phi \wedge \psi \mid \Box \phi \mid \diamondminus \phi \mid \boxplus \phi$$

[8] I was inspired to look at the fragment $\mathcal{L}_A(\Phi)$ by thinking about the continuous fragment of μPML (17).

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:*

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

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

Theorem 2.2.5 (Duality)**.** *Observe that for all* $\phi \in \mathcal{L}_A(\Phi)$ *and* $\psi \in \mathcal{L}_B(\Phi)$*,* $(\phi^B)^A = \phi$ *and* $(\psi^A)^B = \psi$*. Moreover, we have the following validities:* $\VDash \neg(\phi^B) \leftrightarrow \phi$ *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, I present what I feel is the natural consequence of the logical equivalence given in Lemma 2.2.1:

Definition 2.2.6. *If* $\phi \in \mathcal{L}_A(\Phi) \cup \mathcal{L}_B(\Phi)$ *then let* ϕ^* *be the same formula, with all instances of* $\boxplus$*,* $\boxminus$*,* $\diamondminus$ *and* $\diamondplus$ *eliminated. That is,* $(\cdot)^*$ *has the following recursive definition:*

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

Theorem 2.2.7 (EviL Elimination)**.** *For all* $\phi \in \mathcal{L}_A(\Phi)$ *or* $\phi \in \mathcal{L}_B(\Phi)$*, we have the following validity:*

$$\VDash \phi \leftrightarrow \phi^*$$

Proof. The proof proceeds in three steps.

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

$$\VDash \boxminus\phi \leftrightarrow \phi \qquad \VDash \diamondplus\phi \leftrightarrow \phi$$

- 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 $\diamondplus$ for the connective $\vee$. This is because in each case one may simply use distribution, such as can be done here:

$$\begin{aligned} \VDash \boxminus(\phi \wedge \psi) &\leftrightarrow \boxminus\phi \wedge \boxminus\psi \\ &\leftrightarrow \phi \wedge \psi \end{aligned}$$

  On the other hand, $\vee$ is more interesting for $\boxminus$, and dually $\wedge$ for $\diamondplus$. Using induction, Lemma 2.2.1, and subsitution, and distribution, we have the line of reasoning:

$$\begin{aligned} \VDash \boxminus(\phi \vee \psi) &\leftrightarrow \boxminus(\diamondplus\phi \vee \diamondplus\psi) \\ &\leftrightarrow \boxminus\diamondplus(\phi \vee \psi) \\ &\leftrightarrow \diamondplus(\phi \vee \psi) \\ &\leftrightarrow \diamondplus\phi \vee \diamondplus\psi \\ &\leftrightarrow \phi \vee \psi \end{aligned}$$

- Case $\Diamond$: Once again, this follows immediately from the validities of Lemma 2.2.1, namely $\VDash \boxminus\Diamond\phi \leftrightarrow \Diamond\phi$ and $\VDash \diamondplus\Diamond\phi \leftrightarrow \Diamond\phi$

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

$$\begin{array}{ll} \VDash \boxminus\diamondplus\phi \leftrightarrow \diamondplus\phi & \VDash \diamondplus\diamondplus\phi \leftrightarrow \diamondplus\phi \\ \VDash \boxminus\boxminus\phi \leftrightarrow \boxminus\phi & \VDash \diamondplus\boxminus\phi \leftrightarrow \boxminus\phi \end{array}$$

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

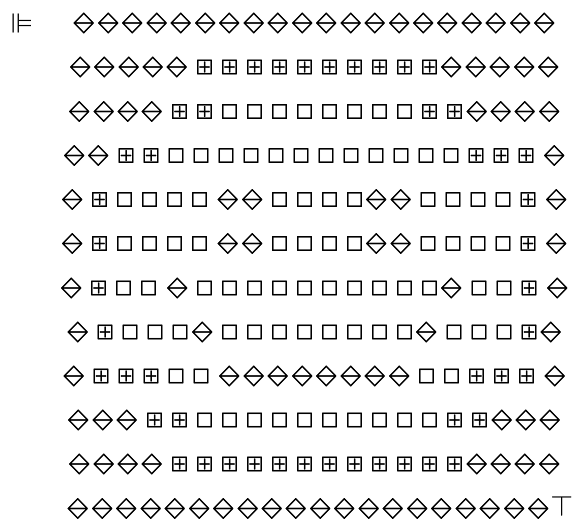
$$\begin{aligned} \VDash \psi &\leftrightarrow \neg(\psi^A) \\ &\leftrightarrow \neg((\psi^A)^*) \\ &\leftrightarrow \neg((\psi^*)^A) \\ &\leftrightarrow \neg(\neg(((\psi^*)^A)^B)) \\ &\leftrightarrow \neg\neg\psi^* \\ &\leftrightarrow \psi^* \end{aligned}$$

QED

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

$$\VDash \boxminus\diamondplus\Diamond t \vee \boxplus\Box\diamondminus\neg t$$

$$\VDash ((\diamondminus\boxplus\Box q \wedge \boxplus\diamondminus\Box q) \vee \Box\boxplus\diamondminus q) \wedge ((\boxplus\Box\diamondminus q \vee \Box\diamondminus\boxplus q) \wedge \diamondminus\Box\boxplus q) \leftrightarrow \Box q$$



The way I 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, I should remark that Theorem 2.2.7 reflects one of the basic themes of EviL - the interplay between belief, reflected by $\Box$, and imagination, reflected by $\Diamond$. I feel that these two phenomena are just two sides of the same coin - furthermore, one couldn't have more natural opposites. Belief and imagination exemplify what I naturally feel are 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. However, I admit that I do not understand what philosophical light is shed by Theorem 2.2.7, if there is indeed any at all.

2.2.2 Multiple Agents

In this section I turn to extending 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. But I will freely admit that my EviL intuitions are principally grounded in the single agent case – I recommend thinking about the multi-agent case as just a generalization of 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:*

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

...where $X \in \mathcal{A}$ 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

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

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.

I shall now provide a formal definition of the semantics for the multi-agent $(\Vdash)$: [9]

Definition 2.2.10.

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

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 $\Box$, $\Diamond$, $\boxminus$, $\boxplus$, $\Diamond\!\!\!\!-$, $\Diamond\!\!\!\!+$ and $\circlearrowleft$ are all replaced with $\Box_X$, $\Diamond_X$, $\boxminus_X$, $\boxplus_X$, $\Diamond\!\!\!\!-_X$, $\Diamond\!\!\!\!+_X$ and $\circlearrowleft_X$ respectively, for any fixed $X \in \mathcal{A}$. Furthermore, compactness still fails, just as presented in §2.3.1.

Finally, there is are two novel validities that these semantics give rise to:

$$\Vdash \Box_X \phi \to \Box_X \boxminus_Y \phi$$
$$\Vdash \Box_X \phi \to \Box_X \boxplus_Y \phi$$

This is just to say, that as the EvIL agent's deliberative process was opaque to her beliefs in the single agent case, as expressed by (2.1.8) and (2.1.9) in §2.1.3, in a similar fashion she cannot read anyone else's mind, nor anyone else hers.

I will admit that little use shall be made of EvIL generalized this way, since the single agent case is far more natural to think about for me. However, it will be of interest to the main theorems of EvIL, which I shall present in §2.3, that they be proved in the widest generality.

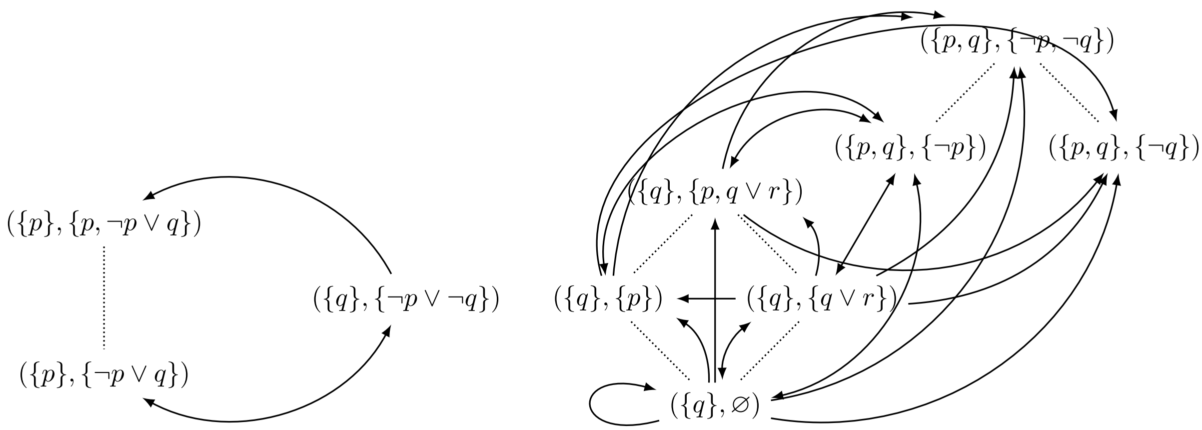
2.2.3 Kripke Structures

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, I will

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

demonstrate that every EviL model corresponds to some highly structured Kripke model, with a minor modification on the standard definition. However, it will turn out that this correspondence is one way - the class of Kripke models for which EviL is strongly complete do not, in general, possess corresponding EviL models.

To elucidate my intuition for understanding EviL models as Kripke models, I would like to return to the visualization technique for EviL models I introduced in §1.8. This involved, roughly, thinking of the EviL models as *posets* with arrows, as I first presented in Fig. 3. I have given additional examples in Figs. 6(a) and 6(b). In all of these depictions, the implicit relational structure of EviL models is given visual expression. So it seems only natural to me 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 $\mathbb{M} = \langle W^{\mathbb{M}}, R^{\mathbb{M}}, \sqsubseteq^{\mathbb{M}}, \sqsupseteq^{\mathbb{M}}, V^{\mathbb{M}}, P^{\mathbb{M}}_{\circlearrowleft}\rangle$ where*[10]*:*

- *$W^{\mathbb{M}}$ is a set of worlds*
- *$R^{\mathbb{M}} : \mathcal{A} \to \wp(W \times W)$, $\sqsubseteq^{\mathbb{M}}: \mathcal{A} \to \wp(W \times W)$, and $\sqsupseteq^{\mathbb{M}}: \mathcal{A} \to \wp(W \times W)$ are $\mathcal{A}$-indexed sets of relations*[11]
- *$V : \Phi \to \wp(W)$ is a predicate letter valuation*
- *$P_{\circlearrowleft} : \mathcal{A} \to \wp(W)$ are sets of worlds indexed by agents*

[10]Where the context is clear, I shall drop $\mathbb{M}$ from the superscripts I am employing.
[11]I will abbreviate $R(X)$, $\sqsubseteq (X)$ and $\sqsupseteq (X)$ as $R(X)$, $\sqsubseteq (X)$ and $\sqsupseteq (X)$ as R_X, $\sqsubseteq_X$ and $\sqsupseteq_X$ respectively.

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

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

Definition 2.2.12. *Let $\mathbb{M}$ be in the class $\mathcal{K}_{\Phi,\mathcal{A},I}$*

$$
\begin{aligned}
\mathbb{M}, w \Vdash p &\iff w \in V^{\mathbb{M}}(p)\\
\mathbb{M}, w \Vdash \phi \to \psi &\iff \mathbb{M}, w \Vdash \phi \textit{ implies } \mathbb{M}, w \Vdash \psi\\
\mathbb{M}, w \Vdash \bot &\iff \textit{False}\\
\mathbb{M}, w \Vdash \Box_X \phi &\iff \forall v \in W^{\mathbb{M}}. w R^{\mathbb{M}}_X v \textit{ implies } \mathbb{M}, v \Vdash \phi\\
\mathbb{M}, w \Vdash \boxminus_X \phi &\iff \forall v \in W^{\mathbb{M}}. w \sqsupseteq^{\mathbb{M}}_X v \textit{ implies } \mathbb{M}, v \Vdash \phi\\
\mathbb{M}, w \Vdash \boxplus_X \phi &\iff \forall v \in W^{\mathbb{M}}. w \sqsubseteq^{\mathbb{M}}_X v \textit{ implies } \mathbb{M}, v \Vdash \phi\\
\mathbb{M}, w \Vdash \circlearrowleft_X &\iff w \in P^{\mathbb{M}}_{\circlearrowleft}(X)
\end{aligned}
$$

Kripke structures can be observe to have a lot less structure than EviL models. However, EviL models can be understood as Kripke structures in disguise. To illustrate this, observe the following lemma:

Definition 2.2.13 ($\mho^{\mathfrak{M}}$ Translation)**.** *Let $\mathfrak{M}$ be an* EviL *model. Define $\mho^{\mathfrak{M}} := \langle \mathfrak{M}, R^{\mathfrak{M}}, \sqsubseteq^{\mathfrak{M}}, \sqsupseteq^{\mathfrak{M}}, V^{\mathfrak{M}}, P^{\mathfrak{M}}_{\circlearrowleft} \rangle$, where*

- $(a, A) R^{\mathfrak{M}}_X (b, B) \iff \forall \psi \in A_X . b \models \psi$
- $(a, A) \sqsupseteq^{\mathfrak{M}}_X (b, B) \iff a = b$ *and* $A_X \supseteq B_X$
- $(a, A) \sqsubseteq^{\mathfrak{M}}_X (b, B) \iff a = b$ *and* $A_X \subseteq B_X$
- $(a, A) \in P^{\mathfrak{M}}_{\circlearrowleft}(X) \iff \forall \psi \in A_X . a \models \psi$

Lemma 2.2.14. *For all $\mathfrak{M}$ and all $(a, A) \in \mathfrak{M}$, $\mathfrak{M}, (a, A) \Vdash \phi$ if and only if $\mho^{\mathfrak{M}}, (a, A) \Vdash \phi$*

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

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

Proposition 2.2.15. *For any* EviL *model $\mathfrak{M}$, $\mho^{\mathfrak{M}}$ has the following properties*[12]*:*

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

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

(III) $\sqsupseteq^{\mathfrak{M}}_X$ *is anti-symmetric*

(IV) $w \sqsupseteq^{\mathfrak{M}}_X v$ *if and only if* $v \sqsubseteq^{\mathfrak{M}}_X w$

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

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

[12]Note that in this we have that $\{w, v\} \subseteq \wp\Phi \times \wp\mathcal{L}_0$ in the subsequent discussion

(VII) $(\sqsubseteq^{\mathfrak{M}}_Y \circ R^{\mathfrak{M}}_X) = R^{\mathfrak{M}}_X = (\sqsupseteq^{\mathfrak{M}}_Y \circ R^{\mathfrak{M}}_X)$

(VIII) $w \in P^{\mathfrak{M}}_{\circlearrowleft}(X)$ *if and only if* $wR^{\mathfrak{M}}_X w$

. . . the situation in (VI) can be visualized in 7(a), while (VII) can be split into Figs. 7(b) and 7(c).

$$\begin{tikzcd}
(a, A') \arrow[r] \arrow[d, dash, dotted] & (b, B) \\
(a, A) \arrow[ur, dashed] &
\end{tikzcd}$$

(a) *(VI)* - $(R^{\mathfrak{M}}_X \circ \sqsubseteq^{\mathfrak{M}}_X) \subseteq R^{\mathfrak{M}}_X$ and $R^{\mathfrak{M}}_X \subseteq (R^{\mathfrak{M}}_X \circ \sqsupseteq^{\mathfrak{M}}_X)$

$$\begin{tikzcd}
 & (b, B') \arrow[d, dash, dotted] \\
(a, A) \arrow[r] \arrow[ur, dashed] & (b, B)
\end{tikzcd}$$

(b) *(VII)* - $(\sqsubseteq^{\mathfrak{M}}_Y \circ R^{\mathfrak{M}}_X) \subseteq R^{\mathfrak{M}}_X$ and $R^{\mathfrak{M}}_X \subseteq (\sqsupseteq^{\mathfrak{M}}_Y \circ R^{\mathfrak{M}}_X)$

$$\begin{tikzcd}
(a, A) \arrow[r] \arrow[dr, dashed] & (b, B) \arrow[d, dash, dotted] \\
 & (b, B'')
\end{tikzcd}$$

(c) *(VII)* - $(\sqsupseteq^{\mathfrak{M}}_Y \circ R^{\mathfrak{M}}_X) \subseteq R^{\mathfrak{M}}_X$ and $R^{\mathfrak{M}}_X \subseteq (\sqsubseteq^{\mathfrak{M}}_Y \circ R^{\mathfrak{M}}_X)$

Figure 7: Visualizations of the relationships in Proposition 2.2.15

Proof. Everything except *(VI)* follows directly from the definitions - so I shall only demonstrate $R^{\mathfrak{M}}_X \subseteq R^{\mathfrak{M}}_X \circ \sqsupseteq^{\mathfrak{M}}_X$.

Suppose that $(a, A)R^{\mathfrak{M}}_X(b, B)$, then evidently $\forall\psi \in A_X.b \models \psi$. Now assume that $(a, A) \sqsupseteq^{\mathfrak{M}}_X (c, C)$. Then we know that $A_X \supseteq C_X$. But then $\forall\psi \in C_X.b \models \psi$, so $(c, C)R^{\mathfrak{M}}_X(b, B)$, which suffices to show the claim. QED

Definition 2.2.16. *A Kripke structure is called* EviL *if it makes true the above properties (I) through (VIII) (with the exception of (III), which is optional).*

These properties are definitive - as shall be demonstrated in §2.3.5, EviL is sound and strongly complete for EviL models.

The Kripke semantics also serve to provide proper intuition behind EviL models. I 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 at world x, agent X's assumptions (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's easier for a world to be incompatible with an agent's evidence.

Finally, while Prop. 2.2.15 presents itself as a sort of representation lemma, the relationship between EviL semantics and Kripke semantics is not reciprocal. Not every Kripke model can be represented as an EviL model. Example 2.2.18 presents an elementary example of this failure of representation. It turns on the following observation:

Lemma 2.2.17. *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. This is an elementary result in the semantics of propositional logic. QED

Example 2.2.18. *Consider a single agent* EviL *Kripke structure* $\mathbb{M} := \langle W, R, \sqsubseteq, \sqsupseteq, V, P\rangle$ *where*

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

...this structure is indicated in Fig. 2.2.18. No EviL *model corresponds to* $\mathbb{M}$*.*

$$w \longrightarrow v$$

Figure 8: $\mathbb{M}$ is a Kripke Structure with no EviL representation

Observe that the above model 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, I can assert that it is impossible for there to be an EviL *structure* $\mathfrak{M}$ *with a world* (a, A) *such that* $\mathbb{M}, w \Vdash \phi$ *if and only if* $\mathfrak{M}, (a, A) \Vdash\!\!\models \phi$*.*

For suppose there were, then I 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)*, I know that for the* b *mentioned in* (*1*)*, 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* $a \not\models A$*, since otherwise I would have* $\mathfrak{M}, (a, A) \Vdash\!\!\models \Diamond\Diamond\top$

This is absurd... since $a = b = \varnothing$ *and* $b \models A$ *then it must be that* $a \models A$*.* $\lightning$

The above one way correspondence is inconvenient - it means that while EviL only enjoys some features from modal logic, it is denied others. Despite this, I hold that EviL enjoys *most* of the benefits of basic modal logic. Furthermore, even though EviL is quite formal in nature, it might rightly be considered an application of traditional modal logic rather than a novel logic in of itself. If the language of EviL had been first order rather than modal, I doubt I would have been tempted to call a logic at all. The novelty of EviL is that it presents semantics that automatically connect truth and derivability (as expressed in Theorem 2.1.8, the Theorem Theorem).

2.3 EviL Completeness

2.3.1 Failure of Compactness

In this section I shall demonstrate that EviL is not compact by giving an example of an infinite set of formulae for which every finite subset is satisfiable while the entirety is not.

Lemma 2.3.1. *Let $\tau : \Phi \to \mathcal{L}$ be defined as follows:*

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

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

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

So let $S \subseteq \tau[\Phi]$ be finite. I shall provide a model that satisfies S. First observe that there is a finite $\Psi \subset \Phi$ such that $S = \tau[\Psi]$. Let $\{q, r, s\} \subseteq \Phi \setminus \Psi$ contain three distinct letters - this can be done, since $\Phi \setminus \Psi$ is infinite. Let $\mathfrak{M} := \{(a, \{q\}), (b, \{r\}), (c, \{s\})\}$, where

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

Using the same visualization convention I introduced in §1.8, $\mathfrak{M}$ can be visualizedin Fig. 9. It is straightforward to check that $\mathfrak{M}, (a, \{q\}) \Vdash \tau(p)$ for all $p \in \Psi$.

On the other hand, suppose there was some model $\mathfrak{N}$ such that $\mathfrak{N}, (a, A) \Vdash \tau(p)$ for all $p \in \Psi$. This implies that $\mathfrak{N}, (a, A) \Vdash p$ for each $p \in \Phi$. Moreover, let $(b, B) \in \mathfrak{N}$ be such that $b \models A$ (one exists since by hypothesis $\mathfrak{N}, (a, A) \Vdash \Diamond p$). By the semantics of $\tau(p)$, it is evident that $\mathfrak{N}, (b, B) \Vdash \neg p$ and that there's a $(c, C) \in \mathfrak{N}$ such that $c \models B$ and $\mathfrak{N}, (c, C) \Vdash p \wedge \Diamond p$. But then from this it must be that a and c both contain exactly the same sentence letters, so $a = b$ which means that $a \models B$ as well, since $B \subseteq \mathcal{L}_0$ (as per the grammar restriction). However it cannot be that $\mathfrak{N}, (a, A) \Vdash \Diamond p$ since $\mathfrak{N}, (a, A) \Vdash \Box\neg p$, which follows from the assumption that $\mathfrak{N}, (a, A) \Vdash \tau(p)$. Thus it is impossible that $\mathfrak{N}, (a, A) \Vdash \tau[\Phi]$. $\lightning$ QED

The above argument illustrates that while EviL semantics present themselves as similar to the traditional Kripke Semantics for modal logic, they aren't the same. The similarity particularly

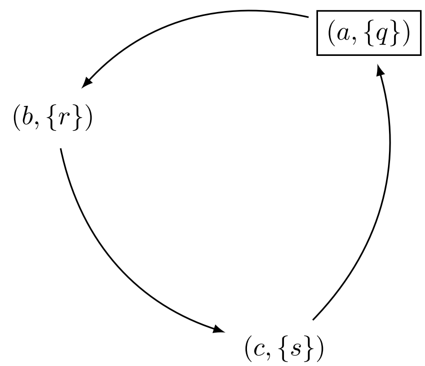


Figure 9: $\mathfrak{M}, (a, \{q\}) \Vdash \tau[\Psi]$

manifests itself when thinking about visualizations like Fig. 9. But if for some reason two world-pairs (a, A) and (b, B) make true the same sentence letters, then $a = b$. Moreover, for any $C \subseteq \mathcal{L}_0$, we have that $a \models C$ if and only if $b \models C$ in this case. This is crucial - even though EvIL is modal, it has no accessibility relations; the role of accessibility relations is instead picked up by basic sets of beliefs C. It is necessarily the case that for a basic set of beliefs C that $C \subseteq \mathcal{L}_0$, due to grammar restriction I decided upon to ensure the semantics of EvIL were well-defined. It is from this basic design choice that my unusual failure of compactness manifests itself. I shall return to considering the relationship of Kripke semantics and EvIL models in §2.2.3.

The failure of compactness, while a fairly basic result in the model theory of EvIL, is far reaching. As a consequence, there is no hope of achieving completeness using infinitary Lindenbaum constructions that are typically employed in modal logic, such as is done in (7, chapter 4), for instance. Hence the completeness theorem for EvIL presented in §2.3 will necessarily have to be finite in nature.

2.3.2 Axiom Systems

In Table 1, I have provided a Hilbert-style axiom system for EvIL. In addition to giving each axiom, I have also provided my own philosophical reading of what each axiom says. One unusual feature of this logic is that it is not *normal*, that is it is not closed under variable substitution.

This logic makes true a variety of relationships between the various modalities, which are given in the following lemma:

Lemma 2.3.2. *We have the following provable equivalences:*

$$\vdash \Box_X \phi \leftrightarrow \boxminus_X \Box_X \phi \qquad \vdash \Box_X \phi \leftrightarrow \Box_X \boxminus_Y \phi \qquad \vdash \Box_X \phi \leftrightarrow \Box_X \boxplus_Y \phi$$
$$\vdash \boxminus_X \phi \leftrightarrow \boxminus_X \boxminus_X \phi \qquad \vdash \boxplus_X \phi \leftrightarrow \boxplus_X \boxplus_X \phi \qquad \vdash \circlearrowleft_X \leftrightarrow \boxminus_X \circlearrowleft_X$$

In addition to the main system presented above, it can be understood to contain two subsystems, corresponding to two fragments of the main grammar:

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

Table 1: A Hilbert style axiom system for EvIL

(1)	$\vdash \phi \to \psi \to \phi$
(2)	$\vdash (\phi \to \psi \to \chi) \to (\phi \to \psi) \to \phi \to \chi$
(3)	$\vdash (\neg\phi \to \neg\psi) \to \psi \to \phi$
(4)	$\vdash \boxminus_X \phi \to \phi$
(5)	$\vdash \boxminus_X \phi \to \boxminus_X \boxminus_X \phi$
(6)	$\vdash p \to \boxminus_X p$
(7)	$\vdash \neg p \to \boxminus_X \neg p$
(8)	$\vdash \Diamond_X \phi \to \boxminus_X \Diamond_X \phi$
(9)	$\vdash \Box_X \phi \to \Box_X \boxminus_Y \phi$
(10)	$\vdash \phi \to \boxminus_X (\circlearrowleft_X \to \Diamond_X \phi)$
(11)	$\vdash \circlearrowleft_X \to \boxminus_X \circlearrowleft_X$
(12)	$\vdash \Box_X(\phi \to \psi) \to \Box_X \phi \to \Box_X \psi$
(13)	$\vdash \boxminus_X(\phi \to \psi) \to \boxminus_X \phi \to \boxminus_X \psi$
(I)	$\dfrac{\vdash \phi \to \psi \qquad \vdash \phi}{\vdash \psi}$
(II)	$\dfrac{\vdash \phi}{\vdash \Box_X \phi}$
(III)	$\dfrac{\vdash \phi}{\vdash \boxminus_X \phi}$

(1)	$\vdash \phi \to \psi \to \phi$
(2)	$\vdash (\phi \to \psi \to \chi) \to (\phi \to \psi) \to \phi \to \chi$
(3)	$\vdash (\neg\phi \to \neg\psi) \to \psi \to \phi$
(4)	$\vdash \boxplus_X \phi \to \phi$
(5)	$\vdash \boxplus_X \phi \to \boxplus_X \boxplus_X \phi$
(6)	$\vdash p \to \boxplus_X p$
(7)	$\vdash \neg p \to \boxplus_X \neg p$
(8)	$\vdash \Box_X \phi \to \boxplus_X \Box_X \phi$
(9)	$\vdash \Box_X \phi \to \Box_X \boxplus_Y \phi$
(10)	$\vdash \phi \to \boxplus_X (\circlearrowleft_X \to \Diamond_X \phi)$
(11)	$\vdash \neg \circlearrowleft_X \to \boxplus_X \neg \circlearrowleft_X$
(12)	$\vdash \Box_X(\phi \to \psi) \to \Box_X \phi \to \Box_X \psi$
(13)	$\vdash \boxplus_X(\phi \to \psi) \to \boxplus_X \phi \to \boxplus_X \psi$
(I)	$\dfrac{\vdash \phi \to \psi \qquad \vdash \phi}{\vdash \psi}$
(II)	$\dfrac{\vdash \phi}{\vdash \Box_X \phi}$
(III)	$\dfrac{\vdash \phi}{\vdash \boxplus_X \phi}$

Table 2: Axiom system EVIL$^{\boxminus}$ and EVIL$^{\boxplus}$ respectively

Definition 2.3.3. *Define $\mathcal{L}^{\boxminus}(\Phi, \mathcal{A})$ as the fragment:*

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

And define $\mathcal{L}^{\boxplus}(\Phi, \mathcal{A})$ as the fragment:

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

Table 2 gives the axioms systems for these two fragments. For now, we shall observe that EVIL extends EVIL$^{\boxminus}$ and EVIL$^{\boxplus}$. In §2.3.7 we shall make this precise.

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

Lemma 2.3.4 (Soundness)**.** *If $\vdash \phi$ then for any model $\mathfrak{M}$ and any $(a, A) \in \mathfrak{M}$ we have that $\mathfrak{M}, (a, A) \models \phi$*

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

- The first step is to construct a Kripke model $\dagger^\phi$ consisting of finite maximally consistent sets of formulae related to ϕ where $\dagger^\phi, w \not\Vdash \phi$ for some world $w \in W^{\dagger^\phi}$. This model will be shown to make true nine properties.

- The second step is to construct a model $⸸^{\dagger^\phi}$ which is bisimular to $\dagger^\phi$. This model also makes true these nine properties as well as an additional tenth property.

- The final third step is to construct an EVIL model $\pentagram_{\phi}^{\dagger^{\dagger^{\phi}}}$. I shall then show that for each $w \in W^{\dagger^{\dagger^{\phi}}}$ there is a corresponding $(a, A) \in \pentagram_{\phi}^{\dagger^{\dagger^{\phi}}}$ such that $\dagger^{\dagger^{\phi}}, w \Vdash \psi$ if and only if $\pentagram_{\phi}^{\dagger^{\dagger^{\phi}}}, (a, A) \models \psi$ for all subformulae ψ of ϕ.

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

2.3.3 Subformula Model Construction

In this section we provide definitions and lemmas related to the subformula construction $\dagger^{\phi}$. I consciously imitate (8) in my approach, as well as the "Fischer-Ladner Closure" used in the completeness theorem of PDL (7).

Definition 2.3.5.

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

Lemma 2.3.6. *By Lemma 2.3.2 we have*

$$\vdash \sim \phi \leftrightarrow \neg\phi \qquad \vdash \boxslash_X \phi \leftrightarrow \boxminus_X \phi \qquad \vdash \boxtimes_X \phi \leftrightarrow \boxplus_X \phi$$

Moreover,

$$\boxslash_X \phi = \boxslash_X \boxslash_X \phi \qquad \boxtimes_X \phi = \boxtimes_X \boxtimes_X \phi$$

Definition 2.3.7. *Let $\delta(\phi) \subseteq \mathcal{A}$ be the set of agents that occur in ϕ*[13]

Definition 2.3.8. *Define $\Sigma(\Delta, \phi)$ using primitive recursion as follows:*

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

Lemma 2.3.9. *$\Sigma(\delta(\phi), \phi)$ is finite. Moreover, we have the following:*

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

[13] In natural language, we read $\delta(\phi)$ as "the dudes mentioned by ϕ."

- *If* $\boxminus_X \phi \in \Sigma(\delta(\phi), \phi)$ *then* $\boxslash_X \phi \in \Sigma(\delta(\phi), \phi)$
- *If* $\boxplus_X \phi \in \Sigma(\delta(\phi), \phi)$ *then* $\boxtimes_X \phi \in \Sigma(\delta(\phi), \phi)$

Definition 2.3.10. *Let* $At(\Psi)$ *denote the maximally consistent subsets of* Ψ

Lemma 2.3.11 (Lindenbaum Lemma)**.** *If* $\Gamma \nvdash \phi$ *and* $\Gamma \subseteq \Sigma(\delta(\phi), \phi)$*, then there is a* $\Gamma' \in At(\Sigma(\delta(\phi), \phi))$ *such that* $\Gamma \subseteq \Gamma'$ *and* $\Gamma' \nvdash \phi$

Definition 2.3.12. *Define* $\dagger^\phi := \langle W^{\dagger^\phi}, V^{\dagger^\phi}, P_X^{\dagger^\phi}, R_{\Box_X}^{\dagger^\phi}, R_{\boxminus_X}^{\dagger^\phi}, R_{\boxplus_X}^{\dagger^\phi} \rangle$ *where:*

$W^{\dagger^\phi} := At(\Sigma(\delta(\phi), \phi))$
$V^{\dagger^\phi}(p) := \{w \in W^{\dagger^\phi} \mid p \in w\}$
$P_X^{\dagger^\phi} := \{w \in W^{\dagger^\phi} \mid \circlearrowleft_X \in w\} \cup \{w \in W^{\dagger^\phi} \mid X \notin \delta(A)\}$
$R_{\Box_X}^{\dagger^\phi} := \{(w, v) \in W^{\dagger^\phi} \times W^{\dagger^\phi} \mid \{\psi \mid \Box_X \psi \in w\} \subseteq v\}$
$R_{\boxminus_X}^{\dagger^\phi} := \{(w, v) \in W^{\dagger^\phi} \times W^{\dagger^\phi} \mid \bigcup\{\{\psi, \boxslash_X \psi\} \mid \boxslash_X \psi \in w\} \subseteq v \wedge \bigcup\{\{\psi, \boxtimes_X \psi\} \mid \boxtimes_X \psi \in v\} \subseteq w\}$
$R_{\boxplus_X}^{\dagger^\phi} := \{(v, w) \in W^{\dagger^\phi} \times W^{\dagger^\phi} \mid \bigcup\{\{\psi, \boxslash_X \psi\} \mid \boxslash_X \psi \in w\} \subseteq v \wedge \bigcup\{\{\psi, \boxtimes_X \psi\} \mid \boxtimes_X \psi \in v\} \subseteq w\}$

Lemma 2.3.13 (Truth Lemma)**.** *For any subformula* $\psi \in \Sigma(\delta(\phi), \phi)$ *and any* $w \in W^{\dagger^\phi}$*, we have that* $\dagger^\phi, w \Vdash \psi$ *if and only if* $\psi \in w$

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

I shall demonstrate the right to left direction for $\boxminus_X$. Assume that $\boxminus_X \psi \notin w$, then $w \nvdash \boxminus_X \psi$. By Lemma 2.3.6 this is true if and only if $w \nvdash \boxslash_X \psi$. Now abbreviate:

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

Now suppose towards a contradiction that $\{\sim \psi\} \cup A \cup B \vdash \bot$. Then $A \cup B \vdash \psi$, and furthermore by Lemma 2.3.6 and rule (III) from the axioms we have that $\boxslash_X A \cup \boxslash_X B \vdash \boxslash_X \psi$.[14] But then let

$$A' := \{\boxslash_X \chi \mid \boxslash_X \chi \in w\}$$
$$B' := \{\sim \chi \mid \sim \chi \in w\}$$

Since $\boxslash_X \boxslash_X \chi = \boxslash_X \chi$ by Lemma 2.3.6, we have $A' = \boxslash_X A$. Moreover, by Lemma 2.3.6, axiom 13, and classical logic we can see that

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

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 contradicts what has been stipulated. $\lightning$

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

Hence it must be that $\{\sim \psi\} \cup A \cup B \nvdash \perp$. In addition, from the fact that $w \subseteq \Sigma(\delta(\phi), \phi)$ with Lemma 2.3.9 and the hypothesis we have that $\{\sim \psi\} \cup A \cup B \subseteq \Sigma(\delta(\phi), \phi)$. Hence by the Lindenbaum Lemma we have that there is some $v \in At(\Sigma(\delta(\phi), \phi))$ such that $\{\sim \psi\} \cup A \cup B \subseteq v$. By the inductive hypothesis we have that $\dagger^{\phi}, v \nVdash \psi$.

To complete the argument, we have to show that $wR^{\dagger^{\phi}}_{\boxminus_X} v$. Since $A \subseteq v$ we just need to check that $\bigcup\{\{\psi, \boxtimes_X \psi\} \mid \boxtimes_X \psi \in v\} \subseteq w$. Suppose that $\boxtimes_X \psi \in v$ but $\psi \notin w$. Since w is maximally consistent we have then that $\notin \psi \in w$. Thus $\sim \boxtimes_X \psi \in v$, which 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 2.3.6 that $\boxtimes_X \boxtimes_X \psi = \boxtimes_X \psi$, which once again implies that v is inconsistent. $\lightning$ QED

Lemma 2.3.14 ($\dagger^{\phi}$ is Partly EviL). $\dagger^{\phi}$ *makes true the following properties:*

(1) $R^{\dagger^{\phi}}_{\boxminus_X} \subseteq W^{\dagger^{\phi}} \times W^{\dagger^{\phi}}$

(2) $W^{\dagger^{\phi}}$ *is finite*

(3) *For all* $w \in W^{\dagger^{\phi}}$ *we have* $wR^{\dagger^{\phi}}_{\boxminus_X} w$

(4) *If* $wR^{\dagger^{\phi}}_{\boxminus_X} v$ *and* $vR^{\dagger^{\phi}}_{\boxminus_X} z$ *then* $vR^{\dagger^{\phi}}_{\boxminus_X} z$

(5) $R^{\dagger^{\phi}}_{\boxplus_X} = (R^{\dagger^{\phi}}_{\boxminus_X})^{-1}$

(6) *If* $wR^{\dagger^{\phi}}_{\boxminus_X} v$ *then* $w \in V^{\dagger^{\phi}}(p)$ *if and only if* $v \in V^{\dagger^{\phi}}(p)$

(7) *If* $wR^{\dagger^{\phi}}_{\boxminus_X} v$ *and* $vR^{\dagger^{\phi}}_{\Box_X} u$ *then* $vR^{\dagger^{\phi}}_{\Box_X} u$

(8) *If* $wR^{\dagger^{\phi}}_{\boxminus_X} v$ *then* $uR^{\dagger^{\phi}}_{\Box_X} w$ *if and only if* $uR^{\dagger^{\phi}}_{\Box_X} v$

(9) *If* $w \in P^{\dagger^{\phi}}_X$ *then* $wR^{\dagger^{\phi}}_{\Box_X} w$

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

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

2.3.4 Bisimulation

I first introduce a Backus-Naur form grammar for the Either type constructor, which may be viewed as a coproduct in category theory (in the category of Sets)[15]:

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

Definition 2.3.15. *Let* $\mathbb{M}$ *be a Kripke model, then define* $\ddagger^{\mathbb{M}}$ *as a model*

$$\langle W^{\ddagger^{\mathbb{M}}}, V^{\ddagger^{\mathbb{M}}}, P^{\ddagger^{\mathbb{M}}}_X, R^{\ddagger^{\mathbb{M}}}_{\Box_X}, R^{\ddagger^{\mathbb{M}}}_{\boxminus_X}, R^{\ddagger^{\mathbb{M}}}_{\boxplus_X} \rangle$$

Where:

[15] Either is taken from the functional programming language `Haskell`

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

Lemma 2.3.16. *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 between $\mathbb{M}$ and $\dagger^{\mathbb{M}}$:*

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

Lemma 2.3.17. *If $\mathbb{M}$ is partly* EviL *then $\dagger^{\mathbb{M}}$ is partly* EviL *as well. It also makes true another, novel property:*

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

Any partly EviL *Kripke model that makes true this tenth property is said to be **completely* EviL**.

2.3.5 Abstract Completeness

2.3.6 Translation

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

Definition 2.3.18. *Let $\mathbb{M}$ be a partly* EviL *Kripke structure. I shall make the following definition:*

$$\ulcorner w \urcorner^{\mathbb{M}} := \{v \mid w(R_{\boxminus_X}^{\mathbb{M}} \cup R_{\boxplus_X}^{\mathbb{M}})^* v\}$$

Here R^* is the reflexive transitive closure of R.

Lemma 2.3.19 (Column Lemma)**.** *The following hold if $\mathbb{M}$ is partly* EviL*:*

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

Definition 2.3.20. *Let $L(\phi) := \{p \in \Phi \mid p$ is a subformula of $\phi\}$*
Let $\Lambda^{\mathbb{M}} := \bigcup\{\{\{w\}, \ulcorner w \urcorner^{\mathbb{M}}\} \mid w \in W^{\mathbb{M}}\}$
Let $\rho_\phi^{\mathbb{M}} : \Lambda^{\mathbb{M}} \rightarrowtail \Phi \setminus L(\phi)$ be an injection
Let $\theta_\phi^{\mathbb{M}} : W^{\mathbb{M}} \to \wp\Phi \times \wp(\mathcal{L}|_{\mathsf{Prop}}(\Phi))$ be defined such that:

$$\begin{aligned}\theta_\phi^{\mathbb{M}}(w) := (\{p \in L(\phi) \mid \mathbb{M}, w \Vdash p\} \cup \{\rho_\phi^{\mathbb{M}}(\ulcorner w \urcorner^{\mathbb{M}})\}, \lambda X.\{\neg\rho_\phi^{\mathbb{M}}(\ulcorner v \urcorner^{\mathbb{M}}) \mid \neg wR_{\Box_X}^{\mathbb{M}} v\} \\ \cup\{\bot \to \rho_\phi^{\mathbb{M}}(\{v\}) \mid wR_{\boxminus_X}^{\mathbb{M}} v\})\end{aligned}$$

Let $\star_\phi^{\mathbb{M}} := \theta_\phi^{\mathbb{M}}[W^{\mathbb{M}}]$

Lemma 2.3.21. *Let* $\mathbb{M}$ *be a completely* EviL *Kripke structure. Then for any subformula* ψ *of* ϕ *and any* $w \in W^{\mathbb{M}}$, *we have* $\mathbb{M}, w \Vdash \psi$ *if and only if* $\star_\phi^{\mathbb{M}}, \theta_\phi^{\mathbb{M}}(w) \models \psi$

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

Assume that $\mathbb{M}, w \nVdash \Box_X \psi$, then there's some $v \in W^{\mathbb{M}}$ such that $w R_{\Box_X}^{\mathbb{M}} v$ $\mathbb{M}, v \nVdash \psi$. Let $(a, A) := \theta^{\mathbb{M}}(w)$ and $(b, B) := \theta^{\mathbb{M}}(v)_\phi$. By the inductive hypothesis it suffices to show that $\star_\phi^{\mathbb{M}}, (b, B) \models A_X$. But the only things in A_X are tautologies or formulae of the form $\neg \rho_\phi^{\mathbb{M}}(\ulcorner u \urcorner^{\mathbb{M}})$ where $\neg w R_{\Box_X}^{\mathbb{M}} u$. But then Lemma 2.3.19 it can't be that $\neg \rho_\phi^{\mathbb{M}}(\ulcorner v \urcorner^{\mathbb{M}}) \in A_X$, and this suffices.

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

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

2.3.7 Completeness

Theorem 2.3.22. *If* $\nvdash \phi$ *then there is some model* $\mathfrak{M}$ *and some* $(a, A) \in \mathfrak{M}$ *such that* $\mathfrak{M}, (a, A) \not\models \phi$

Proof. Assume $\nvdash \phi$, then by Lemmas 2.3.13 and 2.3.14 we have some partly EviL Kripke structure and world such that $\mathbb{A}, a \nVdash \phi$. By Lemma 2.3.16 we have that there is a completely EviL Kripke structure $\mathbb{B}$ such that $\mathbb{A} \underline{\leftrightarrow} \mathbb{B}$, thus there is some world b such that $\mathbb{B}, b \nVdash \phi$. Finally by Lemma 2.3.21 we have that there's a model $\mathfrak{C}$ in EviL semantics and a pair $(c, C) \in \mathfrak{C}$ such that $\mathfrak{C}, (c, C) \not\models \phi$. QED

2.3.8 Conservativity, Decidability & Complexity

In this section, we discuss basic computability results for EviL. I demonstrate that all of the fragments of EviL are decidable, and establish a lower bound on the computational complexity.

I shall first prove the following lemma:

Lemma 2.3.23. EviL, EviL$^{\boxminus}$ *and* EviL$^{\boxplus}$ *with a single agent are all conservative extensions of the basic modal logic with just axiom K. That is, if* $\nvdash_K \phi$ *then* $\nvdash_{\text{EviL}} \phi$ *and similarly for the fragments* EviL$^{\boxminus}$ *and* EviL$^{\boxplus}$.

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

Proof. Assume that $\nvdash_K \phi$, then we know from modal logic that there's a finite Kripke Structure $\mathbb{M} := \langle W, V, R \rangle$ such and a world $w \in W$ such that $\mathbb{M}, w \nvdash \phi$. 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 $\mathbb{M}', w \nvdash \phi$. Hence by the Lemma 2.3.21 we have a model $\mathfrak{M}$ and $(a, A) \in \mathfrak{M}$ such that $\mathfrak{M}, (a, A) \not\models \phi$; so by soundness for EviL we have the desired result.

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

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

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. 10.

$$\begin{tikzcd}
& \text{EviL} & \\
\text{EviL}^{\boxminus} \arrow[ur] & & \text{EviL}^{\boxplus} \arrow[ul] \\
& K \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

Figure 10: EviL conservative extensions of K

Lemma 2.3.24. 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

3 Applications

3.1 Collapse

3.2 Epistemic Plurality

3.2.1 Different Kinds of Knowledge

3.2.2 Moore's Paradox

3.2.3 Fitch's Paradox

3.3 Intuitionistic Logic

3.3.1 The Gödel Tarski McKinsensy Embedding

3.3.2 Knowledge

3.3.3 Imagination

3.3.4 van Benthem S4

3.3.5 $ImK_{\Box}$

4 Epilogue

4.1 Comparison to Other Approaches

4.2 Failures

There are several points of failure of EvIL that I feel must be addressed:

(I) EvIL is not really a logic, because it is non-normal and non-compact, so it therefore any kind of reasonable algebraic duality is impossible (for details on this, see 7, chapter 5)

(II) EvIL is not dynamic and therefore fails to conform to the prevailing paradigm for epistemic logics

(III) EvIL is not completely computer verified - only the completeness theorem for the central axiom system for EvIL has been produced; none of the many auxiliary results have been verified

(IV) EvIL only partly accommodates irrationality

(V) EVIL is inhuman - the assumptions it makes for the nature of knowledge and EVIL agent's cognitive abilities are unrealistic

A Grammars

$\mathcal{L}_{\text{therm}}$	$\phi ::= x$ Pascals $\mid y$ moles $\mid z$ Kelvin $\mid \phi \to \psi \mid \bot \mid \Box\phi$	pg. 6
$\mathcal{L}_K(\Phi)$	$\phi ::= p \in \Phi \mid \phi \to \psi \mid \bot \mid \Box\phi$	pg. 8
$\mathcal{L}_0$	$\phi ::= p \in \Phi \mid \phi \to \psi \mid \bot$	pg. 9
$\mathcal{L}_A(\Phi)$	$\phi ::= p \mid \neg p \mid \top \mid \bot \mid \circlearrowleft \mid \phi \wedge \psi \mid \phi \vee \psi \mid \Diamond\phi \mid \boxminus\phi \mid \diamondplus\phi$	pg. **??**
$\mathcal{L}_B(\Phi)$	$\phi ::= \neg p \mid p \mid \bot \mid \top \mid \neg\circlearrowleft \mid \phi \vee \psi \mid \phi \wedge \psi \mid \Box\phi \mid \diamondminus\phi \mid \boxplus\phi$	pg. **??**

B Alternate Semantics

In this section, I 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 (42), chapter 2.

First, recall the basic modal grammar $\mathcal{L}_K(\Phi)$:

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

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 (6) for an elaboration on these connections.

Definition B.0.1. *Define by recursion the following two truth relations:*

First relation:

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

$$\mathfrak{M}, w \Vdash \phi \wedge \psi \iff \textit{both } \mathfrak{M}, w \Vdash \phi \textit{ and } \mathfrak{M}, w \Vdash \psi$$

$\mathfrak{M}, w \Vdash \phi \vee \psi \iff$ *either* $\mathfrak{M}, w \Vdash \phi$ *or* $\mathfrak{M}, w \Vdash \psi$

$\mathfrak{M}, w \Vdash \neg\phi \iff \mathfrak{M}, w \Vvdash \phi$

$\mathfrak{M}, w \Vdash \Box\phi \iff \beta(w) \vdash^* \phi$

Where $\vdash^*$ *is a sequent that is closed under reflection and resolution:*

$$\frac{\phi \in \Gamma}{\Gamma \vdash^* \phi} \qquad \frac{\Gamma \vdash^* \neg\phi \vee \psi \quad \Delta \vdash^* \phi}{\Gamma \cup \Delta \vdash^* \psi}$$

Second relation:

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

$\mathfrak{M}, w \Vvdash \phi \wedge \psi \iff$ *either* $\mathfrak{M}, w \Vvdash \phi$ *or* $\mathfrak{M}, w \Vvdash \psi$

$\mathfrak{M}, w \Vvdash \phi \vee \psi \iff$ *both* $\mathfrak{M}, w \Vvdash \phi$ *and* $\mathfrak{M}, w \Vvdash \psi$

$\mathfrak{M}, w \Vvdash \neg\phi \iff \mathfrak{M}, w \Vdash \phi$

$\mathfrak{M}, w \Vvdash \Box\phi \iff$ *there is some* $v \in \iota(w)$ *such that* $\mathfrak{M}, v \Vvdash \phi$

I feel it is necessary to motivate the intuition behind these semantics. Informally, I think of these two truth relations correspond to two players, whom I 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 I am trying to model; I imagine epistemic agents modeled by this system to be embroiled in internal conflict. I feel this sort of dissension between reason and imagination rages on within us all – it's fundamental to human nature.

These semantics are not naturally bivalent; that is it doesn't hold that either $\mathfrak{M}, w \Vdash \phi$ or $\mathfrak{M}, w \Vvdash \phi$, exclusively. To see this consider a model where $\beta(w) = \iota(w) = \varnothing$; then evidently $\mathfrak{M}, w \nVdash \Box p$ and $\mathfrak{M}, w \not\Vvdash \Box p$.

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* $\Vvdash$ *are bivalent if and only if* $\mathfrak{M}, w \Vdash \phi \iff \mathbb{M}^{\mathfrak{M}}, w \Vdash_\Box \phi$.

Proof. ($\Longrightarrow$) Assume that $\Vdash$ and $\Vvdash$ are bivalent and consider any $\phi \in \mathcal{L}_K(\Phi)$. The proof that $\mathfrak{M}, w \Vdash \phi$ is equivalent to $\mathbb{M}^{\mathfrak{M}}, w \Vdash_\Box \phi$ 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:

$$\begin{aligned}\mathbb{M}^{\mathfrak{M}}, w \Vdash_{\Box} \neg\phi \iff & \quad \mathbb{M}^{\mathfrak{M}}, w \nVdash_{\Box} \phi \\ \iff & \quad \mathfrak{M}, w \nVdash \phi \qquad \text{(inductive step)} \\ \iff & \quad \mathfrak{M}, w \Vvdash \phi \qquad \text{(bivalence)} \\ \iff & \quad \mathfrak{M}, w \Vdash \neg\phi\end{aligned}$$

Modality: We have another chain of equivalences:

$$\begin{aligned}\mathfrak{M}, w \Vdash \Box\phi \iff & \quad \mathfrak{M}, w \not\Vvdash \Box\phi & \text{(bivalence)} \\ \iff & \quad \forall v \in \iota(w).\mathfrak{M}, w \not\Vvdash \phi & \text{(definition)} \\ \iff & \quad \forall v \in \iota(w).\mathfrak{M}, w \Vdash \phi & \text{(bivalence)} \\ \iff & \quad \forall v \in \iota(w).\mathbb{M}^{\mathfrak{M}}, w \Vdash_{\Box} \phi & \text{(inductive step)} \\ \iff & \quad \forall v. wR^{\mathfrak{M}}v \implies \mathbb{M}^{\mathfrak{M}}, w \Vdash_{\Box} \phi & \text{(definition)} \\ \iff & \quad \mathbb{M}^{\mathfrak{M}}, w \Vdash_{\Box} \Box\phi\end{aligned}$$

This completes the induction.

($\Longleftarrow$) Assume that $\mathfrak{M}, w \Vdash \phi$ and $\mathbb{M}^{\mathfrak{M}}, w \Vdash_{\Box} \phi$ are always equivalent. We have:

$$\begin{aligned}\mathfrak{M}, w \Vdash \phi \iff & \quad \mathbb{M}^{\mathfrak{M}}, w \Vdash_{\Box} \phi \\ \iff & \quad \mathfrak{M}, w \nVdash_{\Box} \neg\phi \\ \iff & \quad \mathfrak{M}, w \nVdash \neg\phi \qquad \text{(hypothesis)} \\ \iff & \quad \mathfrak{M}, w \not\Vvdash \phi\end{aligned}$$

QED

Corollary B.0.3. *If* $\Vdash$ *and* $\Vvdash$ *are bivalent, then* $\beta(w) \vdash^* \phi$ *for all* $\phi \in \mathrm{Th}_{\Vdash}(\mathfrak{M})$ *for all* $w \in W^{\mathfrak{M}}$*, where* $\mathrm{Th}_{\Vdash}(\mathfrak{M}) = \{\phi \in \mathcal{L}_K(\Phi) \mid \mathfrak{M}, w \Vdash \phi \text{ for all } w \in W^{\mathfrak{M}}\}$*.*

Evidently bivalence of $\Vdash$ and $\Vvdash$ 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:

$$\begin{aligned}\beta^{\mathbb{M}}(w) := & \quad \{\phi \in \mathcal{L}_K(\Phi) \mid \mathbb{M}, w \Vdash_{\Box} \Box\phi\} \\ \iota^{\mathbb{M}}(w) := & \quad \{v \in W^{\mathbb{M}} \mid wR^{\mathbb{M}}v\}\end{aligned}$$

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, reflecting on my remarks in §1.2, I think that it's 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 in my opinion. However, this is evidently not entirely necessary. Call a belief/imagination model *reasonable* if the following two constraints are satisfied:

- $\beta(w) \vdash^* \phi$ for all $\phi \in \mathrm{Th}_{\Vdash}(\mathfrak{M})$ for all $w \in W^{\mathfrak{M}}$, where $\mathrm{Th}_{\Vdash}(\mathfrak{M}) = \{\phi \in \mathcal{L}_K(\Phi) \mid \mathfrak{M}, w \Vdash \phi$ for all $w \in W^{\mathfrak{M}}\}$
- $\mathrm{Mod}^{\mathfrak{M}}_{\nVdash}(\beta(w)) \subseteq \iota(w)$, where $\mathrm{Mod}^{\mathfrak{M}}_{\nVdash}(\beta(w)) = \{v \in W^{\mathfrak{M}} \mid \mathfrak{M}, v \nVdash \phi$ for all $\phi \in \beta(w)\}$
- $\beta(w) \setminus \mathrm{Th}_{\Vdash}(\mathfrak{M})$ 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 $\mathbb{M}^{\mathfrak{M}}, w \Vdash_{\Box} \phi$ then $\mathfrak{M}, w \Vdash \phi$

(ii) If $\mathbb{M}^{\mathfrak{M}}, w \nVdash_{\Box} \phi$ then $\mathfrak{M}, w \Vvdash \phi$

Hence we have $\Vdash$ and $\Vvdash$ are bivalent.

Proof. The propositional, disjunctive and conjunctive cases are all straightforward; I shall focus on negation and modality.

Negation: In the case of (i), we know that

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

The proof for (ii) is similar.

Modality: In the case of (i), assume that $\mathbb{M}^{\mathfrak{M}}, w \Vdash_{\Box} \Box\phi$. 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 $\mathfrak{M}, v \Vdash \phi$.

From this and the fact that $\mathfrak{M}$ is reasonable we can infer that $\bigvee_{\psi \in \beta(w) \setminus \mathrm{Th}(\mathfrak{M})} \neg\psi \vee \phi \in \mathrm{Th}_{\Vdash}(\mathfrak{M})$. 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 $\beta(w) \vdash^* \phi$, which just means that $\mathfrak{M}, w \Vdash \Box\phi$, 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. *$\vdash_K \phi$ if and only if $\mathfrak{M}, w \Vdash \phi$ 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 $\nvdash_K \phi$. 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 $\mathbb{M}, w \nVdash_{\Box} \phi$. Now consider a slightly modified model $\mathbb{M}' := \langle W^{\mathbb{M}}, V', R^{\mathbb{M}} \rangle$ where

$$V'(p) := \begin{cases} \{v\} & p = \rho(v) \\ V(p) & o/w \end{cases}$$

A proof by induction on subformulae ψ of ϕ verifies that $\mathbb{M}, w \Vdash_{\Box} \psi$ if and only if $\mathbb{M}', w \Vdash_{\Box} \psi$.

So now consider $\mathfrak{M} := \langle W^{\mathbb{M}'}, V^{\mathbb{M}'}, \tau, \lambda x. R^{\mathbb{M}'}[x] \rangle$ such that

$$\tau(w) := \mathrm{Th}(\mathbb{M}') \cup \left\{ \bigvee_{v \in R^{\mathbb{M}'}[w]} \rho(v) \right\},$$

where $\mathrm{Th}(\mathbb{M}') := \{\psi \in \mathcal{L}_K(\Phi) \mid \mathbb{M}', v \Vdash \psi \text{ for all } v \in W^{\mathbb{M}'}\}$. A proof by induction on ψ shows that $\mathbb{M}', w \Vdash_{\Box} \psi$, $\mathfrak{M}, w \Vdash \psi$ and $\mathfrak{M}, v \nVdash \psi$ are equivalent for all $\psi \in \mathcal{L}_K(\Phi)$. Thus we have that for all $v \in W^{\mathfrak{M}}$ that $\mathfrak{M}, v \nVdash \psi$ for all $\psi \in \mathrm{Th}(\mathbb{M}')$. Moreover, evidently $wR^{\mathbb{M}'}v$ if and only if $\mathbb{M}', v \Vdash_{\Box} \bigvee_{u \in R^{\mathbb{M}'}[w]} \rho(u)$, whence we have that $wR^{\mathbb{M}'}v$ if and only if $\mathfrak{M}, v \nVdash \chi$ for all $\chi \in \tau(w)$. With this we can employ induction and establish that $\mathbb{M}', w \Vdash_{\Box} \psi$ if and only if $\mathfrak{M}, w \models \psi$ for all $\psi \in \mathcal{L}_K(\Phi)$. Since $\mathbb{M}', w \nVdash_{\Box} \phi$, we have that $\mathfrak{M}, w \not\models \phi$. 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, I do not consider them adequate. These models are only reasonable in the sense that they indeed model agents providing nontrivial proofs for their beliefs. However, they aren't reasonable in the sense that they are simple to reckon with. So while the semantics provided in §2 requires a grammar restriction, which might be considered inelegant, I have settled on it, rather than the formulation given above, precisely because I consider this to be less manageable.

C An Application of Pure Model Theory to EviL Semantics

Recall that (*VI*), presented in Prop. 2.2.15 in §2.2.3 states:

Proposition C.0.7. *For any* EviL *model* $\mathfrak{M}$, $\mho^{\mathfrak{M}}$ *has the following property:*

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

In other words, if $(a, A) R^{\mathfrak{M}}_X (c, C)$ and $(a, A) \sqsupseteq^{\mathfrak{M}}_X (b, B)$, then $(b, B) \sqsupseteq^{\mathfrak{M}}_X (c, C)$

Recall that along with this principle, the following philosophical reading was offered:

"If the agent assumes fewer things, more things are imaginable, since it's easier for a world to be incompatible with an agent's evidence."

In fact, in light of Theorem 2.1.8, the Theorem Theorem, (VI) follows from a general model theoretic relationship. For a given Kripke structure $\mathbb{M}$, define two operators $Mod^{\mathbb{M}} : \wp\mathcal{L}(\Phi, \mathcal{A}) \to \wp(W^{\mathbb{M}})$ and $Th^{\mathbb{M}} : \wp(W^{\mathbb{M}}) \to \wp\mathcal{L}(\Phi, \mathcal{A})$

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

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

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

Hence we 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 (37, chapter 3) that we have the following two properties:

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

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

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:

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

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

References

[1] N. Agray, W. Van Der Hoek, and E. De Vink, *On BAN logics for industrial security protocols*, Lecture notes in computer science (2002), 2936.

[2] 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** (1985), no. 2, 510–530, ArticleType: primary_article / Full publication date: Jun., 1985 / Copyright 1985 Association for Symbolic Logic.

[3] S. N. Artemov, *Justification logic*, CUNY Graduate Center, New York (2007).

[4] S. N. Artemov and E. Nogina, *Introducing justification into epistemic logic*, Journal of Logic and Computation **15** (2005), no. 6, 1059.

[5] Alexandru Baltag and Sonja Smets, *Probabilistic dynamic belief revision*, Synthese **165** (2008), no. 2, 179–202.

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

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

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

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

[10] Clapeyron, *Mmoire sur la puissance motrice de la chaleur*, J. lecole polytechnique **14** (1834), 153190.

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

[12] Daniel Dennett, *The intentional stance*, 7th ed., MIT Press, Cambridge Mass, 1998.

[13] Michael Raymond DePaul, *Resurrecting old-fashioned foundationalism*, Rowman & Littlefield, 2001.

[14] Ralph Waldo Emerson, *Essays first series*, December 2008, LoC Class PS: Language and Literatures: American and Canadian literature.

[15] M. Fitting, *A logic of explicit knowledge*, Logica Yearbook (2004), 1122.

[16] ———, *The logic of proofs, semantically*, Annals of Pure and Applied Logic **132** (2005), no. 1, 125.

[17] Galle Fontaine, *Continuous fragment of the mu-Calculus*, Computer Science Logic, 2008, pp. 139–153.

[18] Jeffrey Friedl, *Mastering regular expressions*, 3rd ed. ed., O'Reilly, Sebastopol CA, 2006.

[19] Dov M. Gabbay and F. Guenthner, *Handbook of philosophical logic: Volume 6*, 2nd ed., Springer, Dordrecht [u.a.], May 2002.

[20] Joseph Y. Halpern, *Set-theoretic completeness for epistemic and conditional logic*, Annals of Mathematics and Artificial Intelligence **26** (1999), no. 1-4, 1–27.

[21] V. F Hendricks and J. Symons, *Wheres the bridge? epistemology and epistemic logic*, Philosophical Studies **128** (2006), no. 1, 137167.

[22] David Hilbert, *Neubegrndung der mathematik. erste mitteilung*, Abhandlungen aus dem Mathematischen Seminar der Universitt Hamburg **1** (1922), no. 1, 157–177.

[23] Jaakko K. Hintikka, *Knowledge and belief*, Cornell Univ. Pr., 1969.

[24] A. Hommersom, John-Jules Ch Meyer, and E. de Vink, *Toward reasoning about security protocols: A semantic approach*, Electronic Notes in Theoretical Computer Science **126** (2005), 5375.

[25] A. Hommersom, John-Jules Ch Meyer, and E. De Vink, *Update semantics of security protocols*, Synthese **142** (2004), no. 2, 229267.

[26] Barteld P. Kooi, *Probabilistic dynamic epistemic logic*, Journal of Logic, Language and Information **12** (2003), no. 4, 381–408.

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

[28] 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** (1959), 23–56, ArticleType: primary_article / Full publication date: 1959 / Copyright 1959 The Aristotelian Society.

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

[30] H. J Levesque, *A logic of implicit and explicit belief*, Proceedings of the National Conference on Artificial Intelligence, 1984, p. 198202.

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

[32] John-Jules Ch Meyer and Wiebe van der Hoek, *Epistemic logic for AI and computer science*, Cambridge University Press, 1995.

[33] Plato, *The republic*, October 1998, LoC Class PA: Language and Literatures: Classical Languages and Literature.

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

[35] W. V. Quine, *Two dogmas of empiricism*, The Philosophical Review **60** (1951), no. 1, 20–43, ArticleType: primary_article / Full publication date: Jan., 1951 / Copyright 1951 Cornell University.

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

[37] Steven Roman, *Lattices and ordered sets*, 1 ed., Springer, September 2008.

[38] Ariel Rubinstein, *Modeling bounded rationality*, MIT Press, 1998.

[39] 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.

[40] Johan van Benthem, *Reflections on epistemic logic*, Logique Anal., Nouv. Sr. **34** (1991), no. 133-134, 514.

[41] ———, *Conditional probability meets update logic*, Journal of Logic, Language and Information **12** (2003), no. 4, 409–421.

[42] ———, *Modal logic for open minds*, Center for the Study of Language and Inf, February 2010.

[43] Johan van Benthem, Jelle Gerbrandy, and Barteld Kooi, *Dynamic update with probabilities*, Studia Logica **93** (2009), no. 1, 67–96.

[44] Johan van Benthem and F. R Velzquez-Quesada, *Inference, promotion, and the dynamics of awareness*, ILLC Amsterdam. To appear in Knowledge, Rationality and Action (2009).

[45] F. R Velzquez-Quesada, *Inference and update*, Synthese **169** (2009), no. 2, 283300.

[46] John Vietch, David B. Manley, Charles S. Taylor, and Ren Descartes, *Descartes' meditations*, http://www.wright.edu/cola/descartes/, July 2005.

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