In the preamble:

Set \wp to powerset

 $Set \ominus to diamondminus$

Set \oplus to diamondplus

Set ♦ to MNSDiamond

Set ⊪ to VDash

EviL

Introduction to Epistemic Logic

- Epistemic Logic (EL) is a branch of applied modal logic
- Since its conception in Hintikka's , EL traditionally takes knowledge as a primitive notion
- The language of basic EL is the common language of modal logic:

$$\phi \quad := \quad p \in \Phi \quad | \quad \phi \to \psi \quad | \quad \bot \quad | \quad K\phi$$

framebreak

- Modern EL can be thought of as a general intentional framework for reasoning about information; that being the case it naturally finds applications beyond philosophy. Specifically, EL finds application in:
 - Computer Science in multi-agent systems and security protocol analysis
 - Economics EL is particularly suited to game theory; the notion of information states are shared between the two frameworks

The Axioms of EL

- Basic EL of one agent is generally assumed to be the modal logic S5
- The axioms of \$5 are:
 - All classical propositional tautologies
 - $K\phi \to \phi$ Axiom T Truth: If the agent knows something, then its true
 - $K\phi \to KK\phi$ Axiom 4 Positive Introspection: If the agent knows something, she knows that she knows it
 - $\neg K\phi \to K\neg K\phi$ Axiom Solution Introspection: There are no unknown unknowns (despite Donald Rumsfeld)
 - $K(\phi \to \psi) \to K\phi \to K\psi$ Axiom K Logical Omniscience: The agent knows all the logical consequences of her knowledge

framebreak

• EL is also closed under the following rules:

$$\frac{\vdash \phi \quad \vdash \phi \to \psi}{\vdash \psi} \qquad \qquad \frac{\vdash \phi}{\vdash K \sigma}$$

The Semantics of EL

- The semantics of EL are Kripke Structures with equivalence relations
- An equivalence relation \sim over a set W satisfies the following rules for all $x, y, z \in W$:
 - \sim is reflexive $-x \sim x$
 - \sim is transitive if $x \sim y$ and $y \sim z$ then $x \sim z$
 - \sim is symmetric if $x \sim y$ then $y \sim x$

framebreak

- Let $\mathbb{M} := \langle W, \sim, V \rangle$ be a Kripke model where \sim is an equivalence relation.
- The semantic truth predicate (\Vdash) of EL is defined as follows:
 - $\mathbb{M}, w \Vdash p \iff p \in V(w)$
 - $\mathbb{M}, w \Vdash \phi \to \psi \iff$ either $\mathbb{M}, w \nvDash \phi$ or $\mathbb{M}, w \Vdash \psi$
 - $\mathbb{M}, w \Vdash \bot$ never
 - $\mathbb{M}, w \Vdash K\phi \iff$ for all $v \in W$, if $w \sim v$ then $\mathbb{M}, v \Vdash \phi$
- Such a model is called an \$5 model

Thermometer in a Box

- Basic EL ascribes knowledge to inanimate objects as readily as it models agents
- To understand this, consider the case of a thermometer in a 1 m³ box:

FIXME Picture of Thermometer in a Box

Figure 1:

• If the thermometer reads 290 Kelvin, how many moles of gas are in the chamber?

Ideal Gas Law

• There is no one answer for how many moles of gas are in the chamber. Rather, the answer is governed by the *ideal gas law*:

$$PV = nRT$$

• Here P is the pressure, V is the volume, n stands for moles, R is the ideal gas constant, and T is the temperature

Thermometer Models

• We may use epistemic logic to model the beliefs of this thermometer in a box. The following minor adaptation of the basic epistemic logic grammar is suited to this purpose:

$$\phi := x \text{ Pascals} \mid y \text{ moles} \mid z \text{ Pascals} \mid \phi \rightarrow \psi \mid \bot \mid K\phi$$

framebreak

- With appropriate interpretation we can think of an EL model $\langle W, V, \sim \rangle$ for the thermometer in a box, expressed as follows (remember that the volume in constant!):
 - W is pairs (P, n) where P is some positive pressure in Pascals and
 - V is defined so:
 - $(P, n) \in V(x \text{ Pascals}) \text{ if } P = x$
 - $(P, n) \in V(y \text{ moles}) \text{ if } n = y$
 - $(P, n) \in V(z \text{ Kelvin}) \text{ if } z = \frac{P}{n \cdot R}$
 - Finally, $(P, n) \sim (P', n')$ if and only if $z = \frac{P}{n \cdot R}$

Thermometer Models (Visualization)

• We can visualize the information states of the thermometer in a box as follows:

FIXME Picture of Thermometer information states

Figure 2:

Thermometers are People Too... Sort of

- The view suggests that, under the view put forward by modern EL, objects like thermometers have mental states just like people.
- This view originates with Daniel Dennet in his book The Intentional Stance:

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.

Unreasonable

- Sadly, while the above framework of EL is a powerful system for analyzing informational aspects of knowledge, it does not do a very good job at modeling justifications
- That is to say, S5 Kripke structures, like thermometers, can convey the informational interplay of knowledge for an epistemic agent, but they cannot adequately model anything recognizable as a reason for possessing knowledge

Desiderata

• Roughly, one way model reasons in epistemic logic would be to enforce the following:

 $\mathbb{M}, w \Vdash K\phi$ if and only if the agent has some kind of proof of ϕ at w

- The purpose of *Evidentialist Logic*, or EVIL, is to provide a framework which enforces this
- Analogous approaches are taken by *Justification Logic*, developed at CUNY, and in Fernando Velazquez-Quesada's PhD thesis

The Language of EVIL

• Define the language \mathcal{L}_0 as basic propositional logic over an infinite set of proposition letters Φ :

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

ullet The language \mathcal{L}_{EviL} extends this as follows:

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

Reading the \mathcal{L}_{EVIL}

- Before jumping into the semantics, here is how to read these operations:
- $\Box \phi$ means the agent can deduce ϕ from her basic beliefs
- means that all of the agent's basic beliefs are true
- $\ominus \phi$ holds if there is a way the agent could cast some of her experience and background assumptions into doubt, and after which ϕ holds
- $\oplus \phi$ holds if there is a way that the agent could extend her beliefs (or perhaps remember something) in some way, after which ϕ holds
- The following deserves special attention:
- $\Box \phi$ means the agent can deduce ϕ from her basic beliefs
- This is in line with a *foundationalist* epistemology namely that there are some beliefs that get the privilege of being *basic* or *grounded*. Other beliefs are then deduced appropriately from these basic beliefs.
- This is not so unreasonable to help develop one's intuitions regarding this, here are some examples of Basic and Non-Basic beliefs which some people might hold

Basic	Non-Basic
I have a hand on the end of my arm	I cannot go to Albert Heijn right now
	(because I'm in a Logic Tea talk)
Peano Arithmetic is consistent	Evil is consistent
	God is not benevolent
God is benevolent	(because there is so much suffering
	in the world)

Charlotte Example

- To get a feel of how EVIL works, consider an agent named Charlotte. In this toy model, Charlotte has two basic beliefs:
 - ✓ If Abelard has tried to kill Alex, then Alex has survived
 - ✓ Abelard has tried to kill Alex
- Assume these two statements are true. Then we have that

• Moreover, logical conclusions drawn from true premises are true. In this little example, it is also the case that

"Alex has survived"

A First Stab a Definition of Knowledge

• One first idea of how to define knowledge in this system would then be:

$$K\phi := \circlearrowleft \land \Box \phi$$

- This is not an adequate analysis, however
- But suppose that Charlotte also believes the following:
 - ✓ If Abelard has tried to kill Alex, then Alex has survived
 - ✓ Abelard has tried to kill Alex
 - × Vietnam is south of Malaysia
- The last statement is false...
- Moreover, now not all of Charlotte's basic beliefs are true, so under the previous definition we have

 $\neg K$ "Alex has survived"

- This is not in line with intuition however, since it seems like Vietnam's position with relation to Malaysia is *irrelevant* to conclusions Charlotte's might draw about Alex
- If Charlotte could *cast into doubt* the idea that "Vietnam is south of Malaysia", and then carry on deducing with only relevant information, she might reasonably be expected to have Knowledge
- A refined definition of knowledge that accommodates the above observation is:

$$K\phi := \ominus(\circlearrowleft \land \Box \phi)$$

Logical Omniscience

• This definition of knowledge has the feature that the following is *invalid* (known as logical omniscience):

$$\frac{K\phi - K(\phi \to \psi)}{K\psi} \times$$

Example

A – It's the harvest season for cranberries

B – If it's the harvest season for cranberries, there is a risk of bear attacks

C – A study in 2008 showed that black bear attacks in cranberry bogs in New England have been in steady decline

Pretend that a field biologist holds the beliefs above.

She knows the first statement given her background in botony, and the second statement given her background in biology

framebreak

 $\sqrt{}$ It's is the harvest season for cranberries

✓ – If it's the harvest season for cranberries, there is a risk of bear attacks

 \times – A study in 2008 showed that black bear attacks in cranberry bogs in New England have been in steady decline

Now assume that while the first two statements are true, the study is in fact erroneous – perhaps not enough evidence was gathered

framebreak

A – It's the harvest season for cranberries

B – If it's the harvest season for cranberries, bears will be attracted, hence there is a risk of bear attacks

C – A study in 2008 showed that black bear attacks in cranberry bogs in New England have been in steady decline

And moreover, imagine that in the case of this biologist, she cannot think about bear attacks in cranberry bogs in New England without appealing to the above 2008 study

That is just to say, when thinking about bears in New England, she cannot cast the 2008 study into doubt and just use her other beliefs as a biologist

framebreak

- The biologist will not have knowledge of the risk of bear attacks in this scenario, only inconsistent ideas on the subject
- Despite the fact that she knows that it is harvest season for cranberries and if it is harvest season for cranberries, there's a risk of bear attacks, she cannot formulate a sound argument for there being a risk of bear attacks

EVIL Semantics

- EVIL models are sets $\mathfrak{M} \subseteq \wp \Phi \times \wp \mathcal{L}_0$; that is to say they are sets of pairs (a, A) where:
 - \bullet a is a set of proposition letters
 - A is a set of proposition formulae the agent's basic beliefs

framebreak

- The Evil truth predicate \parallel is defined recursively as follows:
 - $\mathfrak{M}, (a, A) \Vdash p \iff p \in a$
 - $\mathfrak{M}, (a, A) \Vdash \phi \to \psi \iff$ either $\mathfrak{M}, (a, A) \not\Vdash \phi$ or $\mathfrak{M}, (a, A) \Vdash \psi$
 - $\mathfrak{M}, (a, A) \Vdash \bot$ never
 - $\mathfrak{M}, (a, A) \Vdash \Box \phi \iff \forall (b, B) \in \mathfrak{M}.$ if $\mathfrak{M}, (b, B) \Vdash \alpha$ for all $\alpha \in A$, then $\mathfrak{M}, (b, B) \Vdash \phi$
 - $\mathfrak{M}, (a, A) \Vdash \exists \phi \Longleftrightarrow \forall (b, B) \in \mathfrak{M}. \text{ if } a = b \text{ and } A \supseteq B, \text{ then } \mathfrak{M}, (b, B) \Vdash \phi$
 - $\mathfrak{M}, (a, A) \Vdash \boxplus \phi \iff \forall (b, B) \in \mathfrak{M}.$ if a = b and $A \subseteq B$, then $\mathfrak{M}, (b, B) \Vdash \phi$
 - $\mathfrak{M}, (a, A) \Vdash \circlearrowleft \iff \mathfrak{M}, (a, A) \Vdash \alpha \text{ for all } \alpha \in A$

Theorem Theorem

$$\mathfrak{M}, (a, A) \Vdash \Box \phi \iff \operatorname{Th}(\mathfrak{M}) \cup A \vdash_{\text{EVIL}} \phi$$

This illustrates that the semantics of EviL reflect the intuitions previously presented

Conservative Extension

Let K denote basic modal logic. EviL is a conservative extension of K; that is, for any modal formula ϕ

$$\vdash_K \phi \iff \vdash_{\text{EVIL}} \phi$$

But K is not the same as EVIL, since in modal logic you do not have \boxplus, \boxminus and \circlearrowleft

EVIL Is Not Compact

Suppose the set of proposition letters Φ is infinite. Consider the infinite set of formulae, $\tau[\Phi]$, where $\tau: \Phi \to \mathcal{L}_{Evil}$ is:

$$\tau(p) := p \land \Box p \land \Diamond \top \land \Box \Box \bot$$

Every finite subset of $\tau[\Phi]$ is satisfiable in EVIL, but not the entirity

Hence, Evil is not compact

EVIL Axioms

EVIL is not normal; it is not closed under replacement of proposition letters with arbitrary formulae. Hence the axioms of EVIL are schematic:

(*) indicates a non-normal axiom

EVIL is closed under modus ponens, and Necessitation for \Box , \Box and \Box

EVIL Soundness and Weak Completeness

For all formulae ϕ in \mathcal{L}_{EVIL}

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

Moreover, EviL has the finite model property (where all of the basic beliefs sets are finite too) Complexity

EVIL is decidable; specifically, we have the following bounds on its complexity:

$$PSPACE \subseteq EviL \subseteq EXP2$$

Elimination Theorem

One way to read $\Box \phi$ as "the agent believes ϕ " and $\Diamond \phi$ as "the agent can imagine ϕ " It is well known that they are each other's *dual*, so they exhibit quite a bit of symmetry Since \boxplus is associated with going up in basic beliefs, it may be thought of as sort of symmetrical

to \boxminus , which is associated with going down in basic beliefs

framebreak

This symmetry is exhibited in the following validities:

$$\begin{array}{lll} \vdash_{\mathrm{EviL}} \boxminus p \leftrightarrow p & \vdash_{\mathrm{EviL}} \boxminus p \leftrightarrow p \\ \vdash_{\mathrm{EviL}} \boxminus \neg p \leftrightarrow \neg p & \vdash_{\mathrm{EviL}} \boxminus \neg p \leftrightarrow \neg p \\ \vdash_{\mathrm{EviL}} \boxminus \Diamond \phi \leftrightarrow \Diamond \phi & \vdash_{\mathrm{EviL}} \boxminus \Box \phi \leftrightarrow \Box \phi \\ \vdash_{\mathrm{EviL}} \boxminus \Diamond \phi \leftrightarrow \Diamond \phi & \vdash_{\mathrm{EviL}} \boxminus \Box \phi \leftrightarrow \Box \phi \\ \vdash_{\mathrm{EviL}} \boxminus \Box \phi \leftrightarrow \Box \phi & \vdash_{\mathrm{EviL}} \boxminus \Box \phi \leftrightarrow \Box \phi \\ \vdash_{\mathrm{EviL}} \boxminus \Box \phi \leftrightarrow \ominus \phi & \vdash_{\mathrm{EviL}} \boxminus \Box \phi \leftrightarrow \ominus \phi \\ \vdash_{\mathrm{EviL}} \boxminus \circlearrowleft \leftrightarrow \circlearrowleft & \vdash_{\mathrm{EviL}} \boxminus \neg \circlearrowleft \leftrightarrow \neg \circlearrowleft \\ \\ \vdash_{\mathrm{EviL}} \boxminus \circlearrowleft \leftrightarrow \circlearrowleft & \vdash_{\mathrm{EviL}} \boxminus \neg \circlearrowleft \leftrightarrow \neg \circlearrowleft \\ \end{array}$$

Note that all of these validities are reductions – the formula on left of the biimplication is always more complex than the formula on the right

framebreak

Now consider the following two dual fragments of $\mathcal{L}_{\text{EVIL}}$:

$$\phi := p \mid \neg p \mid \top \mid \bot \mid \circlearrowleft \mid \phi \lor \psi \mid \phi \land \psi \mid \diamondsuit \phi \mid \boxminus \phi \mid \oplus \phi
\phi := \neg p \mid p \mid \bot \mid \top \mid \neg \circlearrowleft \mid \phi \land \psi \mid \phi \lor \psi \mid \Box \phi \mid \ominus \phi \mid \boxminus \phi$$

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

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

The observed reductions give rise to *elimination theorems* on these fragments framebreak

Specifically, recursively define an elimination operation $(\cdot)^*$ such that

$$\begin{array}{lll} p^* := p & (\neg p)^* := \neg p \\ \top^* := \top & \bot^* := \bot \\ \circlearrowleft^* := \circlearrowleft & (\neg \circlearrowleft)^* := \neg \circlearrowleft \\ (\phi \lor \psi)^* := \phi^* \lor \psi^* & (\phi \land \psi)^* := \phi^* \land \psi^* \\ (\diamondsuit \phi)^* := \diamondsuit (\phi^*) & (\Box \phi)^* := \Box (\phi^*) \\ (\boxminus \phi)^* := \phi^* & (\boxminus \phi)^* := \phi^* \\ (\oplus \phi)^* := \phi^* & (\boxminus \phi)^* := \phi^* \end{array}$$

framebreak

For $\phi \in \mathcal{L}_A \cup \mathcal{L}_B$ we have

$$\vdash_{\text{EVIL}} \phi \leftrightarrow \phi^*$$

This is a consequence of the previous observed reductions, along with duality

framebreak

This gives rise to the following EVIL validity

(FIXME: Insert very silly picture)

EVIL Intuitionism

In Reason, Truth and History (1981), Hilary Putnam writes:

To claim a statement is true is to claim that it could be justified (pg. 56)

This intuition presents two things:

- 1. It suggests an anti-realist, constructivist perspective on truth
- 2. Reading $\Box p$ as "p can be justified" is natural in EviL. This suggests that truth conditions in models for intuitionistic logic can be translated into deductions an EviL agent might perform

framebreak

Recall the grammar of intuitionistic logic:

$$\phi := p \mid \bot \mid \phi \lor \psi \mid \phi \land \psi \mid \phi \to \psi \tag{\mathcal{L}_{Int}}$$

Define $(\cdot)^{\text{EviL}}: \mathcal{L}_{\text{Int}} \to \mathcal{L}_{\text{EviL}}$ to be a variation on the Gödel Tarski McKinsey embedding:

$$p^{\text{Evil}} := \Box p \qquad \qquad \bot^{\text{Evil}} := \bot \\ (\phi \lor \psi)^{\text{Evil}} := (\phi^{\text{Evil}} \lor \psi^{\text{Evil}}) \qquad \qquad (\phi \land \psi)^{\text{Evil}} := (\phi^{\text{Evil}} \land \psi^{\text{Evil}}) \\ (\phi \to \psi)^{\text{Evil}} := \boxplus (\phi^{\text{Evil}} \to \psi^{\text{Evil}})$$

framebreak

We have the following:

$$\vdash_{\text{Int}} \phi \iff \vdash_{\text{EVIL}} \phi^{\text{EVIL}}$$

Intuitionist Logic as EVIL Epistemic Logic

The above embedding involves equating truth conditions in Intuitionistic Kripke Structures with beliefs in EviL models

It may be modified to $(\cdot)^{\text{EviiL}}:\mathcal{L}_{\text{Int}} \to \mathcal{L}_{\text{EviL}}$

$$\begin{array}{ll} p^{\text{Eviil}} := Kp & \bot^{\text{Eviil}} := \bot \\ (\phi \lor \psi)^{\text{Eviil}} := (\phi^{\text{Eviil}} \lor \psi^{\text{Eviil}}) & (\phi \land \psi)^{\text{Eviil}} := (\phi^{\text{Eviil}} \land \psi^{\text{Eviil}}) \\ (\phi \to \psi)^{\text{Eviil}} := \boxplus (\phi^{\text{Eviil}} \to \psi^{\text{Eviil}}) & \end{array}$$

Where $K\phi:=\ominus(\circlearrowleft\wedge\Box\phi),$ the formulation of knowledge previously suggested framebreak

One again, we have:

$$\vdash_{\text{Int}} \phi \iff \vdash_{\text{EVIL}} \phi^{\text{EVIIL}}$$

Intuitionistic Logic as the Logic of Imagination

One way to read $\Box \phi$ is "can deduce ϕ ", and its dual $\Diamond \phi$ as "can imagine ϕ "

Because of the symmetry of \boxplus and \boxminus , and their interplay with \square and \diamondsuit as previously illustrated, we have yet another embedding $(\cdot)^{\text{EVIIIL}}: \mathcal{L}_{\text{Int}} \to \mathcal{L}_{\text{EVIL}}$ of intuitionistic logic into EVIL:

$$\begin{array}{ll} p^{\text{EVIIIL}} := \Diamond p & \bot^{\text{EVIIIL}} := \bot \\ (\phi \lor \psi)^{\text{EVIIIL}} := (\phi^{\text{EVIIIL}} \lor \psi^{\text{EVIIIL}}) & (\phi \land \psi)^{\text{EVIIIL}} := (\phi^{\text{EVIIIL}} \land \psi^{\text{EVIIIL}}) \\ (\phi \to \psi)^{\text{EVIIIL}} := \boxminus (\phi^{\text{EVIIIL}} \to \psi^{\text{EVIIIL}}) & \end{array}$$

framebreak

As before, we have:

$$\vdash_{\text{Int}} \phi \iff \vdash_{\text{Evil}} \phi^{\text{EviiiL}}$$

Modal Intuitionistic Logic

Furthermore, this embedding can be extended to the modal intuitionistic logic ${\rm Im} K_\square$:

Furthermore, this embedding can be extended to the modal intuitionistic logic ImK_{\pi}:
$$\phi := p \mid \bot \mid \phi \lor \psi \mid \phi \land \psi \mid \phi \rightarrow \psi \mid \Box \phi \qquad (\mathcal{L}_{\text{ImK}_{\square}})$$
This gives rise to $(\cdot)^{\text{EixL}} : \mathcal{L}_{\text{Int}} \rightarrow \mathcal{L}_{\text{EviL}}$

$$\begin{array}{ll} p^{\text{EixL}} := \Box p & \bot^{\text{EixL}} := \bot \\ (\phi \lor \psi)^{\text{EixL}} := (\phi^{\text{EixL}} \lor \psi^{\text{EixL}}) & (\phi \land \psi)^{\text{EixL}} := (\phi^{\text{EixL}} \land \psi^{\text{EixL}}) \\ (\phi \to \psi)^{\text{EixL}} := \boxplus (\phi^{\text{EixL}} \to \psi^{\text{EixL}}) & (\Box \phi)^{\text{EixL}} := \Box \phi^{\text{EixL}} \end{array}$$

Not surprisingly, we have:

$$\vdash_{\operatorname{ImK}_{\square}} \phi \iff \vdash_{\operatorname{EviL}} \phi^{\operatorname{EixL}}$$

Conclusion

EVIL admittedly presents a somewhat different perspective on knowledge than traditional epistemic logic

Regardless, the tools of EVIL and EL are the same; they are after all both modal logics It is my sincerest hope that EVIL will help the field to progress into new areas of investigation