Deontic Logic

Jennifer Carr

PIKSI Logic 2018

A deductive system for modal logics

A DEDUCTIVE SYSTEM FOR MODAL LOGICS

Basics: we'll use an axiomatic derivation system:

- Every line in a derivation is meant to follow from a set of axioms. If the logic is a sound one, every line will be a logical truth.
  - In other kinds of proof systems, this isn't the case:

1	$A \wedge B$	
2	A	$\wedge$elim, 1

- A starting set of axioms:

K $\quad \Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$

DUAL $\quad \Diamond A \leftrightarrow \neg\Box\neg A$

TAUT $\quad$ All tautologies $\qquad (A \wedge B) \rightarrow A$

A DEDUCTIVE SYSTEM FOR MODAL LOGICS

These are the axioms for the most basic modal logic, which is also called **K**.

K $\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$

- *K* lets us distribute the '$\Box$' over the conditional.

DUAL $\Diamond A \leftrightarrow \neg\Box\neg A$

- DUAL simply gives the definition of '$\Diamond$'

TAUT All tautologies are axioms of the system.

- A tautology is any sentence guaranteed to be true by its truth table, i.e. by its truth-functional form.

We can add further axioms to make more interesting modal logics, including deontic logic.

Our system has only two inference rules!

1. **MP** (modus ponens): if you have A and you have $A \rightarrow B$, you can infer $\mathcal{B}$.
2. **NEC** (necessitation): if you have A, you can infer $\Box A$.

it's raining or it's not raining

$\Box$ (it's raining or not)

A SIMPLE EXAMPLE

Let's derive a theorem of **K**:

$$\mathbf{K} \vdash \Box A \to \Box(B \to A)$$

1. $A \to (B \to A)$ TAUT
2. $\Box(A \to (B \to A))$ NEC, 1
3. $\Box(A \to (B \to A)) \to (\Box A \to \Box(B \to A))$ K
4. $\Box A \to \Box(B \to A)$ MP 2,3

ANOTHER EXAMPLE

With just two inference rules, derivations are sometimes unwieldy. Consider, e.g., $(\Box A \wedge \Box B) \to \Box(A \wedge B)$.

1. $A \to (B \to (A \wedge B))$ TAUT
2. $\Box(A \to (B \to (A \wedge B)))$ NEC, 1
3. $\Box(A \to (B \to (A \wedge B))) \to (\Box A \to \Box(B \to (A \wedge B)))$ K
4. $\Box A \to \Box(B \to (A \wedge B))$ MP 2, 3
5. $\Box(B \to (A \wedge B)) \to (\Box B \to \Box(A \wedge B))$ K
6. $(\Box A \to \Box(B \to (A \wedge B))) \to (\Box(B \to (A \wedge B)) \to (\Box B \to \Box(A \wedge B))) \to (\Box A \to (\Box B \to \Box(A \wedge B))))$ TAUT
7. $(\Box(B \to (A \wedge B)) \to (\Box B \to \Box(A \wedge B))) \to (\Box A \to (\Box B \to \Box(A \wedge B)))$ MP 4, 6
8. $\Box A \to (\Box B \to \Box(A \wedge B)))$ MP 5, 7
9. $(\Box A \to (\Box B \to \Box(A \wedge B)))) \to ((\Box A \wedge \Box B) \to \Box(A \wedge B))$ TAUT
10. $(\Box A \wedge \Box B) \to \Box(A \wedge B)$ MP 8, 9

SOME DERIVED RULES

We can make these proofs shorter and easier with some **derived rules**.

Example: notice that in the previous proofs, we had to introduce tautological conditionals to get new lines that follow from previous lines by propositional logic.

For example, to get from lines of the form '$A \rightarrow B$' and '$B \rightarrow C$' to '$A \rightarrow C$', we have to introduce the tautology:

$$(A \rightarrow B) \rightarrow ((B \rightarrow C) \rightarrow (A \rightarrow C))$$

$A \rightarrow B$
$B \rightarrow C$

$A \rightarrow C$

...and then apply MP twice.

Instead, we can introduce a derived rule that allows us to skip these steps.

SOME DERIVED RULES

PL

If $\mathbf{K} \vdash A_1, \ldots, \mathbf{K} \vdash A_n$, and B follows from $A_1, \ldots, A_n$ by propositional logic, then $\mathbf{K} \vdash B$.

Another convenient derived rule: It's often convenient to go directly from $\mathbf{K} \vdash A \to B$ to $\mathbf{K} \vdash \Box A \to \Box B$, without having to add an instance of K and then apply MP. More generally:

RK

If $K \vdash A_1 \to (A_2 \to \ldots (A_{n-1} \to A_n) \ldots)$ then $K \vdash \Box A_1 \to (\Box A_2 \to \ldots (\Box A_{n-1} \to \Box A_n) \ldots)$.

$A \to (B \to C)$

$\Box(A \to (\Box B \to \Box C))$

NOT

$\vdash A \vee \neg A$

$\vdash \Box A \vee \Box \neg A$

SOME DERIVED RULES

Substitution

Substitution: if our logic entails that two sentences are logically equivalent, then you can substitute one in for the other, even when they occur in the context of a larger sentence.

For example: $\neg\Diamond A \leftrightarrow \Box\neg A$

$$\mathbf{K} \vdash A \leftrightarrow \neg\neg A$$

$\Diamond A \leftrightarrow \neg\Box\neg A$

$$\mathbf{K} \vdash \mathbf{P}A \leftrightarrow \neg\mathbf{O}\neg A$$

So, the following inferences are permissible:

1. $\neg\neg\mathbf{O}\neg A \leftrightarrow \neg\neg\mathbf{O}\neg A$
2. $\mathbf{O}\neg A \leftrightarrow \neg\neg\mathbf{O}\neg A$ — 1, sub. $\neg\neg p$ for p
3. $\mathbf{O}\neg A \leftrightarrow \neg\mathbf{P}A$ — 2, sub. $\neg\mathbf{O}\neg p$ for $\mathbf{P}p$

EXAMPLES

$\mathbf{K} \vdash (\Box A \wedge \Box B) \rightarrow \Box(A \wedge B)$ again!

FROM K TO STANDARD DEONTIC LOGIC

Standard deontic logic (SDL) is built from **K**.

- Instead of '$\Box$', we use '**O**' for *obligatory*
- Instead of '$\Diamond$', we use '**P**' for *permissible*

SDL includes an extra axiom:

D $\mathbf{O}A \rightarrow \mathbf{P}A$

SOME THEOREMS OF SDL

SDL = KD

$$\mathbf{O}\top$$

$$\neg\mathbf{O}\bot$$

$$(\mathbf{O}A \wedge \mathbf{O}B) \rightarrow \mathbf{O}(A \wedge B)$$

$$\mathbf{O}(A \wedge B) \rightarrow (\mathbf{O}A \wedge \mathbf{O}B)$$

$$\neg(\mathbf{O}A \wedge \mathbf{O}\neg A)$$

NO DILEMMAS

Let's prove $\neg(\mathbf{O}A \wedge \mathbf{O}\neg A)$. (This means: no logically conflicting obligations or dilemmas.)

Problems with NEC for deontic modals

WORRIES ABOUT NECESSITATION

NEC has the immediate result that all tautologies are obligatory.

1.	$\top$	TAUT
2.	$\mathbf{O}\top$	NEC

Not particularly intuitive, but is this a problem?

- You have infinitely many tautologous obligations...
- ...But they're incredibly easy to fulfill (indeed, impossible not to fulfill)!
- Without this, we'd have to start from scratch with deontic logic.

WORRIES ABOUT NECESSITATION

Some potentially more serious worries:

- Imagine a possible world in which there are no intelligent creatures of any kind—perhaps a possible world where life never emerges. In such a world, it seems true to say: "There are no obligations" or "Nothing is obligatory." So shouldn't these sentences be contingent, rather than false as a matter of logical necessity?
- Similarly: it seems like there could be a person of whom it's true to say: "Jim never satisfies any of his obligations." So this sentence ought to be contingent. But Jim either steals or doesn't steal, and by NEC, it's obligatory that Jim either steals or doesn't steal.

More Paradoxes Involving Conditional Obligation

THE GENTLE MURDERER

The **Paradox of the Gentle Murderer** comes from Forrester (1984).

It seems like the following sentences could all be true together:

1. Hannibal Lecter is obligated not to murder his prison guard.
2. Hannibal Lecter is going to murder his prison guard.
3. If Hannibal Lecter is going to murder his prison guard, he is obligated to murder him gently.
4. As a matter of logical necessity, if Hannibal Lecter murders his prison guard gently, then he murders his prison guard.

THE GENTLE MURDERER

1. Hannibal Lecter is obligated not to murder his prison guard.
2. Hannibal Lecter is going to murder his prison guard.
3. If Hannibal Lecter is going to murder his prison guard, he is obligated to murder him gently.
4. As a matter of logical necessity, if Hannibal Lecter murders his prison guard gently, then he murders his prison guard.

1. $\mathbf{O}\neg M$
2. M
3. $M \rightarrow \mathbf{O}G$
4. $\vdash G \rightarrow M$

- From 2. and 3. we can can infer...? OG (MP)
- From 4., we can infer ...? OG → OM (RK)
- Given these two inferences, we get ...? OM (MP)

Is this just a problem for the idea that $G \to M$ entails $\mathbf{O}G \to \mathbf{O}M$—that is, a problem for either **K** or NEC?

Arguably not! Intuitively, we could skip that step and present a simplified version of the puzzle:

1.* Hannibal Lecter is not obligated to murder his prison guard, gently or not.[1]
2. Hannibal Lecter is going to murder his prison guard.
3. If Hannibal Lecter is going to murder his prison guard, he is obligated to murder him gently.

But if these three are all true, then this looks like a problem for Modus Ponens. (!)

[1] Indeed, he's obligated *not* to murder his prison guard gently!

THE GENTLE MURDERER

In response, many have argued that we shouldn't translate 3. into symbolic logic as 3a., but rather as 3b.:

3. If Hannibal Lecter is going to murder his prison guard, he is obligated to murder him gently.

3a. $M \rightarrow \mathbf{O}G$ narrow scope **O**

3b. $\mathbf{O}(M \rightarrow G)$ wide scope **O**

Then, at least in symbolic logic, MP isn't applicable, so the paradox seems to dissolve:

1. M
2. $\mathbf{O}(M \rightarrow G)$
3. $\mathbf{O}\neg M$

THE SAMARITAN PARADOX

The **Samaritan Paradox** (Prior 1958, Åqvist 1967):

R ⊢ P

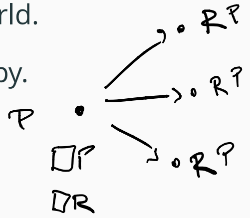
1. You ought to rescue the abandoned, starving puppy.

According to SDL, 1. is true iff 2. is true in all the deontically ideal worlds that are accessible from your world.

2. You rescue the abandoned, starving puppy.

But 2. logically entails 3.

3. There is an abandoned, starving puppy.



So 3. is true in all deontically ideal worlds accessible from your world. So 4. is true at your world. (!)

4. There ought to be an abandoned, starving puppy.

KRATZER'S SAMARITAN PARADOX

Another version of the Samaritan paradox, targeted at conditionals.

1. There ought not be any abandoned, starving puppies. $\mathbf{O}\neg P$
2. If there is an abandoned, starving puppy, someone ought to rescue it. $\mathbf{O}(P \rightarrow R)$

The problem: in SDL, 2. is a trivial consequence of 1. If 1 is true, then so is 2 but—but also 3:

3. If there is an abandoned, starving puppy, you should punch it in the face. $\mathbf{O}(P \rightarrow F)$

Even when 1. is true, 2. should be substantive and 3. shouldn't trivially follow from 1. Indeed, even when 1. is true, 3. should be false.

Logic and Natural Language Semantics

FROM SDL TO NATURAL LANGUAGE SEMANTICS

Natural language semantics for deontic language—for example, *obligatory, permissible, impermissible, ought, should, must, may, can, good, bad, right, wrong*—uses many of the resources of standard deontic logic, but aims to avoid some of its more counterintuitive consequences.

The best-known and most influential natural language semantics for deontic language is Angelika Kratzer's (1977, 1981, 1991) semantics for modals and conditionals: for short, Kratzer semantics.

We'll discuss a simplified version of Kratzer semantics.

WHAT'S WRONG WITH SDL

Many of the paradoxes of SDL have something in common: they suppose that there are some accessible deontically ideal worlds, but then focus on what's next best on the assumption that we won't manage to actualize those ideal worlds.

- Ideally, Lecter won't murder. Next best: he murders gently.
- Ideally, there are no abandoned, starving puppies. Next best: the abandoned, starving puppies get rescued.

One problem with SDL: it only considers what goes on in accessible deontically ideal worlds, ignoring next best worlds.

Can we solve this by saying that, in these cases, only next-best worlds are accessible?

No: then "Lecter shouldn't murder" and "There shouldn't be abandoned, starving puppies" would be false!

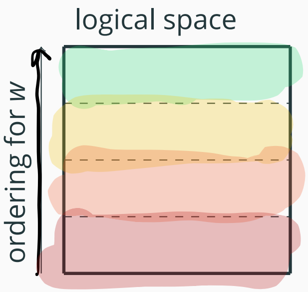
SIMPLIFIED KRATZER SEMANTICS

Kratzer agrees that **O** and **P** quantify over a domain of best accessible worlds.

But these are determined by two parameters, not merely an accessibility relation.

- **Modal base**: we can think of this as determining which worlds are circumstantially accessible: what can be the case, given the current circumstances.
  (This isn't a deontic notion. It's related to the idea of "can" in "'ought' implies 'can'".)
- **Ordering**: a ranking of worlds in terms of deontic ideality.

ORDERING



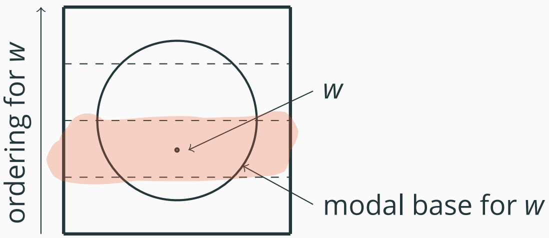
Suppose, e.g., that the highest worlds in the ordering are ones where there are no abandoned puppies.

Second best: worlds where there is an abandoned puppy, but it gets rescued.

Third best: worlds where there is an abandoned puppy and it doesn't get rescued, but also doesn't get punched.

Fourth best: worlds where there is an abandoned, unrescued puppy and you punch it in the face.

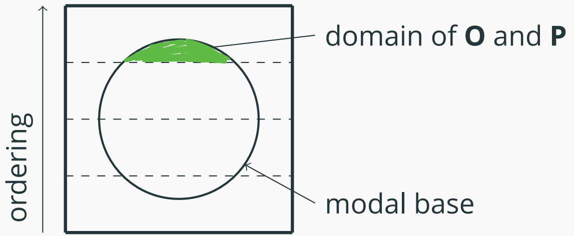
MODAL BASE



The modal base represents which worlds are possible at *w*, given the circumstances.

In natural language, this will be *context-sensitive*: perhaps determined by what the speaker is temporarily assuming is an immutable fact about the world.

THE DOMAIN FOR O AND P

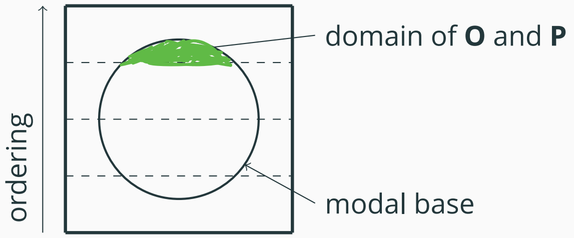


The key to Kratzer semantics: **O** and **P** don't quantify over the set of deontically ideal worlds, or the set of accessible worlds.

Instead, they quantify over the highest ranked accessible worlds.

So, e.g., if we assume that you cannot solve world hunger, then we don't include worlds where you solve world hunger in the modal base. So "You should solve world hunger" isn't true, even though worlds where where you solve world hunger are very highly ranked.

SIMPLIFIED KRATZER SEMANTICS



Kratzer semantics for modals:

"**O**A" is true at *w* iff *A* is true at all the highest ranked worlds in the modal base at *w*.

"**P**A" is true at *w* iff *A* is true at at least one highest ranked worlds in the modal base at *w*.

KRATZER SEMANTICS FOR CONDITIONALS

To address puzzles like Gentle Murderer and Kratzer's Samaritan Paradox, this isn't the whole story: we also need Kratzer's semantics for deontic conditionals.

We saw that there are problems with interpreting "if *A*, ought *B*" as either:

1. $A \rightarrow \mathbf{O}B$
2. $\mathbf{O}(A \rightarrow B)$

What other options are there?

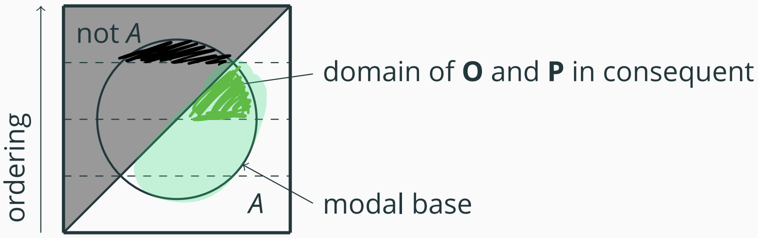
Kratzer: we shouldn't think of "if *A*, ought *B*" as representable with the material conditional $\rightarrow$.

KRATZER SEMANTICS FOR CONDITIONALS

On Kratzer's view, conditionals are a kind of modal claim.

The antecedent of a conditional has the job of restricting the modal base. (Her semantics is a "restrictor analysis of conditionals".)

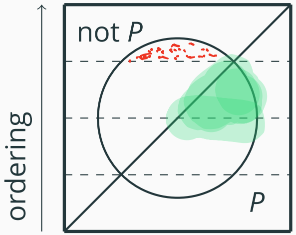
"If *A*, ought *B*" is true iff *B* is true in all highest ranked *A*-worlds in the modal base.



KRATZER'S SAMARITAN PARADOX

P R

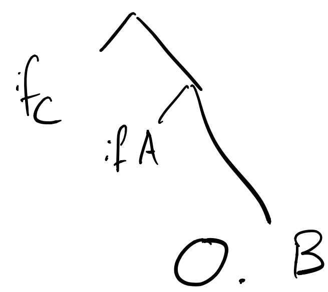
1. There ought not be any abandoned, starving puppies.
2. If there is an abandoned, starving puppy, someone ought to rescue it.
3. If there is an abandoned, starving puppy, you should punch it in the face.



In all top-ranked worlds in modal base, $\neg P$. So 1. is true.
In all top-ranked *P*-worlds in modal base, someone rescues. So 2. is true.
It's not the case that in all top-ranked *P*-worlds in the modal base, you punch. So 3. is false.

IF THERE'S TIME...

1. What about the gentle murderer?
2. What about the original Samaritan paradox?



$A \to (B \to C)$

$\equiv$

$(A \wedge B) \to C$

$O_A \, B$

$O_{A \wedge C} \, B$

$O(A, B)$

DEONTIC DETACHMENT

$O(\text{help})$

$O(\text{help} \to \text{tell})$] → $O(\text{tell})$

⊥

→ ~~~help~~

$\sim\text{help} \to O(\sim\text{tell})$] → $O(\neg\text{tell})$

FACTUAL DETACHMENT

h	+t
h	-t
-h	-t
-h	+t

-h	-t
-h	+t

Malte Willer Phil Imprint

Cat Saint-Croix & Rich Thomason

Proceedings of DEON 2014

Journal of Logic & Computation