Philosophy 244, #2: Modal Syntax and Semantics

The great advantage of ordinary propositional logic is that all of its connectives are truth-functional. It was this truth-functionality that let us abstract away from propositions and define validity in terms of arbitrary truth-value assignments as opposed to arbitrary propositional assignments. It's the truth-functionality that makes ordinary propositional logic "extensional."

Beyond Extensionality

When we get to propositional *modal* logic all that changes, because modal operators are not truth-functional. Whether it is necessary that *p* depends on a lot more than *p*'s truth-value; it depends on what *p* means, what proposition it expresses. If *p* expresses the proposition that all planets are *under a billion miles* across, then *p* is true and $\Box p$ is false; why? If it expresses the proposition that all planets are *over a billionth of a mile* across, then *p* is true and $\Box p$ is (arguably) true too. Same truth-value for *p*, different truth-values for $\Box p$.

But although $\Box$ is not truth-functional, it is *intensional*; $\Box p$'s truth-value is a function of *p*'s intension. By *p*'s intension we mean something like its truth-value across all worlds, not just one. We can think of this truth-profile, if we like, as the proposition *p* expresses, and this is common around these parts Equally we could see the truth-profile as the characteristic function—characteristic functions map members of a set to 1 and non-members to 0—of a set of worlds, and let the proposition be a set of worlds.

Does this mean that the semantics of modality is going to require an ontology of propositions, those "creatures of darkness"? One of the great accomplishments of modern modal logic is to show that propositions aren't required; we can still get by with truth values. The only difference is that now an evaluation of the language is going to involve, not just a single assignment of truth-values to atomic sentences, but a whole bunch of these assignments: one for each "possible world." Intensionality in the metalanguage says that α are β are intersubstitutable in modal contexts ...X... if they have the same truth-values in the same worlds. We could even try formulate intensionality in the object language: $\Box(\alpha \equiv \beta) \supset (....\alpha... \equiv ... \beta))$. This is getting ahead of ourselves but it's good to keep it in the back of your head.

Actually it's oversimplifying to say modal operators are intensional; some are hyperintensional. Take permission: It's permitted in Eden to eat infinitely many apples apart from *this* one (from the Tree of Knowledge of Good and Evil) but not to eat infinitely many apples full stop—yet eating infinitely apples is necessarily equivalent to eating infinitely many not identical to this one. Operators that are not even intensional are hyperintensional. For *our* purposes modal operators are intensional.

Syntax of Propositional Modal Logic (PML)

The syntax of propositional modal logic is easy; it is a slight variant of the syntax for propositional logic. The atomic letters are the same. The rules for parentheses is the same. The connectives too are the same, *except* that we've got one new unary connective $\Box$, which behaves syntactically like $\neg$. The formation rules are the same, except for this one new operator:

FR1 Every letter is an (atomic) wff.
FR2 If α is a wff, so are $\neg\alpha$ and $\Box\alpha$.
FR3 If α and β are wffs, so is $(\alpha\lor\beta)$.

CI Lewis actually built his systems around a *binary* modal operator $\Rightarrow$. This requires a change rather in *FR2*. $\alpha\Rightarrow\beta$ in our terms is $\Box(\alpha\supset\beta)$. Can one likewise define $\Box$ from $\Rightarrow$? Lately some have taken the counterfactual operator $\Box\to$ as basic, defining $\Box\alpha$ as, say, $\alpha\lor\neg\alpha \Box\to \alpha$. The book goes into some of this history in Chapter 11.

As the conjunction operator was defined in terms of $\neg$ and $\lor$, the possibility operator $\Diamond$ is defined in terms of $\neg$ and $\Box$:

[Def $\Diamond$] $\Diamond\alpha =_{df} \neg\Box\neg\alpha$.

It could have been done the other way around. Some examples of well-formed formulas are....? Note that all PC wffs are PML-wffs.

Validity in PML

This is where things get interesting. Which modal formulae are going to count as valid? Part of the answer is clear right off the bat, for some modal wffs are valid purely in virtue of their truth-functional structure, e.g., $(\Box p \,\&\, \Box q) \supset \Box p$. This is bound to be true on all propositional assignments whatever $\Box$ means. But other very similar-looking wffs are valid for distinctively modal reasons, e.g., $\Box(p\&q) \supset \Box p$. Here the meaning of $\Box$ matters. Try to show this.

So: is there a general definition of validity that captures both of these cases, that captures modal validity in general? The intuitive, informal answer is like before: a wff is PML-valid iff it stays true no matter which propositions you let its atomic parts express. The difference is that in the PC-context this led immediately to a precise formal definition in which the quantification was over not propositions but truth-value assignments. Due to the non-truth-functional nature of $\Box$, it's not going to be so easy this time around. But something *like* the PC definition, in eschewing propositions, is still possible. The idea behind it has three parts:

(a) Where the truth-value of a nonmodal statement depends only on how things *actually* are —on the truth-values of atomic statements in one world, the actual world— the truth-value of a modal statement depends also on how things *could* have been — on their truth-values as well in counterfactual worlds.

(b) Associated with each world w is a range of worlds that are possible relative to w: we can think of them as the worlds that *would* have been possible had w been actual, although they may not be possible as matters stand.

(c) $\Box p$ is true in a world w iff p is true in all the worlds that would have been possible had w been actual. $\Diamond p$ is true in a world w iff p is true in at least one of the worlds that would have been possible had w been actual.

This three-part idea can be made tolerably precise by means of an elaboration of the PC-game, in which the role of the many possible worlds is played by a plurality of players; and corresponding to the "accessibility relations" among worlds – that is, which worlds are possible relative to which — we have constraints on which of the players are visible to which others.

The range of these "seeing arrangements" — or, as the book says, "seating arrangements" — is left for the time being wide open. It could be that every player can see every other, or it could be that no player can see anyone, or it could be that every player can see only her or himself, or whatever. If you put the seeing arrangements — who is visible to whom — together with a specification of which atoms are written on which players' sheets, the result is called a *modal setting*, or just a *setting*. (The seeing arrangements are going to become "frames"; "models" are the analogue of seeing arrangements plus sheets with atoms.)

Now for the rules. they are 1, 2, and 3 of the PC-game—remind yourself what they are—together with the following two "new" rules, one for box and one for diamond:

4. When $\Box\alpha$ is called, raise your hand iff *every* player you can see raised their hand when α was called.

5. When $\Diamond\alpha$ is called, raise your hand iff *some* player you can see raised their hand when α was called.

If every player bar none raises their hand for α on a given setting, then α is called successful for that setting. A formula that succeeds for every setting is successful period.

The PC game can be thought of as a special instance of the modal game; it's the case where the seating arrangement involves only a single person — it doesn't matter

who they can see since there are no modal operators. Then PC-validity is defined in terms of settings built on seeing arrangements (albeit a single simple kind of arrangement):

A wff holds for a given PC-setting iff the player raises his/her hand when it's called. A wff is valid full stop iff it holds in every PC-setting based on the one and only seeing arrangement.

Modal validity too will be defined in terms of settings built on seeing arrangements

A wff holds for a given PML-setting iff every player raises his/her hand when it's called. A wff is valid in a PML seeing arrangement iff it holds in every setting based on that arrangement. A wff is valid wrt a type of seeing arrangement iff it is valid in every seeing arrangement of the relevant type.

Informally, α is valid wrt a type of seeing arrangement, just if no matter what players so arranged find on their sheets of paper, they all wind up raising their hands when α is called.

Why are we stopping with a *relativized* notion of validity? Why not define validity full stop as validity in every seeing arrangement whatever, rather than every one of a certain type? Good question! Part of the answer is that $\Box$ is not *quite* a logical constant; it's a little bit substantive. This being so there is no notion of pure modal validity to define or analyze. Take a formula like $\Box\alpha \supset \Box\Box\alpha$, or $\alpha\supset\Box\Diamond\alpha$. Are they valid? It depends; on some construals of $\Box$ they are and on others they are not. There is no single, canonical notion of validity because there is no canonical interpretation of $\Box$. The guiding thought of modern, post-Kripkean, modal logic is that

for each interpretation $\Box_k$ of $\Box$, there is a corresponding type C_k on seeing arrangement, such that α is valid with $\Box$ interpreted as $\Box_k$ iff it is valid in every seeing arrangement of type C_k.

This idea has in fact been borne out in many cases although not in all. This course focuses on the good news; we'll mostly be ignoring the parts of the book that construct notions of validity not amenable to the seeing arrangement idea. Suffice it to say that they tend to be "special" and less necessity-like than the ones that are amenable.

Reflexive and Absolute Necessity

To see how it works, there is a *very* natural, very common, idea about $\Box$ according to which it is a "strengthening" operator: $\Box\alpha$ is always a stronger (meaning, not weaker) statement than α, so that the conditional $\Box\alpha \supset \alpha$ comes out valid. $\Box\alpha \supset \alpha$ Is not valid on *all* seeing arrangements. Imagine a two-player game in which each player sees only the other; player 1 keeps her hand down for p while player 2 raises his. Player 1 accordingly raises her hand for $\Box p$ (it's true in all the worlds she can see) despite not raising it for p. So here is an arrangement on which $\Box p \supset p$ is not valid.

Can we find a type of seeing arrangements that corresponds to the (obvious-seeming!) idea of necessity as strengthening? Consider this obvious-seeming constraint on arrangements:

C All of the players can see themselves.

This is called *reflexivity*. We have to convince ourselves that whenever a player raises her hand for $\Box\alpha$ on such an arrangement, she also raises her hand for α. Why would this be? If every hand she sees is up for α, and she can see her own hand, then her own hand is up for α. So $\Box\alpha \supset \alpha$ is valid on every reflexive seating arrangement.

Kind of like the way $\neg$ would not quite be a logical constant if we were giving a semantics flexible enough to accommodate intuitionistic negation.

Even $\Box\alpha\supset\alpha$ is not valid on all interpretations of $\Box$ (belief, provability, obligation).

Now we want to show conversely that if $\Box\alpha \supset \alpha$ is valid on a seeing arrangement, the arrangement is reflexive. This is a little trickier. Suppose for contradiction that the arrangement is not reflexive, and consider a player Aunty Narcissus who cannot see herself. Surely the following could happen: each of the players that Aunty Narcissus can see has p on their sheet, so that she raises her hand for $\Box p$. But she doesn't have p on her own sheet, so when p is called she keeps her hand down. Since she raises her hand for the antecedent but not the consequent, Aunty has got to keep her hand down when $\Box p \supset p$ is called. So, as the book says, "any sense of 'necessary' in which whatever is necessary is true can be reflected by restricting the seating arrangements to those in which all players can at least see themselves (20)."

Suppose for argument's sake that we do not restrict the seeing arrangements in any way. Are there any wffs that are valid in every seeing arrangement whatsoever: that are *absolutely* valid? It turns out there are. Though essentially just one, in the sense that this one can serve as the sole modal axiom for the set of absolutely valid wffs. The formula in question is

K $\Box(\alpha \supset \beta) \supset (\Box\alpha \supset \Box\beta)$

Suppose K were not absolutely valid. Then by the rules for $\supset$ there would have be a setting in which in which some player, call them #1,

(a) raises a hand for $\Box(\alpha \supset \beta)$
(b) raises a hand for $\Box\alpha$
(c) doesn't raise a hand for $\Box\beta$.

How would this come about? From (c) we know that #1 can see a player #2 whose hand stayed down when β was called. From (b) we know that #2's hand was up when α was called. From these two facts it follows that #2's hand was down when $\alpha \supset \beta$ was called. But since #1 can see #2's unraised hand, #1's hand should not be up when $\Box(\alpha \supset \beta)$ is called. By (a), #1 hand *was* up, though. Contradiction. So K is absolutely valid, in honor of which the logic of *absolutely* valid formulas is system K.

Players and Worlds

All that was done with players and paper, but the intended analogy is of course this:

players	worlds
player #n	world w_n
letters on player #n's sheet	atomic sentences true in w_n
#n sees #m	w_m is accessible from w_n
seeing arrangements	accessibility-structures aka frames
settings on a seeing arrangement	models over a frame
success in a setting	truth in a model
success in a seeing arrangement	truth in all models over a frame, validity in the frame
success in all arrangements whatsoever	absolute validity, validity in all frames
success in reflexive arrangements	reflexive validity, validity over reflexive frames
success in type BLAH arrangements	*BLAH*-validity, validity over BLAH-type frames

It's this last that points the way to what's cool about possible worlds semantics, that it turns a miscellany of confusingly related modal logical systems into a series of cleanly related constraints on binary (accessibility) relations.

Some Logical Systems

Time to start looking at some specific logical systems, starting with our "base" system K – the system such that theoremhood in it corresponds to "absolute" validity, ie. validity in all seating arrangements whatsoever. Logical systems can be presented in a number of ways, but we'll do it axiomatically. This involves specifying

(a) the language in which wffs of the system are framed
(b) a selected set of wffs known as the axioms
(c) a set of transformation (inference) rules.

The system's *theorems* will be the closure of the axioms under the inference rules, ie., everything you can get from the former by repeated application of the latter.

Crucially all of the notions (a), (b), and (c) have got to be *effective* in the sense that there's a mechanical procedure for determining what counts as a wff of the language, what counts as an axiom, and what counts as an (instance of) an inference rule. These procedures have also got to be non-semantic in nature. One needn't know what anything stands for or what is true to figure out what the wffs, axioms, and permissible inferences, are. For this reason the axiomatic approach to logical systems is sometimes called the (or a) syntactic approach; the intended contrast being with the semantil approach via structures and validity.

System K

Earlier we talked about absolute validity – validity in all seating arrangements, or all frames – and said that this was the notion of validity captured by system *K*. What is meant by this? That the theorems of *K* are precisely and without exception the absolutely valid wffs. Another word for absolute validity is thus *K*-validity.

To present *K* I've got to tell you (a) what the wffs are, (b) what the axioms are, and (c) what the inference rules are.

(a) The wffs, and indeed the wffs of all systems to be considered for quite a while, are just the wffs of modal propositional calculus.

(b) The axioms are of two kinds; *PC* as written is a "schema" that lays out a whole bunch of axioms at once.

PC If α is a PC-valid wff, then α is an axiom.

K $\Box(p \supset q) \supset (\Box p \supset \Box q)$ is an axiom.

Note that *PC* stands in for an i*nfinite number* of axioms. How does that fit with the idea that the axioms have to be specified effectively or algorithmically? *K* is just one particular sentence, built on the atomic sentences *p* and *q*. Anything of *K*'s form will be a *theorem*, only the others are arrived at using an inference rule.

(c) *K* has three rules:

US Substitution: Replacing $p_1..p_n$ in a theorem by $\alpha_1..\alpha_n$ yields a theorem.
MP or Modus Ponens or Detachment: If α and $\alpha \supset \beta$ are theorems, so is β.
N or Necessitation: If α is a theorem, so is $\Box\alpha$.

To represent the result of systematically substituting the β_is for the p_is in β, we write $\alpha[\beta_1/p_1,...,\beta_n/p_n]$. Where α is a wff and S is an axiomatic system, we write $\vdash_S \alpha$ to express that α is theorem of S. When $\vdash\alpha$ is derivable from $\vdash\alpha$ in the going modal system we write $\vdash\alpha \to \vdash\beta$. With this we can state the rules more simply as

US gets us $\Box(p\&r \supset p\equiv r) \supset (\Box(p\&r) \supset \Box(p\equiv r))$. What else?

What can we get with *N*'s help from $\Box(p\&r \supset p\equiv r) \supset (\Box(p\&r) \supset \Box(p\equiv r))$?

$US \vdash \alpha \rightarrow \alpha[\beta_1/p_1,...,\beta_n/p_n].$

$MP \vdash \alpha, \vdash \alpha \supset \beta \rightarrow \vdash \beta$

$N \vdash \alpha \rightarrow \vdash \Box\alpha$

Notice that only the third of these is a specifically *modal* rule. *N* may strike you as funny; why should α suddenly be entitled to sprout a $\Box$ at the front? Remember that α is only a theorem if it is *valid*; if α is valid, meaning something like true in virtue of the meaning of its logical symbols alone, then it could hardly fail (it seems) to be true in all possible worlds in virtue of those same meanings, which means its necessitation is valid. Later when we get to actuality we'll see a possible hold in this reasoning. (*The President = the actual President* is true in virtue of the meaning of its logical symbols, but it doesn't hold in all worlds.)

Proofs and Provability

A proof in *S* is a finite sequence of wffs, each of which is either (i) an axiom of *S* or (ii) derived from earlier wffs by (iia) applying one of the transformation rules or (iib) making a definitional substitution. Each line of a proof contains three items:

  a wff,
  a reference number for that wff, written immediately before it
  a justification for the wff, written on the far right (not the far left as in the book)

Justifications are basically explanations of how it is that the wff satisfies condition (i) or condition (ii).

As for (i), axioms falling under *PC* are justified by writing "*PC*," perhaps with a number to indicate which of the sample *PC* theorems the wff happens to be. Axioms falling under *K* (there's only one of these!) are justified by writing "*K*."

As for (ii), you should list both the inference rule (or definition) involved and the lines that served as premises. Applications of *US* are justified by writing the earlier line number and then indicating with the square bracket notation the substitution that was made. Applications of *MP* and *N* are marked by "?*MP*" and "?*N*" respectively.

So, let's do it. Like the book we'll first prove two theorems in full detail, and then give some methods for abbreviating proofs. *K1* and *K2* are proved on p. 27; we'll do theorem *K1*.

K1	$\Box(p\&q) \supset \Box p\&\Box q$	
1.	$p\&q \supset p$	PC1
2.	$\Box(p\&q \supset p)$	(1)×N
3.	$\Box(p \supset q) \supset (\Box p \supset \Box q)$	K
4.	$\Box(p\&q \supset p) \supset (\Box p\&q \supset \Box p)$	(3),[$p\&q/p$, p/q]
5.	$\Box p\&q \supset \Box p$	(2),(4)×MP
6.	$p\&q \supset q$	PC2
7.	$\Box(p\&q \supset q)$	(5)×N
8.	$\Box(p\&q \supset q) \supset (\Box p\&q \supset \Box q)$	(3),[$p\&q/p$]
9.	$\Box p\&q \supset \Box q$	(7),(8)×MP
10.	$(p \supset q) \supset ((p \supset r) \supset (p \supset (q\&r)))$	PC3
11.	$(\Box(p\&q) \supset \Box p) \supset ((\Box(p\&q) \supset \Box q) \supset (\Box(p\&q) \supset (\Box p\&\Box q)))$	(10), [$\Box(p\&q)/p$, $\Box p/q$, $\Box q/r$]
12.	$(\Box(p\&q) \supset \Box q) \supset (\Box(p\&q) \supset (\Box p\&\Box q))$	(5), (11)×MP
13.	$\Box(p\&q) \supset (\Box p\&\Box q)$	(5), (11)×MP

You should go through the proof of *K2* yourself. Here's the proof of *K3*, which shows how earlier theorems can be appealed to in new proofs.

K3 $\Box(p\&q) \equiv \Box p\&\Box q$

1. $\Box(p\&q) \supset (\Box p\&\Box q)$ K1
2. $\Box(p\&q \supset p)$ (1)×N
3. $\Box(p \supset q) \supset (\Box p \supset \Box q)$ K
4. $\Box(p\&q \supset p) \supset (\Box p\&q \supset \Box p)$ (3),[*p&q/p, p/q*]
5. $\Box p\&q \supset \Box p$ (2),(4)×MP
6. $p\&q \supset q$ PC2
7. $\Box(p\&q \supset q)$ (5)×N
8. $\Box(p\&q \supset q) \supset (\Box p\&q \supset \Box q)$ (3),[*p&q/p*]
9. $\Box p\&q \supset \Box q$ (7),(8)×MP
10. $(p \supset q) \supset ((p \supset r) \supset (p \supset (q\&r)))$ PC3
11. $(\Box(p\&q) \supset \Box p) \supset ((\Box(p\&q) \supset \Box q) \supset (\Box(p\&q) \supset (\Box p\&\Box q)))$ (10), [$\Box(p\&q)/p$, $\Box p/q$, $\Box q/r$]
12. $(\Box(p\&q) \supset \Box q) \supset (\Box(p\&q) \supset (\Box p\&\Box q))$ (5), (11)×MP
13. $\Box(p\&q) \supset (\Box p\&\Box q)$ (5), (11)×MP

24.244 Modal Logic
Spring 2015