Philosophy 244: #4—Adequacy and Extensions

Metalogic

To be a theorem of K is to be derivable from the K-axioms by the K-inference rules. This is a purely syntactical notion, which pays no attention at all what the symbols might mean. To be K-valid, or as we put it, absolutely valid is to be successful in all settings bar none. This is a semantical notion which has a great deal to do with what the symbols mean; soon we will make it more semantical still by restating the definition of validity in the vocabulary of models.

The point for now is that the notion of K-theoremhood and the notion and the notion of absolute validity are prima facie as different as they could be. If we want to claim a connection between the two notions the connection will have to be proved. Two things will have to be proved.

Soundness: Every K-theorem is K-valid.
Completeness: Every K-valid formula is a K-theorem.

Soundness is easier; that's what we'll do today. Completeness is left for later. The *strategy* for proving these things is worth sketching now, though, because they're distinctive and interesting and convey some of the flavor of metalogical reasoning.

To establish soundness, we use mathematical induction. The basis step has us proving that K's axioms are all K-valid. To carry out the inductive step, we must show that the property of K-validity is preserved under application of K's inference rules.

Mathematical induction works like this. Suppose you want to show that all Xs are P. Suppose that every X can be reached from a small number of "seeds" or "generators" by repeated application of a few operations. First show—this is the basis step—that all of the generators are P. Then show—this is the inductive step—that if some things are P, then applying the operations to them can never produce a non-P. From these two steps the conclusion follows. Why? Examples?

To establish completeness, we look at the contrapositive: every *non*-theorem β of K is K-invalid, that is, β has a countermodel. A non-theorem is something whose negation (or negatum) α is K-consistent, that is, you can't derive a contradiction from α. It's equivalent then to prove that every K-consistent α has a model: K-consistency implies K-satisfiability, in the jargon.

To prove this we start with α, and then pile on other wffs, taking care to preserve consistency, until your pile of wffs can't be expanded any further without becoming inconsistent. You then have a *maximal consistent* set with α in it. A maximal consistent set has so much information in it that it tells us how to construct a model of K in which all the set's members, α included, are true. That α holds in this model means that α is K-satisfiable, which is what we were trying to prove.

This is why we try to make do with a very few axioms and rules. Which axioms and rules often itself reflects the needs of these proofs. That is, axioms and rules are *chosen*, sometimes, less with a view to *ordinary* reasoning than *metalogical* reasoning.

Validity

Soundness and completeness both rely on the notion of K-validity. The account in terms of games was intuitive, which is good; but the features that made it so were inessential to our purposes and in some cases positively misleading, which is bad. Let's try to strip some of these features away.

The set of players gives way to an arbitrary nonempty set *W*, which for motivational reasons is referred to as the set of worlds.

Non-emptiness is important if $\Box\alpha$ is to imply $\Diamond\alpha$. Just as non-emptiness is crucial in quantificational logic if $\forall x\alpha(x)$ is to imply $\exists x\alpha(x)$. Empty domains are allowed in "free logic."

Instead of the seeing-relation, we have an arbitrary binary relation *R* on *W*, that is, an arbitrary set of ordered pairs whose elements are drawn from the set of worlds. *R* is an *accessibility relation* on *W*.

Corresponding to the notion of a seeing arrangement we'll now have frames <*W*,*R*>.

Corresponding to the inscribed sheets of paper, one per player, we'll now have a value-assignment *V* which maps each world and propositional variable taken together to a truth-value: $V(p,\ w) = 0$ or 1.

Corresponding to the notion of a setting (on a seeing arrangement) we'll now have a model <W, R, V>, a "model based on the frame <W, R>." V is said to be based on the frame too.

So: a model M is an ordered triple <W,R,V> where W is an arbitrary nonempty set, R is a binary relation on W (a subset of $W \times W$), and V is a value-assignment. Next is to explain what is is for a formula to be true (false) at an M-world; this is done by expanding the domain of V from propositional variables to all wffs of the language. It's given that for all propositional variables and $w \epsilon W$, $V(p,w) = 0$ or 1. That is the basis clause of a recursive definition. Now we add the recursion clause.

How do recursive definitions relate to inductive proois?

(V$\neg$) $V(\neg\alpha,\ w) = 1$ - $V(\alpha,\ w)$, that is,......

(V$\vee$) $V(\alpha\vee\beta) = \max(V(\alpha,\ w), V(\alpha,\ w))$

(V$\Box$) $V(\Box\alpha, w) = 1$ iff $V(\alpha, u) = 1$ for all u such that wRu, otherwise 0

Which sets up the definition of validity.

α is valid on frame <W,R> iff $V(\alpha,w) = 1$ for all V based on <W,R> and $w\ \epsilon\ W$,

α is K-valid, aka absolutely valid, iff α valid on every frame.

This may seem unnecessarily complicated! Why not just say: α is absolutely valid iff it's valid in every model? Why not indeed? Stay tuned for logics where collapsing the levels leads to the wrong results.

Prop. 2.1 Every theorem of K is K-valid.

To get this we prove a more general result. Let K + Λ be the system obtained from K by adding as extra axioms all the wffs in Λ, and keeping the transformation rules unchanged.

Prop 2.2 If each $\alpha\epsilon\Lambda$ is valid on <W,R>, all theorems of K+Λ is <W,R>-valid too.

Prop. 2.1 is just the special case of **2.2** where Λ is the empty set. (Why?)

The proof of 2.2 uses two lemmas.

Lemma 2.3 All valid PC wffs and axiom K are valid on all frames.
Lemma 2.4 For any frame F, the set of wffs valid on F is closed under K's rules.

This is enough for **2.2**. Why? Now the proofs.

Proof of 2.3: Let <W,R> be given. A valid PC-wff is true on every (regular, non-modal) valuation V of its variables; so $V(\alpha,w)$ is true for every modal V, hence every V based on <W,R>; so it is valid on <W,R>. Suppose for contradiction that K is not valid on <W,R>. Then there is a V on <W,R> and a $w\epsilon W$ such that (i) $V(\Box(p\supset q),w) = 1$, (ii) $V(\Box p,w) = 1$, but (iii) $V(\Box q,w) = 0$. By (iii), w bears R to a u such that $V(q,u) = 0$. By (ii), $V(p,u) = 1$. But then w bears R to a u such that $V(p\supset q,u) = 0$, contrary to (i). $\Box$

Proof of 2.4: Let <W,R> be given. [US] Suppose that $\alpha[\beta_1/p_1...]$ is not valid on <W,R>. Then there's a model based on <W,R> such that for some $u\epsilon W$, $V(\alpha,u) = 0$. Define V* so that for all $w\epsilon W$ $V^*(p_i,w) = V(\beta_i,w)$. Then $V^*(\alpha,w) = 0$ — proof by induction — so α was not valid on <W,R> in the first place. [MP] If α and $\alpha\supset\beta$ are valid on <W,R>, then both are true in every world of every model based on <W,R>; hence by (derived evaluation rule) ($V\supset$), the same applies to β. [N] If α is valid on <W,R>, then for all models based on <W,R> and all $u\epsilon W$, $V(\alpha,u) = 1$; hence for all $w\epsilon W$, α is true in all worlds w bears R to; hence for all $w\epsilon W$, $V(\Box\alpha,w) = 1$; hence $\Box\alpha$ is valid on <W,R>. $\Box$

Extensions of K

Now, clearly the most noteworthy thing about system K is that the notion of necessity it captures is not "strengthening": it's not a notion according to which whatever is necessary has got to be true. How do we know this? Proved earlier that $\Box p \supset p$ isn't K-valid, and by soundness, if it isn't K-valid it isn't a theorem of K. We look now at an extension of K that adds "strengthening" as an additional axiom. System T is just like system K except its axioms are PC, K, and

So named by the Belgian logician Feys in 1937, working off of a system devised by Godel. Feys wrote a book in 1958 with William Craig. Craig went on to teach at Berkeley for many years and is now a Professor Emeritus there.

T $\Box p \supset p$

This is called the axiom of necessity, not to be confused with the rule of necessitation. This is probably the weakest system that anyone seriously regards as having a chance of capturing the notion of metaphysical necessity. Here are a couple of important theorems of T that are not theorems of K.

Later we might look at a paper of Nathan Salmon's arguing that nothing stronger than T is philosophically defensible.

T1 $p \supset \Diamond p$

1.	$\Box\neg p \supset \neg p$	T
2.	$p \supset \neg\Box\neg p$	(1)xPC
3.	$p \supset \Diamond p$	(2)xDefN

T2 $\Diamond(p \supset \Box p)$

1	$\Box p \supset \Diamond\Box p$	T1[/]
2	$\Diamond(p \supset \Box p) \equiv (\Box p \supset \Diamond\Box p)$	K7[/]
3	$\Diamond(p \supset \Box p)$	(1),(2)x*Eq*

Suppose we wanted to show that these weren't theorems of K. How would we do it? Can you think of countermodels? The main reason for mentioning T2 is to show that the following, which might *seem* like just a dual-ish counterpart of the necessitation rule, is not a rule of system T:

$\vdash \Diamond\alpha \rightarrow \vdash \alpha$

Shouldn't we be able to get this from necessitation by contraposition?

$\vdash\alpha \rightarrow \vdash\Box\alpha$..........Necessitation
$\neg\vdash\Box\alpha \rightarrow \neg\vdash\alpha$.....Contraposition
$\vdash\Diamond\neg\alpha \rightarrow \vdash\neg\alpha$.....Not-must=maybe-not
$\vdash\Diamond\beta \rightarrow \vdash\beta$......$\beta$ subbed in for $\neg\alpha$

Hmmm...

If this were a rule, then from T2 you'd get

$\vdash(p \supset \Box p)$.

From which it follows by necessitation that

$\vdash \Box(p \supset \Box p)$,

in other words, *necessarily, any truth is necessarily true*. Hopefully this is not a theorem of T! Why intuitively would $\vdash\Diamond\alpha \rightarrow \vdash\alpha$ fail, though? How can it be a theorem that α is possible when it's not a theorem that α is true?

Validity for T

Think back to the argument a few days back that $\Box p \supset p$ isn't *absolutely* valid. To get a countermodel we had to use a non-reflexive frame. If that "had to" holds up, then $\Box p \supset p$ is valid on every frame that's reflexive, that is, every world in *W* bears *R* to itself.

Call a formula T-valid iff its valid on every reflexive frame. **Propn. 2.2** shows, if you think about it, that system T is sound relative to this definition of validity. (Now we see the role of Λ is playing in that proposition.) Later we'll see that the reflexive semantics is complete as well.

For now let's go back to the declaration above that $\Box(p \supset \Box p)$ had better not be a theorem of T. How do we show it isn't? Let's think of it semantically. What do we need to find? A reflexive frame on which $\Box(p \supset \Box p)$ is not valid. To get *that* we need a model on such a frame where it isn't true. The simplest idea would be a one-world model.....you take it from there.

System D

T is clearly an extension of K. Now we consider a weaker extension that lies midway between K and T. If $\Box$ expresses obligatoriness, we don't want $\Box p \supset p$ to be a theorem. Leibniz thought he could prove that our world is the best possible; if to be the best possible it can't have any unmet obligations, then Leibniz maybe would want *p ought to be the case, so it is the case* to be a theorem. Leibniz was wrong about this, though, as shown by Voltaire. Nobody today wants *p ought to be the case, so it is the case* to come out a theorem.

But a weaker claim may still seem right: whatever ought to be the case is permitted, that is, nothing is both obligatory and forbidden. This leads us to system D (for deontic), defined as K with the additional axiom

D $\Box p \supset \Diamond p$.

Someone might I suppose want to disagree with even this, on the ground that it rules out moral dilemmas: you ought to save Alice, and you ought to save Bert, but you can't do both. Are you permitted to save Bert? Maybe not, since then you won't be able to fulfill your obligation to Alice.

Our first theorem is to the effect that not absolutely everything is forbidden.

D1	$\Diamond(p \supset p)$	
1	$p \supset p$	PC
2	$\Box(p \supset p)$	(1)xNE
3	$\Box(p \supset p) \supset \Diamond(p \supset p)$	D[$p \supset p/p$]
4	$\Diamond(p \supset p)$	(2),(3)xMP

A derived rule is $\vdash \alpha \rightarrow \vdash \Diamond\alpha$. Does that mean every truth is permitted?! Let's hope not!

To place D with respect to our other systems, we ask: is D contained in T? Equivalently is $\Box p \supset \Diamond p$ provable in *T*? It is if $p \supset \Diamond p$ is provable in T. Hint: it's more or less the contrapositive of $\Box p \supset p$. The next question is whether D is stronger than K. Clearly so. The book shows that D is the weakest extension of K which has any theorems of the form $\Diamond\alpha$.

How do we get this? Let K+ be an extension of K with $\Diamond\alpha$ as a theorem. We have to show that K+ contains D, which comes down to having $\Box\alpha \supset \Diamond\alpha$ as a theorem

D	$\Box p \supset \Diamond p$	
1	$\Diamond\alpha$	K+
2	$q \supset (p \supset p)$	PC
3	$\alpha \supset (p \supset p)$	US[/]
4	$\Diamond\alpha \supset \Diamond(p \supset p)$	(3)xDR3
5	$\Diamond(p \supset p)$	(1),(4)xMP
6	$\Diamond(p \supset p) \equiv (\Box p \supset \Diamond p)$	K7[/]
7	$\Box p \supset \Diamond p$	(5),(6)xMP

Validity for D

By a *dead end* let's mean a world that can't see anything. A frame is*serial* if it contains no dead ends; every world can see at least one world — maybe itself, maybe a different world. D-validity is validity on every serial frame. To establish soundness it suffices by **Prop. 2.2** to show that axiom D is valid on every serial frame. Suppose not; then some model based on a serial frame has a world *w* in which $\Box p$ holds and $\Diamond p$ fails. But the only way for that to happen is for *w* to be a dead end; if it could see any world *u*, *u* would have to both be and not be a *p*-world which is contradictory. Completeness is left for later. Next time: S4, B, and S5; what they're good for and their relations to K, D, and T. The ordering is going to be: K < D < T < B, S4 < S5.

Chisholm's Paradox of Contrary-to-Duty Imperatives

1. Smith ought to go help his neighbors: $\Box g$
2. He ought to tell them he is coming, if he does: $\Box(g \supset t)$
3. If he doesn't go, he ought not to tell them he is coming: $\neg g \supset \neg\Box t$
4. He is not going to help them: $\neg g$

These claims ought to be consistent! But, (2) implies $\Box g \supset \Box t$ (by K), which with (1) implies $\Box t$, while (3) and (4) imply $\neg\Box t$ by Modus Ponens.