Lecture 3: A Case for the Rational Revisability of Logic.

Earlier discussions on rational revisability of logic:

Putnam: to understand QM, accept counterinstances to distributive law.

In 2-slit experiment,

accept that the particle went through either Slit A or through Slit B and hit the screen at a certain point,

but deny that it went through Slit A and hit the screen at that point, and also that it went through Slit B and hit the screen at that point.

"Quantum logic" was to be our all-purpose logic: classical logic holds only as an approximation, in macroscopic contexts where quantum superposition can be safely ignored.

Dummett: adopt intuitionist logic, **not just as a special "logic" for mathematics but as an all purpose logic:** classical logic is appropriate only for sentences that are in-principle decidable.

It's hard to totally sympathize with either Putnam or Dummett, because their cases for change of logic seem exceptionally thin.

(Putnam also tried to use the example of quantum logic to argue not only that we can rationally revise logic, but that we can rationally revise it *on empirical grounds*. This extension raised a lot of additional problems.)

But might we do better?

I don’t underestimate the epistemological problems here: it is difficult to accommodate rational change of logic into a serious epistemological model (e.g. a probabilistic model).

Typical such models *leave no room for cognitive states that don’t accord with classical logic*.

> They seem to say that logic is immune to rational revision, and to declare anyone who has a non-classical logic automatically irrational.

Maybe after altering logic, we can alter probability theory to accommodate it.

> But: This **gives only an *after-the-fact* revision** of the epistemology, and **one that only accommodates *one particular* revision of logic**.
>
> It gives no illumination about how the debate to change the logic in this way might have been conducted.
>
> And the resulting theory would still hold that *its* logic is immune to rational revision.

Are there non-probabilistic accounts that do better?

Vague slogans:

Quinean holism

Goodman: appeal to reflective equilibrium.

The slogans are appealing, but attempts to fill them out with even a modicum of precision tend to presuppose a logic.

By rationally revising logic I mean

> rationally revising our basic logical modes of reasoning,

not (just)

> rationally revising our opinions about which modes of reasoning preserve truth.

Doing the first usually involves doing the second, but not necessarily conversely. (Examples later.)

As we saw last time, change of modes of deductive reasoning typically brings with it changes in much more.

For instance, if we modify our basic modes of reasoning so as to give up disjunctive syllogism (the inference $A \vee B, \neg A \vdash B$), we're almost sure to also give up the rule of explosion ($A, \neg A \vdash B$). And then we'd better alter a standard constraint on degrees of belief, that the degrees of belief in A and in ¬A can never add to more than 1. This will mean that we will have to make changes in basic modes of *inductive* reasoning as well.

That the issue is the first is important for understanding why it's hard to get an account of rational change of logic.

I'll make a case for restricting certain classical laws, primarily the law of excluded middle, in order to deal with the semantic paradoxes and the paradoxes of naive property theory.

> What I'll be advocating *isn't* intuitionist logic: that's no better than classical in dealing with the paradoxes (and also strikes me as quite counterintuitive).
>
> Instead, a logic that was only invented in the last few years, though it is of a broad type that has some well-known examples.
>
> The logic reduces to classical when certain assumptions are made. These assumptions seem plausible within ordinary mathematics and physics; so unlike Putnam's and Dummett's proposals, *there is no need to make modifications in ordinary mathematics and physics*.

(Strictly, the case for restricting classical laws isn't *yet* a case that we should *change* our logic, for a *possible* view is that the logic I'll be advocating is the one we've really been employing all along at some deep level.)

All I need for present purposes is to convince you that there is *a serious case to be made* for solving the paradoxes by means of this logic, and that

> *it would not be irrational to alter one's logic* on the basis of this case;

or even just

> *there can be a rational debate about whether to alter one's logic* on the basis of this case.

****

Heterologicality paradox:

(TO) $\forall x$[‘$F(v)$’ is true of $x \leftrightarrow F(x)$]

In particular

‘$F(v)$’ is true of itself $\leftrightarrow F$(‘$F(v)$’)

Apply to case where ‘$F(v)$’ is ‘v is not true of itself’. We get

‘v is not true of itself’ is true of itself $\leftrightarrow$ it is not true of itself.

This has the form $B \leftrightarrow \neg B$, and hence is classically contradictory.

Turning the argument around: we get a classical argument for a claim of form

Either

(i) ‘$F(v)$’ is true of c, but not $F(c)$ [**Overspill**]

or

(ii) $F(c)$, but ‘$F(v)$’ is not true of c. [**Underspill**]

There must be either overspill, or underspill, or both. (Not just in this example, but in many others.)

Interesting.

Russell's summary of his discussion of Hegel's philosophy (*History of Western Philosophy*):

> "The worse your logic, the more interesting the consequences to which it gives rise."

Analog for truth: Assuming classical logic, we must have either

True(⟨*A*⟩), but ¬*A*	**[Overspill]**
A, but ¬True(⟨*A*⟩)	**[Underspill]**

(E.g., because of Liar sentences.)

Three types of classical solutions:

(1) Positing underspill

(2) Positing overspill

(3) Neither positing underspill nor positing overspill, but demanding that there be one or the other.

(Non-exhaustive: some solutions fall under both (1) and (2).)

(1) Paradigmatic **Underspill** theories: posit **truth-value gaps.**

Let **L** be the name of a sentence that asserts its own untruth. The gap theory asserts

L is not true

but also

$\langle$**L** is not true$\rangle$ is not true.

It asserts something, but in the same breath, asserts the untruth of what it has asserted.

Motivation: gap theories allow one to accept with full generality the “left to right half of the Tarski biconditionals”: the schema

(T-OUT) $\quad$ $\mathrm{True}(\langle A\rangle) \rightarrow A$.

(Or the analog for ‘true of’:

$\langle F(x)\rangle$ is true of $o \rightarrow F(o)$.)

But the price is high: not only are there sentences that the theory asserts while in the same breath asserting to be untrue, it turns out that this is so *even for instances of (T-OUT).*

That is, the theory accepts all instances of (T-OUT), but also declares that not all instances of (T-OUT) are true.

Essentially, **the theory asserts that not all of its own axioms are true**.

For a theory to declare some of its axioms untrue is not just weird, it goes against the whole point of the notion of truth.

It's often said that the point of the notion of truth is to enable one to express agreement and disagreement in cases where one is otherwise unable to.

I'll argue in a few minutes that this *underestimates* the point, but it is certainly part of the point.

Illustration: Suppose I'm considering a theory given by a set of axioms that is recursive but infinite. I'm pretty sure that it goes wrong somewhere, but don't know precisely where. How can I express my disagreement with it, other than by saying "Not every axiom of the theory is true"?

But this will serve its purpose only if for each axiom, the claim that the axiom is true is equivalent to the axiom itself.

And this equivalence fails, according to the gap theorist.

> A dramatic case: the gap theorist obviously agrees with his own gap theory, but thinks that not all of its axioms are true.

Can we get around this, say by expressing disagreement in terms of falsehood rather than untruth, or something of the sort?

No. Suppose a gap theorist hears Jones say

(J) The Liar sentence is not true,

and hears Smith say

(S) The Liar sentence is true.

The gap theorist

(i) agrees with Jones and disagrees with Smith.

but

(ii) regards both their utterances as untrue and regards neither of them as false.

(J) and (S) are to be evaluated in exactly the same way by the gap theorist as regards truth and falsity, even though he agrees with (J) and disagrees with (S).

Using untruth or falsity to express disagreement is clearly impossible in the theory.

(2) Paradigmatic **Overspill** theories: posit **truth value gluts**. (Not dialetheism as standardly understood: logic is classical.)

Most of them accept the other half of the Tarski truth theory,

(T-IN) $A \rightarrow \text{True}(\langle A \rangle)$.

Obviously they avoid one problem of the gap theories: they do declare their own axioms true.

But

(i) they declare some of their axioms false as well as true.

(ii) **they declare that some of their own rules of inference**, such as modus ponens, **don't preserve truth.**

E.g., as classical theories, they accept the inference

L is true → 0=1

L is true

So 0=1.

They also accept that both premises are true, but that the conclusion isn't true!

I think that this is actually slightly less bad than declaring one's axioms untrue, but this would take too long to explore.

Glut theories are as bad as gap theories as regards agreement and disagreement. Go back to the example of

(J) The Liar sentence is not true,

and

(S) The Liar sentence is true.

The glut theorist

(i) agrees with (S) and disagrees with (J),

but

(ii) regards them both as true and both as false.

Again, ***the glut theorist can't distinguish between (J) and (S) as regards their truth-theoretic properties, even though he agrees with one and disagrees with the other.***

Again, using untruth or falsity to express disagreement is impossible in the theory.

(3) Theories that neither posit underspill nor posit overspill, but demand that there be one or the other.

This may seem an uninteresting position: just agnosticism between the two positions we've already considered and found problematic.

But there's a version that isn't straightforward agnosticism: intuitively, it's a position according to which it is *indeterminate* whether there is underspill or overspill in any given case. (Need to give this content.)

The theories accept (most or all of) the following truth rules:

(T-Introd) $A \vDash \text{True}(\langle A \rangle)$

(T-Elim) $\text{True}(\langle A \rangle) \vDash A$

(¬T-Introd) $\neg A \vDash \neg\text{True}(\langle A \rangle)$

(¬T-Elim) $\neg\text{True}(\langle A \rangle) \vDash \neg A$.

They avoid paradox because, though they accept all classical rules of inference, *they restrict certain classical meta-rules*.

They reject conditional proof, so (T-Introd) doesn't lead to $A \rightarrow \text{True}(\langle A \rangle)$, and (T-Elim) doesn't lead to $\text{True}(\langle A \rangle) \rightarrow A$.

They also reject *reductio* (¬-Elimination).

They also reject reasoning by cases (∨-Elimination).

It's natural to call these theories *weakly classical*.

The rejection of reasoning by cases plays an essential role:

Let CONT be the contradiction

⟨**L** is not true⟩ is true and ⟨**L** is not true⟩ is not true.

1. ⟨**L** is not true⟩ is true ⊨ CONT

   [Use (T-Elim) and the identity **L** = ⟨**L** is not true⟩.]

2. ⟨**L** is not true⟩ is not true ⊨ CONT

   [Use (T-Introd) and the identity.]

3. ⊨ ⟨**L** is not true⟩ is true or ⟨**L** is not true⟩ is not true.

So: the disjunction of two things that imply contradictions is a logical truth!

This may seem absurd. But that feeling rests on the principle of reasoning by cases: the principle that if A implies C and B implies C then $A \vee B$ implies C.

Yogi Berra's advice.

We've seen that classical (including weakly classical) theories imply that there is either Overspill or Underspill.

But by the weakly classical theorist's truth rules, the existence of Overspill is contradictory, and so is the existence of Underspill.

These theories reject both Underspill and Overspill as contradictory, but say that there must be one or the other!

Back to agnosticism v. indeterminacy:

Presumably when we are merely agnostic about the truth of *A*, we will allow the meta-inference from

$$\text{True}(\langle A \rangle) \vDash C$$

and

$$\neg\text{True}(\langle A \rangle) \vDash C$$

to

$$\text{True}(\langle A \rangle) \vee \neg\text{True}(\langle A \rangle) \vDash C$$

and hence to

$$\vDash C.$$

So the failure to accept the inference is a sign of something beyond mere agnosticism. "Indeterminacy" seems a natural term, though this is a label rather than an explanation.

Though I think these views counterintuitive, they have a prima facie advantage:

> By accepting the four truth rules, these theories avoid the problem of agreement and disagreement, as I stated them.

BUT: The problem of agreement and disagreement was really just the tip of a bigger iceberg.

Talk of truth isn't *just* a means of expressing agreement and disagreement, for the same reason that talk of goodness isn't *just* a means of expressing approval and disapproval: 'true', like 'good', occurs in embedded contexts (contexts embedded more deeply than a negation).

In particular, 'true' is used inside conditionals. And in order for it to serve its purpose, it needs to be well-behaved there.

That is: inside conditionals as in unembedded contexts, 'true' needs to serve as a device of infinite conjunction or disjunction (or more accurately, a device for allowing quantification).

Suppose I can't remember exactly what was in the Conyers report on the 2004 election, but say

(X) If everything that the Conyers report says is true then the 2004 election was stolen.

Suppose that what the Conyers report says is $A_1,...,A_n$. Then relative to this last supposition, (X) better be equivalent to

(Y) If A_1 and ... and A_n then the 2004 election was stolen.

And this requires True($\langle A \rangle$) to be intersubstitutable with A even when A is the antecedent of a conditional.

The point generalizes to more complex constructions: in order for the notion of truth to serve it's purposes, we need the following ***Intersubstitutivity Principle***:

> If *C* and *D* are alike except that (in some transparent context) one has "*A*" where the other has "⟨*A*⟩ is true", then one can legitimately infer *D* from *C* and *C* from *D*.

And weakly classical theories, like classical gap and glut theories, have to reject this.

> Since such theories accept $A \leftrightarrow A$, Intersubstitutivity would lead to $\text{True}(\langle A \rangle) \leftrightarrow A$, which we know that no classical theorist can consistently accept.

So though weakly classical theories may handle the problem of agreement and disagreement, the significance of this is open to question.

How can we keep the Intersubstitutivity Principle?

One possibility: Use Kripke's Kleene-based fixed point constructions *in the following way*:

> Take the set of sentences contained in any such fixed point as your theory of truth.

(*Not* a gap theory.)

This gives a theory of truth based on a non-classical logic without excluded middle: the so-called *strong Kleene logic*.

The downside: this logic, and the theory of truth based on it, are extraordinarily weak.

In particular, the only conditional it contains will be the one defined from ¬ and ∨ in the usual way, so that $A \to A$ will just be an instance of excluded middle. So

> $A \to A$ won't be a general law.
>
> Neither will lots of other natural laws, such as $A \to A \vee B$, or $(A \to B) \to (\neg B \to \neg A)$.
>
> We also won't get either direction of the Tarski biconditionals. E.g., True(⟨*A*⟩) $\to A$ is equivalent to $A \to A$, which we don't have.

A second (related) difficulty: there is no way in this logic to say that sentences such as the Liar are defective.

Contrary to Kripkean rhetoric, we can't say that they are neither true nor false.

The natural substitute is to say that they're neither determinately true nor determinately false. But *no such notion of determinateness is available in the logic*.

To address the first difficulty, I propose that we try to strengthen Kleene logic by adding a new and better-behaved conditional, while still retaining Intersubstitutivity.

It turns out that this will handle the second difficulty as well: we'll be able to use the conditional to define defectiveness.

But how is the conditional to work? There are well-known paradoxes that need to be circumvented (e.g. the Curry-Löb paradox), and circumventing them proves hard.

One logic which “almost works” is **Lukasiewicz’s continuum-valued logic**, on which sentences take on semantic values in the unit interval [0,1]. 1 is the sole “designated value”.

Logical operators correspond to functions on these values: e.g.,

$$|\neg A| = 1 - |A|$$

$$|A \rightarrow B| = \begin{cases} 1, & \text{if } |A| \leq |B| \\ 1 - (|A| - |B|), & \text{if } |A| > |B|. \end{cases}$$

(This last yields that $|A \rightarrow A| = 1$, unlike in Kleene semantics.)

It turns out that *if we restrict the use of quantifiers in certain ways*, we can get the Intersubstitutivity Principle and the Tarski biconditionals to hold in this logic.

> (This has a rather nifty proof, using the Brouwer fixed point theorem on the Hilbert cube.)

Good, but not good enough: what we need is a logic that does this, without ad hoc restriction on the quantifiers.

This can be done.

To give you some idea how, without the details, I’ll say more about the Lukasiewicz case.

An important feature of the Lukasiewicz conditional is that it can be used to define a natural notion of determinateness:

'It is determinately the case that *A*' is defined as $\neg(A \rightarrow \neg A)$.

This defines a function from semantic values to semantic values, given by the following graph:

DRAW

This has the properties we'd expect of a notion of determinateness:

(ia) If $|A|=1$, $|DA|$ should be 1

(ib) If $|A|=0$, $|DA|$ should be 0

(ic) If $0<|A|<1$, $|DA|$ should be strictly less than $|A|$

(ii) If $|A|\leq|B|$, $|DA|$ should be less than or equal to $|DB|$.

Indeed, we can strengthen (ib), to

(ib-s) If $|A| \leq |\neg A|$, $|DA|$ should be 0;

this says in effect that if $|A|\leq½$, $|DA|$ should be 0.

The Liar sentence will have value ½, so (ib-s) allows us to assert that that the Liar sentence is neither determinately true nor determinately false.

What about the “determinate Liar” paradox involving a sentence L_1 that declares itself not *determinately* true?

Not a problem: such a sentence can be given value 2/3, consistent with Intersubstitutivity.

You can’t say of such a sentence that it isn’t determinately true: that claim has value only 2/3, and you aren’t allowed to assert sentences with value less than 1.

But you *can* say that it isn’t *determinately determinately* true: that has value 1. (You can also say that it isn’t determinately false.)

More generally: if L_n is a Liar-like sentence equivalent to $\neg D^n\text{True}(\langle L_n\rangle)$, where D^n is the n^{th} iteration of D, then it is consistent with Intersubstitutivity to assign L_n the value $2^n/(2^n+1)$. (This is easy to illustrate graphically: the graph of D^n is

DRAW

so the graph of $\neg D^n$ is

DRAW

and the place where it crosses the drawn diagonal is the value of L_n.)

This will license the assertion of

$$\neg D^{n+1}\text{True}(\langle L_n\rangle),$$

but not of

$$\neg D^n\text{True}(\langle L_n\rangle);$$

so ***D^{n+1}* is always stronger than *D^n***.

The sentences L_n are all non-paradoxical in the theory.

But the introduction of quantifiers poses a problem: they (together with a truth predicate) allow us to construct a limit operator D^ω.

$D^\omega A$ means "For every natural number *n*, the result if prefixing *n* occurrences of *D* to *A* is true".

But this is a **monster operator**: its negation assigns 0 to sentences with value 1 and 1 to sentences with value less than 1, and this re-institutes paradox.

The key to resolving the paradoxes is to modify the Lukasiewicz semantics in such a way that quantification can't produce monsters. It's tricky, but not impossible, to do this.

Here's a **general framework**, of which the Lukasiewicz semantics is a special case.

We introduce: a space V of values and a partial ordering $\preceq$ on it, with a greatest value 1 and a least value 0. (We don't require that the ordering be total.) 1 will be the sole "designated value".

We stipulate the following about the ordering:

> (i) Any two members of V have a least upper bound and a greatest lower bound;
>
> (ii) 1 is not the least upper bound of any two elements both of which are less than 1.

Given (i), we can take $|A \vee B|$ to be the least upper bound of $|A|$ and $|B|$, and $|A \wedge B|$ to be the greatest lower bound. The (ii) guarantees reasoning by cases.

For quantifiers, we do something analogous.

> There's a minor obstacle to getting analogs of (i) and (ii) together, but it can be overcome: e.g. instead of requiring that the space be complete (*every* set of values has a least upper bound), just require that it be "complete enough", i.e. have *sufficiently many* least upper bounds and greatest lower bounds. Properly done, this removes any conflict with (ii).

For negation, we posit an "up-down symmetry" * on the space: an order-reversing operator that when applied twice to any object leads back to that object.

Finally, we need an operator on V corresponding to the conditional. We can demand that it satisfies such laws as these:

(I) $a \Rightarrow b$ is 1 if and only if $a \preceq b$

(IIa) if $b_1 \preceq b_2$, then $(a \Rightarrow b_1) \preceq (a \Rightarrow b_2)$

(IIb) if $a_1 \preceq a_2$, then $(a_2 \Rightarrow b) \preceq (a_1 \Rightarrow b)$

(III) $1 \Rightarrow 0$ is 0

(IV) $(a^* \Rightarrow b^*) = (b \Rightarrow a)$.

We could add many more, but let's not pre-judge which ones ought to be added.

The framework just sketched generalizes the Lukasiewicz semantics.

We want to assign values to sentences (and more generally, to formulas relative to an assignment function) *in a way that accords with the Intersubstitutivity Principle for 'true'*.

> We've seen that in the Lukasiewicz semantics, there is no way to do this.
>
> **But it turns out that there are other instances of the general framework in which this can be done.**

Indeed, we can get a more general Intersubstitutivity Principle, for ‘true of’ or ‘satisfies’.

In this way, **we can get a general resolution of all the semantic paradoxes—and the property-theoretic paradoxes too—which accords with the “naive schemas” of semantics and property-theory.** [Remark on set theory.]

(The value space we need is infinite valued, and only partially ordered.)

To get some sense of the resulting theory, consider **the hierarchy of determinateness operators**.

A problem with the Lukasiewicz theory was that if you iterate the sequence of operators *D*, *DD*, *DDD*, ... into the transfinite, you get a collapse into a monster operator; indeed, the collapse occurs at level ω.

And yet such an iteration is clearly definable in the language, using ordinary set theory plus 'true'.

This is what rendered the Lukasiewicz semantics unsuitable for the paradoxes.

How is the problem avoided in the proposed theory? There is still a determinately operator *D*, defined in almost the same way as before.

Transfinite iterations are allowed. (This requires a theory of ordinal notations, but you can generalize the usual Church-Kleene theory in a way that allows the notations to extend much further, into the non-recursive ordinals.)

For each notation α, we can define "the α^{th} iteration of *D*", and so we can define an "α^{th}-level Liar sentence" L_α that asserts that it itself is not D^α-true.

What the status of L_α? We get the following nice results (generalizing into the transfinite what we had in the finite case for Lukasiewicz): **L_α is not definitely false, and not definitely$^{\alpha+1}$-true.**

The presence of the '+1' is what blocks the paradox.

Can't we restore paradox by a Super-Liar sentence S that says "For every notation α, $\langle S \rangle$ is not D^{α}-true"?

No.

This *can* be formulated in the language.

But it turns out not to produce paradox. The reason is related to the reason why any system of ordinal notations must eventually break down, but this is a rather technical matter that I can't go into today.

[If you define D^{α} on the ordinals itself as opposed to their notations ("There is a notation for α such that ...."), it must *in a sense* collapse.

But it never collapses *to the monster*.

Rather, it collapses to something that is in no sense an iteration of D, because even as applied to sentences with value 1 it yields something with value less than 1.

Moreover, it is indeterminate where that "collapse" occurs.]

Summary:

(1) The theory adumbrated saves the naive principles of truth, satisfaction, property instantiation and so on, in a weakening of classical logic.

(2) It thus avoids the major problems that beset the classical and weakly classical theories.

I think this gives a strong case for weakening classical logic.

The modification needn't make a difference within ordinary mathematics and physics:

> One can assume excluded middle in those domains, and in the presence of excluded middle the conditional behaves just like the classical conditional, so that the logic becomes effectively classical.

I’m fairly happy with this theory. But one might not be, for various reasons: either

> the intuitive attractiveness of keeping excluded middle unrestricted,

or

> something more specific.

So it might seem reasonable to keep to classical or weakly classical resolutions of the paradoxes despite their problems. That’s a matter for rational debate.

But that’s my point**:** as a matter for rational debate, it’s something that we must accommodate in a theory of rational debate: we shouldn’t imagine that classical logic is built into our epistemic norms in such a way as to make rational debate about it impossible.

In the late 1960's and early 1970's there was a great deal of discussion of whether it could ever be rational to change our logic.

- Putnam famously argued that we could best solve the problems in coherently understanding quantum mechanics by accepting counterinstances to the distributive law: in the context of the two-slit experiment, we should accept that the particle went through either Slit A or through Slit B and hit the screen at a certain point, but deny that it went through Slit A and hit the screen at that point and also that it went through Slit B and hit the screen at that point. This "quantum logic" was to be our all-purpose logic: the idea was that classical logic holds only as an approximation, in macroscopic contexts where quantum superposition could be safely ignored.
- Dummett argued that we should adopt intuitionist logic, not just as a special "logic" for mathematics but as an all purpose logic: classical logic was supposedly appropriate only to sentences that are in-principle decidable.

It's hard to totally sympathize with either Putnam or Dummett, because their cases for change of logic seem exceptionally thin. (Putnam also tried to use the example of quantum logic to argue not only that we can rationally revise logic, but that we can rationally revise it *on empirical grounds*. This extension raised a lot of additional problems.) But might we do better?

A common reaction to their proposals was that not only were they wrong in detail, but that the whole idea of rationally revising logic is simply incoherent. Sometimes it was said that we can simply refute Putnam's proposal by pointing out that his assertions violated the distributive law. Even Quine, who had argued in his early work that logic was no more immune to rational revision than any other subject was, seemed to say in his 1970 book that if anyone advocated a change in logic we should translate him as not meaning what he was saying. So Putnam couldn't really have meant that we should change logic, his apparent protests to the contrary notwithstanding!

Such a view strikes me as unfortunate. However, there is one fact which does give it prima facie support: