Propositions, Truth and Paradox

PHIL 428/628, Week 4: Feb. 9 2016

KRIPKE'S THEORY OF TRUTH (CONTD)

1. Limitations of the Model

1.1. Inexpressible Untruth and Ghosts

The idea behind the approach is that a Liar sentence ¬T(c) (where c denotes this sentence) is neither true nor false. In particular, it is not true. But this sense in which this sentence is untrue is not expressible using ¬T(x):

- because ¬T(c) is not in $S_{1,\sigma} \cup S_{2,\sigma}$, so ¬T(x) is not true of it.

So: there is at least a natural notion of untruth that ¬T(x) doesn't express.

One can turn $\mathcal{L}_\sigma$ into a language that can express this notion of untruth: one just adds all of the sentences not in $S_{1,\sigma} \cup S_{2,\sigma}$ to $S_{2,\sigma}$. So in this language T is interpreted by ($S_{1,\sigma}$,D–$S_{1,\sigma}$). (Kripke describes the process of moving from ($S_{1,\sigma}$,$S_{2,\sigma}$) to ($S_{1,\sigma}$,D–$S_{1,\sigma}$) as 'closing-off'.) Call this language $\mathcal{L}_\sigma'$.

In $\mathcal{L}_\sigma'$ ¬T(x) is true precisely of the things that T(x) is *not* true of in $\mathcal{L}_\sigma$. But this is no longer a language that contain its own truth predicate: since ¬T(c) is true in $\mathcal{L}_\sigma'$, but T(x) is not true of this sentence in $\mathcal{L}_\sigma'$.

Of course, it is not simply that in $\mathcal{L}_\sigma$ ¬T(x) fails to be true of everything not in the extension of T(x). Given some apparently innocuous assumptions about self-reference, no formula A(x) can be true precisely of these things.

- For suppose A(x) was true precisely of the things in D–$S_{1,\sigma}$.
- Then consider a sentence A(b) where b denotes this sentence.
- If A(b) is true in $\mathcal{L}_\sigma$, then A(b) must be in $S_{1,\sigma}$ (since this contains precisely the true sentences of $\mathcal{L}_\sigma$).
- But it must also be in D–$S_{1,\sigma}$ (by our assumption that A(x) of precisely those things).
- So A(b) is not true in $\mathcal{L}_\sigma$.
- But then it is not in $S_{1,\sigma}$, and so A(b) *is* true in $\mathcal{L}_\sigma$. Contradiction.

Thus, to express this notion of untruth one must ascend to a metalanguage. And so, as Kripke puts it: the ghost of the Tarski hierarchy is still with us (714).

1.2. Exclusion Negation

Another way of thinking about this issue is this.

In classical semantics, where predicates F(x) are only true or false of things, there seems to be only one natural way of understanding ¬F(x): as being true precisely of the things that F(x) is false of, and false precisely of the things that F(x) is true of.

But once we move to partial interpretations, and have true/false/undefined, there seems to be more than one way of understanding negation.

One way is the strong Kleene way: ¬F(x) is true of everything F(x) is false of, false of everything F(x) is true of, and undefined of everything else.

But another natural way of understanding negation seems to be this: ¬F(x) is true of everything F(x) is not true of (i.e., everything it is false or undefined of), and false of everything that F(x) is true of.

This sort of negation tends to be called *exclusion* negation.

Using ~ for this sort of negation:

Truth-tables

P	¬P
t	f
f	t
u	u

P	~P
t	f
f	t
u	t

Kripke's construction will not work at all if the language contains exclusion negation.

- Since at the first stage T is given partial interpretation (Ø,Ø).
- But now consider a sentence ~T(a), where a denotes this sentence.
- Under this first interpretation, ~T(a) is true.
- But then—following the basic idea of the construction—we'd have ~T(a) going into the extension of T at the second stage.
- But then ~T(a) will no longer true—so then it will be put into the *anti*-extension of T at the third stage.
- And so on: ~T(a) will forever flip between the extension and anti-extension of T, and one will never reach a fixed point.

To relate this to the way we were talking about things last week: if we have exclusion negation in language, then the operation $\phi(S_1,S_2)$, that takes a partial interpretation to another (S_1',S_2'), where S_1' (S_2') is the set of things true (false) when T is interpreted by (S_1,S_2), is not monotone.

(I.e., we do not have: $(R_1,R_2) \leq (S_1,S_2) \Rightarrow \phi(R_1,R_2) \leq \phi(S_1,S_2)$.)

Counterexample: (Ø,Ø) and ({~T(a)},Ø).

The moral might seem to be that while Kripke gives an account on which languages can contain their own truth predicates, this approach allows languages to contain their own truth predicates only if their logical resources are restricted.

If one has exclusion negation in the language, perhaps one can only talk about truth in a metalanguage?

This is of course a somewhat disappointing result! Is there any way of challenging it?

Perhaps one could try to argue that exclusion negation is incoherent? Maybe. But that seems a pretty tough sell!

2. Propositions

Kripke suggests in a footnote that sentences that are neither true nor false fail to express propositions.

Thus, if the world doesn't cooperate, then Nixon and Dean's Watergate utterances might fail to express propositions.

The motivation for this view is that it allows Kripke to in some sense vindicate classical logic.

Thus, there will be classical logical truths that come out as neither true nor false on the construction we are considering. E.g., $T(c) \vee \neg T(c)$ where c denotes $\neg T(c)$.

However, if we say that such sentences simply fail to express propositions then classical logic would hold in the sense that any classical logical truth that expresses a proposition will be true.

Perhaps one could also hope to hold that classical logic holds at the 'level of propositions'?

But how exactly to think about this view?

3. Supervaluationism

An alternative, that might also be motivated by a desire to in some sense preserve classical logic, is to move to some version of the supervaluationist scheme.

A is true (false) in $\mathcal{L}(S_1,S_2)$ under the *supervaluationist* scheme if
A is true (false) in every classical extension of $\mathcal{L}(S_1,S_2)$.

I.e., A is true (false) in every $\mathcal{L}(X,D–X)$ where $S_1 \subseteq X$ and $S_2 \subseteq D–X$.

One can construct a fixed point as before. But now any classical logical truth will come out as true.

A 'feature' of the supervaluationist scheme is that it is non-compositional: suppose $A \notin S_1 \cup S_2$; then T('A') will be neither true nor false in $\mathcal{L}(S_1,S_2)$ under this scheme; as will be T('A') $\vee$ T('A'); but T('A') $\vee \neg$T('A') will be true.