Field on truth: how complex is *too* complex?

Peter Smith

June 30, 2010

In a reading group, we've been working through the first three parts of Field's *Saving Truth from Paradox*, by the end of which he has presented his core proposals. At this point, we've now rather lost the will to continue – for this is an astonishingly badly written book, which makes ridiculous demands on the patience of even a sympathetic reader.

It so happened that it fell to me to introduce the last two chapters in Part III, Ch. 17 in which Field rounds out his key technical construction, and Ch. 18, 'What Has Been Done'. I talked mainly about the latter. Here, for what they are worth, are some very quickly written reflections on what has and hasn't been achieved.

1 Field's game, very schematically indeed

Some preliminary general reminders.

If you are a strong deflationist about truth, then you are going to hold that recognizing the general equivalence of 'True($\langle A\rangle$)' and A is essentially all it takes to grasp the notion of truth (here '$\langle A\rangle$' is a suitable designator for A). But familiarly, the combination of (a) using a language in which modicum of self-reference is possible, (b) having a 'naive' truth predicate where, *universally*, we can infer 'True($\langle A\rangle$)' from A and vice versa, and (c) using classical logic, leads to contradiction. Either you are going to have to qualify the universality of the True($\langle A\rangle$)/A equivalence or modify classical logic. Field gives short shrift to the first option.

Question: can we modify classical logic so as to keep naive truth, and avoid paradox, without unduly hobbling our inferential apparatus? Field thinks we can.

Suppose we start with a language L equipped with the usual classical connectives. L lacks a truth-predicate, but is rich enough to express elementary arithmetic, so – via Gödel coding – it can express its own syntax and self-referential constructions can be executed. Let $\langle A\rangle$ be the Gödel number that codes the sentence A.

Now add to L both (i) a putative truth-predicate Tr (applying of course to Gödel numbers) and (ii) a binary propositional connective '$\longrightarrow$' (and an associated two-way connective '$\longleftrightarrow$'), to get the language L^+.

Then Field's project is to give a procedure for extending any classical two-valued interpretation M of the base language to an interpretation M^+ of the extended language over a lattice of values (with bottom 0 and top 1) such that:

1. M^+ agrees with M in still assigning 0 or 1 as before to all sentences of $L \subset L^+$.

2. $\longrightarrow$ is a 'reasonable' conditional.

3. On any M^+, and for any A of the extended language, '$Tr(\langle A\rangle) \longleftrightarrow A$' has the value 1, so Tr does indeed behave like a naively disquotational truth-predicate.

4. Suppose B is a L^+-sentence, A is a subsentence of B, and B' comes from B by substituting $Tr(\langle A \rangle)$ for A. Then M^+ gives B' the same value as B.

Since our language L^+ inherits the capacity for self-reference, we can form a liar sentence Q in L^+ such that Q is equivalent to $\neg Tr(\langle Q \rangle)$. So we also have

5. There is a Q such that on any M^+, '$\neg Tr(\langle Q \rangle) \longleftrightarrow Q$' has the value 1.

Trouble now threatens, of course, since '$Tr(\langle Q \rangle) \longleftrightarrow Q$' also always has the value 1! So can there possibly be such a procedure for building interpretations M^+ which doesn't collapse into disaster? Field shows that the answer is 'yes': we can do the job by ensuring inter alia that M^+ does not always make $A \vee \neg A$ take value 1, and in particular by ensuring that $Q \vee \neg Q$ doesn't take value 1. With excluded middle not available, the argument goes, you can after all unproblematically have both '$Tr(\langle Q \rangle) \longleftrightarrow Q$' and '$\neg Tr(\langle Q \rangle) \longleftrightarrow Q$' taking value 1.

On the face of it, then, we can have our cake and eat it: we can have a language L^+ capable of self-reference which contains its own truth-predicate for which a naive theory of truth holds. Which looks very clever indeed.[1]

But are appearances in fact deceptive? Does Field's procedure for constructing M^+ actually deliver the advertised goods?

2 A reasonable conditional?

The connective $\longrightarrow$, governed by Field's semantics (yet to be explained) is advertised as a 'reasonable' conditional.

But it is quite unclear what the criterion of reasonableness is. After all, someone could say that, since Field sets up $A \longrightarrow B$ to collapse into the material conditional when A and B are free of occurrences of the truth-predicate, that already damns $\longrightarrow$, since the material conditional (notoriously) isn't a reasonable conditional.

Well, that of course would be a rather captious contribution to the current debate! – but it does point up the problem. If 'reasonable' doesn't mean 'attractive to untutored intuition', what *does* it mean here?

Well, here are a few results concerning $\longrightarrow$ given Field's semantics, as acknowledged by Field himself.

1. The following inferences are *not* validated by Field's semantics:

   (a) $A \longrightarrow (B \longrightarrow C) \quad \therefore \quad B \longrightarrow (A \longrightarrow C)$

   (b) $A \longrightarrow (B \longrightarrow C) \quad \therefore \quad (A \wedge B) \longrightarrow C$

   (c) $(A \wedge B) \longrightarrow C \quad \therefore \quad A \longrightarrow (B \longrightarrow C)$

   (d) $A \longrightarrow (A \longrightarrow C) \quad \therefore \quad A \longrightarrow C$

2. The 'Conditional proof' introduction rule fails: i.e. we don't have the meta-rule that if $\Gamma, A \vDash B$, then $\Gamma \vDash A \longrightarrow B$ (where $\vDash$ of course indicates semantic entailment according to Field's semantics).

Now, to many, in the light of such results, it won't seem *un*reasonable to complain along the following lines.

[1] And there's more: we can use the conditional $\longrightarrow$ to construct something like a determinacy operator within L^+, along the lines of $D(A) =_{\text{def}} A \wedge \neg(A \longrightarrow \neg A)$. But I'm not going to discuss that.

> When in our deflationist innocence, we hanker after a naive theory of truth where '$Tr(\langle A\rangle)$ if and only if A' always holds, those are ordinary conditionals that are supposed to be in play, surely (at least) as strong as the material conditional.
>
> And the failure of (a) to (d), and the fact that we don't even have the canonical introduction rule governing our new connective shows that – whatever else it is – Field's long-arrow connective isn't an ordinary conditional, and is indeed even weaker than material.
>
> So trading in the naively desired biconditionals '$Tr(\langle A\rangle)$ if and only if A' for their simulacra '$Tr(\langle A\rangle) \longleftrightarrow A$' is, when the wraps are off, changing the subject. Having a theory which preserves the latter just isn't having a theory which preserves the former.
>
> We knew all along we couldn't have all our cake and it eat it. What Field gives up on is preserving a genuine version of the naive theory of truth: he, like the rest of us, has to settle for less than he'd ideally like.

The worry here is an absolutely obvious one. And Field has obviously met it before. His response in Ch. 18 is:

> I have heard it charged that the claim to have validated the Tarski biconditionals is an illusion, since the conditional employed "differs in meaning from" the conditional that Tarski employed. The charge seems to me to have no bite, given that in the presence of excluded middle the conditional behaves just like classical logic says a conditional should behave.

But that's quite extraordinary. For a start there a lot of lattice-valued connectives which, in the presence of excluded middle, behave like the material conditional and which differ among themselves: do they all mean the same? Or as Luca Incurvati invited us to do in discussion, compare the claim that the intuitionist conditional doesn't differ from the material conditional because in the presence of excluded middle the conditional behaves just like classical logic says a conditional should behave. Field's response, then, seems just not to meet the charge.

3 Double standards?

Here's another rather remarkable passage in Field's discussion of his quasi-conditional. He writes (§17.4):

> Among the principles that will fail in any theory like this are versions of induction stated as conditionals; e.g.
>
> $$\forall n[(\forall m < n)A(m) \longrightarrow A(n)] \longrightarrow \forall n A(n)$$
>
> ... In practice, it is the rule forms of induction ... that we employ, so there is little impact of the failure of the conditional form.

Extraordinary! For remember, at the outset, Field berates Kripke because Kripke's theory of truth, while allowing the *rule* form of the naive theory – we can legitimately infer A from $Tr(\langle A\rangle)$ and vice versa – doesn't have a 'reasonable' (bi)conditional that gives Kripke all instances of the naive theory in *conditional* form $Tr(\langle A\rangle) \equiv A$.

But if Field is allowed, when it suits him, to say 'Don't worry about the absence of a conditional form of induction: a rule form is good enough and all I need to preserve', why isn't Kripke allowed to say 'Don't worry about the absence of a conditional form of naive truth: a rule form is good enough and all I need to preserve'?

Field's bifurcating attitudes here just seem to involve unprincipled double standards: I can't think of any attractive way to rationalize them.

4 Logical entailment

Two obvious questions about the notion of logical validity:

1. Field is insistent that we don't identify 'real' truth with taking value 1 in some appropriate construction M^+. For a start, model constructions all have set-sized domains and the real universe isn't set-sized. But then why should we accept that preserving-value-1 in any set-sized M^+ is enough to make for a genuine valid entailment?

2. More generally, given that Field's reasoning in exploring his construction is all classical, why should someone who accepts a non-classical logic (of the kind Field wants to defend) believe the results of the Field's enquiries about what does or doesn't follow from what?

Ad q.1: Field originally mentioned Kriesel's squeezing argument as a way of bridging the gap between a model-theoretic construction and 'real' validity. But Kreisel's argument depends on the availability of a completeness proof for an axiomatized logic. And as Field acknowledges (p. 277), Philip Welch has shown that Field's logic isn't appropriately axiomatizable.

Field's response is amazingly insouciant.

> It might be better to adopt the view that what is validated by a given version of the formal semantics outruns "real validity": that the genuine logical validities are some effectively generable subset of those inferences that preserve value 1 in the given semantics. If one adopts this viewpoint then there would doubtless be some arbitrariness in which effectively generable subset to choose, but that seems perfectly acceptable unless one wants to put high (and I think unreasonable) demands on the significance of the distinction between those inferences that are logically valid and those that aren't.

So: Field's semantics warrants inferences that aren't really valid? How do we tell when it overshoots? Well, we aren't to care because the valid/invalid distinction shouldn't be over-rated! (And just *why* shouldn't we over-rate this, while – for example – we should put the top priority on having conditionals rather than just rules in our naive theory of truth? Once more, Field seems to make judgements of what matters and what we can drop or relax, without any obvious deep rationale.)

Ad q.2: Back in §5.6, Field has already touched on this question. He wrote:

> It might seem (a) that the non-classical logician can't accept his or her own explanation of logical validity, since it is given in a classical theory; and (b) that this is a serious embarrassment.

Field responds to (a) by saying that he 'sees no strong reason for rejecting excluded middle for validity claims'. But isn't this getting things exactly upside down? Remember, Field is trying to model, from within, something sufficiently close to the one language we have – if (on pain of paradox) this one language has to be non-classical, then so be it. We should then be, by default, non-classical through and through, unless we have special reason to suppose that excluded middle is in play: and then we surely need a special positive reason for supposing that we can internally define logical consequence using only classical models.

As a fall-back, Field says re (b) that 'the classical model theory might be used as a temporary device, useful for explaining the logic to the classical logician. The non-classical logician might hope to come up with a model theory in her non-classical analogue of set theory, or her non-classical property theory.' Ah: so is the non-classical logician (Field) still just living in hope, not having actually achieved the desired thoroughly non-classical theory of truth after all? Hannes Leitgeb has perhaps given us some handle on the issues here in his paper 'On the Metatheory of Field's "Solving the Paradoxes, Escaping Revenge"'. But as things stand the status of Field's enterprise remains quite murky.

5 Complexity without illumination?

My worries so far are about the seemingly unprincipled nature of Field's explorations. By his own lights, it is unclear how much of his story about validity even someone of his own spectrum of non-classical preferences should believe. And more basically, it is quite unclear why we should accept *that* spectrum of preferences. To repeat, *why* the fierce insistence that a rule-based naive theory isn't enough, combined with a cheerful willingness to so far reduce the strength of the connective that occurs in the required propositional naive theory that it is arguably not a genuine conditional any more?

Still, we could perhaps live with all such initial worries, if the core idea in Field's model construction was elegant and insightful – something that we could work with and develop in a more principled way than he does.

Compare Kripke's now classic paper. There we encounter a quite beautiful idea entirely rooted in ordinary thoughts about how we assign truth-values to propositions involving the truth-predicate and then, of course, developed with a wonderful elegance. I'm not saying that the resulting story is right: but it is compellingly *attractive*. (I remember well the excitement of reading the paper when it very first came out: and also the reaction 'that's so natural! – why didn't someone think of this twenty, thirty years ago?')

Field's story however just looks like a kludge. You don't like the original Kripke fixed point story? You don't like the revision theory of truth? Then – despite the different conceptual motivations of those theories – try mixing the technical tricks from both, alternating revisions (particularly in valuations of the quasi-conditionals) and then building fixed points. And do this not over a three-valued space of valuation, not over a continuum of values, but over a space of 'fine-grained' values which consists in *functions* from a large class of ordinals into $\{0, 1/2, 1\}$. Well, fine: *but what's the conceptual motivation for that*? The further details needn't detain us (as things don't get any more transparent), for the basic question surely arises already: what does this *kind* of complex mathematical trickery buy us by way of philosophical illumination?[2]

[2] And on the subject of complexity, the story isn't just messy, but complex in more technical sense. Suppose we take the ground-level, truth-free language, to be first-order arithmetic. Then Philip Welch has shown that Field's relative consistency proof of his theory requires much more that the predicative arithmetic that arguably suffices for applicable mathematics: even second order number theory with a Δ^1_3-Comprehension Scheme is insufficient.

Am I hankering after too much? Perhaps the aim is not illumination but just an existence proof: showing, by giving a model construction (however fanciful), that it is possible to extend a base language (with self-referential apparatus) with a truth-predicate Tr and a quasi-conditional, so that each Tr-quasi-biconditional comes out valid, and $Tr(\langle A \rangle)$ intersubstitutes for A, while allowing the quasi-conditional at least a modicum of conditional-like logical powers, and all this without falling into contradiction. But if doing this in the end isn't associated with a conceptually attractive story about the significance of the quasi-conditional, and an attractively motivated story about the route by which value assignments to propositions including the truth-predicate and/or the quasi-conditional are reached that echoes thoughts about how we ordinarily ground out assignments of truth, then just what do we now understand about truth that we didn't before?

Field's ill-written book left us baffled about such basics.