Truth, Logical Validity, and Determinateness: a commentary on Field's *Saving Truth from Paradox*

P. D. Welch

December 28, 2009

Abstract

We consider notions of *truth* and *logical validity* defined in various recent constructions of Hartry Field. We try to explicate his notion of *determinate truth* by clarifying the *path dependent hierarchies* of his *determinateness operator*.

Hartry Field in a recent series of papers ([2],[3],[5] in [1]) has indicated how we might approach a theory of truth and associated semantic paradoxes, by providing an inventive alternative to the current semantic theories of Kripke [10], Herzberger [9],[8], Gupta & Belnap [6], *et al.* In his recent book *'Saving Truth from Paradox'* ([4]), building on his earlier papers, he provides both a deep analysis of those currently available constructions of theories of truth, and an account of his own. The book is perhaps the culmination of his thinking and insights in this topic to date. It is a very rich and detailed tapestry that he weaves.

A thumbnail sketch of Field's project might run as follows: (I) the current semantic theories of truth are all lacking in one or more of the following desiderata: (a) they may provide good or interesting theories or semantic truth sets over a model in base languages involving, say, $\neg, \vee, \exists$ but there may not be any viable conditional $\supset$; (b) there may be occurrences of failure of the T-scheme (that for some sentence(s) A we do not have '$T(\ulcorner A \urcorner) \leftrightarrow A$') - where $\leftrightarrow$ is a usual material biequivalence; (c) that there may be failure of *the intersubstitutivity* of $T(\ulcorner A \urcorner)$ for A where the latter is a subformulae of some B. (II) We seek to remedy this by introducing a *binary operator* $\longrightarrow$ to function as a form of generalised conditional. (III) We wish to then provide a semantics that remedies (a)-(c) in a non-classical logic. An attractive feature that has emerged and that has taken on an increasing emphasis is (IV): We may analyse the defectivenenss of the Liar sentence which is made possible in the models produced to date, and which allows for the syntactic expression of what he calls *determinateness*.

That some form of non-classical logic must be employed is a concomitant of Tarski's theorem: otherwise inconsistency looms. In a number of papers Field has given an intricate construction of what he calls a 'G-solution' to this problem. One starts with a base model M with its appropriate language $\mathcal{L}$, typically say that of $\mathbb{N}$ the standard model of arithmetic, but this could also be some level of the ramified hierarchy, some V_κ the collection of all sets of rank less than the ordinal κ, or some such - any model let us say that has a sufficient coding apparatus; this model is then extended to a model M^+ in a larger language $\mathcal{L}^+$ where the latter has added a predicate T and binary operator $\longrightarrow$. The conservative construction of M^+, (in other words the extension of T and the valuation of sentences of $\mathcal{L}^+$ in some partial order) involves a process reminiscent of a *revision theoretic* approach, but in a multi-valued logic. This is done in [2]. Although there the logic used is 3-valued, Field prefers a De Morgan algebra valued semantics where values are taken in such an algebra where the latter has some cardinality greater than that of the ground model M. These semantic values are essentially functions on some ordinal into $\{0, \frac{1}{2}, 1\}$. That the cardinality of the algebra should be greater than that of the ground model seems inessential, and in any case for much of the discussion here it is enough to focus on the coarser grained 3-valued sub-order of $\{0, \frac{1}{2}, 1\}$ alone.

I'd like to take the discussion in two directions. Firstly I'd like to comment on the general framework of the enterprise, and then I'd like to comment on the notion of 'path-hierarchies' that form an essential part of his analysis of a determinateness operator and of any 'revenge immunity'. In order to pin this discussion down onto something concrete, this will rely heavily on what I shall call the principal model construction of Sect 16.2 of the book (that is also essentially that of [2]).

The argument of [4], as I understand it, is intended to take place in some non-classical framework where in particular the law of excluded middle is denied (Field convincingly shows how many of the standard paradoxes can be framed in such a way that a final conclusion of simutaneously deriving a contradiction in the form of obtaining both A and $\neg A$ depends on an application of excluded middle that he would deem illegitimate). He analyses the construction of Kripke using minimal fixed points for firstly the Strong Kleene scheme, and later a supervaluation scheme. If one considers the external logic of the former (often called KF Kripke-Feferman) as really a template for 'real truth' then unacceptable consequences follow.

Field prefers the 'internal logic' formed by abstracting from the sentences within the minimal fixed point itself and there is much to be said for this (and has also been argued by others, *cf.* [7]). However again there are deficiencies in regarding this theory as definitive of 'real truth', even before any attempt to define a conditional in the usual fashion as a material conditional from $\neg, \vee$. This latter attempt also fails. He would take issue with anyone, Kripke or otherwise, who claimed that the laws of 'validity' obtained from this construction, or from any axiomatisation of the internal theory *defined* 'real truth'. Indeed many would follow him there, and also through his analysis of supervaluational and revision-theoretic theories. It is hard to argue in any setting (non-classical or otherwise) for the full adequacy of these accounts: one cannot really claim that such approaches 'define real truth'. I think I agree with him that what Kripke and he have done (Field by providing by a classically defined G model), is to provide a model theory for sentences containing Tr. The G-model semantics, although only defined on set-sized models, then be construed as giving us a sense of semantic or designated value for the 'real world'.

I talk about 'real truth' (my term, not Field's) just to differentiate away from the extensional account of truth from the T-predicate in these various model constructions, on the one hand, and Field's talk of truth, or truth 'as it is', or truth 'in the world.' Just as the models of Kripke, Herzberger *et al.* when defined in a classical meta-theoretic way, provide standard extensional accounts (as a 'fixed point' or 'stable set' of sentences etc.) so Field's G-solutions when built over standard set models provide extensions for a T-predicate and semantic valuations of $\mathcal{L}^+$ in some algebra. However it seems that these are really only constructions demonstrating the consistency of the kind of thing he would like to build. It appears that he would like a theory of truth for the 'real (and essentially) non-classical world' and these model constructions are perhaps in some way only approximating to that by providing a science of 'designated' or 'semantic value.'

In the book, (but not necessarily in earlier papers) he assumes that $\mathcal{L}$ is the language of set theory, and that this is interpreted in a classical fashion - presumably over the universe V of all wellfounded sets of mathematical discourse. When intepreted over some V_κ or some other set-like universe, we may form, using a classical logic, the construction of [2] to obtain interpretations of T and $\longrightarrow$. What is left unspecified is precisely the notion of *(real) logical consequence* $\Gamma \models A$, that then obtains. There are plenty of laws given that this must satisfy, as well as discussion of laws that fail (the classical propositional law of *absorption* for example, as well as the 'Deduction Theorem'-like $\Gamma, B \models A \Rightarrow \Gamma \models B \longrightarrow A$). However we are not given a specific axiomatic theory of truth to characterize this, nor is the non-classical metatheory spelled out in which we are to form our notions of logical consequence. Much seems to be stated somewhat intensionally, and constructions that one might expect to be done meta-theoretically are simply performed 'within the real world'. Particularly in [5], there seems to be an elision between meta-theory and the full theory of truth (the latter for 'everything', rather than any one set-sized model).

This is of concern when considering set theoretic principles. In the G-solutions Field has given, say over some set V_κ, when we try and use some of the validities there as examples of real world logical truths or logical validities, then there are clearly set theoretic statements of intermediate truth value and indeed statements that can only be interpreted 'fuzzily'; there are notable examples, that indicate we must give up the *principle of induction for ordinals* when stated using the new binary operator $\longrightarrow$; indeed it would seem that even the *least number principle* for $\mathbb{N}$ when stated in the same way must go. (I am speeding past all the positive aspects here of the constructions: that induction for ordinals, the least number principle for $\mathbb{N}$, and so on, all remain perfectly valid for any statements or occasions when we are in a position to assert the law of excluded middle for the defining relations.) This is because it is simply 'fuzzy' whether we can assert there is a least ordinal element in a given 'fuzzily defined' class. Again we are to think of these fuzzy definitions *etc*. as taking place in the 'real world' again.

But how are we to make sense of this? How are we to understand the logic of this fuzziness? It seems clear that in the 'real world' the universe of classical sets is embedded in a penumbra of fuzzy classes or sets, which we are to handle in some non-classical manner. There are thus fuzzy subclasses of $\mathbb{N}$ for which it may be indeterminate whether any particular natural number k belongs. I think that we need an account of what this non-classical set theory should consist in. This not because of any particular qualms about set or number theoretic niceties, but it appears to go the heart of the enterprise.

However if 'real truth' is not about classically defined solutions over set sized models, we need a notion of logical consequence that is applicable to *all* (= in 'reality') defined models. Field (Sect 5.3) gives a list of six 'worries' one might have for non-classical approaches to the semantic paradoxes. I'll only mention the first. That is that '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 counters (a) by saying that he 'sees no strong reason for rejecting excluded middle for validity claims'. But the shoe is on the other foot: we have to have a reason, if we are really located in a non-classical 'reality', for defining logical consequence only *via* classical defined models. How can we assess Field's counter, if we cannot tell in which meta-theory, or 'reality' if you will, we are supposed to be working? He observes for (b), granting the worry of (a) for the moment: '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.' Indeed it is this account of non-classical set theory (in which presumably the model theory can be adumbrated) that is what we need. As we query as to what this non-classical set theory or property theory may be like as outlined in [3], Leitgeb points out ([11]p182) that Extensionality is a key issue.

If one argues that the semantics takes a variety of forms, that is classes of G-models in a variety of G-logics, we still need a general description of these classes in order to think about validities. The construction of Section 16.2 (which is that of [2]) yields one particular model (over the ground model for arithmetic for definiteness say). This may or may not be considered as constituting a G-logic in a class of its own; but small variations on this model, thus allowing a wider class, can have not inconsiderable differences on the validites obtained.

As an example of this, it is possible to keep the inductive arrangement of valuation assignment into the three semantic values, at zero, and successor stages, but reformulate the limit rule. To say precisely what that is at this point is to become rather technical, but the idea is that we keep with taking Kleene Strong Fixed points at successor stages; we keep the resetting of conditional assertions to $\frac{1}{2}$, but we take as assignments of values at a limit stage λ not a 'liminf' or 'Σ_2-rule' (which is Fields's rule CONT on p250) for all $\alpha < \lambda$, but instead we take a 'liminf' , or require 'continuity' only on an unbounded subset of λ of 'good' ordinals that somehow 'reflect well the process to date'. This yields then not Field's original model $M_0^+ = M^+$, but a 'longer' model M_1^+, with an acceptable ordinal of the generating process Δ_1 longer than the original Δ_0. If one allows now a G-logic to contain the class of models Γ obtained by generalising this continuity property further on the limit rule, one obtains a class of models constituting a perfectly acceptable (it seems to me) G-logic. In these models all the laws listed by Field continue to hold; the desiderata of (I)-(III) are maintained in the same way, there is a representation of determinateness, so (IV) is fine and determinateness satisfies the same given laws, and the models indeed look extremely similar. However if we were now to take the logical validities as the pairs $B_1 \wedge \cdots B_n \models A$ in which the semantic value of A should be a designated value (say '1') if $B_1 \wedge \cdots B_n$ does, in all the models of the class Γ, then we should lose one of the nice semantic features of any one model. Namely we now have that there is a formula $C(v_0)$ with $\models \exists n C(n)$ but for which there is no m witnessing that $\models C(\underline{m})$. And of course this cannot happen if 'validity' means the weak validity as given in Field's model M_0^+ (or indeed of any of the M_n^+): all of these individually validate the ω-rule.

In the book a somewhat different stance from [5] is taken. Here we are not taking validity as a process that allows us to assert $B_1 \wedge \cdots B_n \models A$ if the semantic value of A has a designated value '1' if $B_1 \wedge \cdots B_n$ does in a G-logic or in all G-models. As above, this would require us to say what a G-model is. Instead (p276) he discusses '*how many* of the inferences that preserve value 1 really ought to be declared logically valid *tout court* (as opposed to say validated by the formal sematics)' (his italics). He then says that if they all should, then one will have to make a decision on some seemingly arbitrary features of the system (the starting valuation for conditionals etc). He notes that it has been shown ([13]) that the logically valid inferences of the construction of [2] have a high degree of computational complexity. He then states

> "it might be better to adopt the view that what is validated by a given version of the formal semantics [*i.e.* a G-model] outruns "real validity": that the genuine logical validities are some *effectively generable subset of those inferences* that preserve value 1"

(italics now mine). He continues:

> "... there would doubtless be some arbitrariness in which effectively generable subset to choose, but that is perfectly acceptable unless one wants to put high (and I think unreasonable) demands on the significance of the distinction between those inferences that are valid and those that are not."

He correctly points out that whatever effectively generable subset one picks out, the G-model semantics of Section 16.2 offer a proof that one will not get into trouble operating with that set of inferences. However it is peculiar to be offering a semantic theory but no prescription as to what the logically valid sentences are, or, if they really are to be an effectively generable proper subset of those in a G-model, no indication of what they could be. I am not sure what to make of the last quotation.

In any case, real truth is now in [4] for Field seemingly far from logical validity, or having semantic value 1 in G-models or indeed any model ("the only account of Truth a non-classical logician should offer consists in the instances of the Tarski schema; or better a more powerful theory ... that implies those instances and certain generalisations ..." (p110): "equating truth with semantic value 1 throws away the virtues of the non-classical approach"). However we are left with the puzzle as to how the 'real truths' in this non-classical reality relate to the 'logical truths' derived from some (or any) particular model. Could the former be a subset of the latter? But let us take baby steps first and try and pin down, in the language not of set theory, but that of $\mathbb{N}$, an explication of how the theory of real truth about $\mathbb{N}$ relates to 'real validities' about $\mathbb{N}$. For this we need to know about how the fuzzily defined models over $\mathbb{N}$ arise. This seems to be a natural demand, not a 'higher demand' on a theory of truth that will seek to claim revenge immunity.

The last displayed quotation above seems to indicate a degree of indifference to logical validity. The discussion of the possibility of non-bivalent validity claims in Ch. 20.5 following, allows the possibility that our "tutored validity' claims (whatever they are) might not coincide with model theoretic validity. Field argues that, notwithstanding the very natural principles that we are forced to give up in order to make the naive theory of truth consistent with a $\longrightarrow$ operator, that tutored validity judgements are "still possible for anyone who has played around with the logic for a bit." However Kreisel's squeezing argument that shows that the intuitive notion of validity coincides with the technical one of logical validity using models, would be problematic in this context. Indeed for Kreisel we need a completeness theorem, which as Field remarks, [13] shows is not possible here.

The book is full of achievements but when I reflect I find it somewhat difficult to delineate them precisely. We do not seem to have a definite view as to what the logical truths or validities are, we do not seem to have a definite extensional notion of 'real truths', nor do we have any axiomatisation of such a theory or a characteristic part of a putative theory. On the other hand what we do have is a model. Moreover a model rich in structure with an elegant collection of validities that serves to show that Field's plan of showing the desiderata (I)-(III) is consistently available. Out of the model springs forth (IV). I should like to spend the rest this paper discussing it and its variants.

We could instead argue here that the G-model of Section 16.2 is *the* model that we should focus our attention on. *This* is the model that provides us with the most sophisticated and successful description so far of a model in $\mathcal{L}^+$ satisfying all the requirements of (I)-(III) with the additonal bonus of (IV). I think there is a lot to be said for that: the model is eminently successful with many pleasing features. However it is a model built as a conservative extension of a set ground model in $\mathcal{L}$. And being defined over a set it does not provide the G-logic of truth 'as it really is' - but again only approximations. If there were some axiomatisation of Field's theory, or some indication as to which 'effectively generable' subset of the validities obtained from the model then we should have something more concrete to go on.

An advantage with the Kripkean Strong Kleene minimal fixed point model is that the model itself (in the words of Martin) 'flows naturally' from the conceptual viewpoint of the logic and our grasp of the connectives (well at least $\neg, \vee$) and quantifiers. The principal construction of 16.2, is a hybrid, involving as it does alternating Kripkean fixed points, and a revision rule at limits, and yields a model of great complexity. I'd like to emphasise this once more (although I am always perplexed to find that it seems to cut little ice with philosophers, but has an electrifying effect on mathematical logicians). To produce even the simplest model over the standard first order model of arithmetic, (as in [2]) one needs a substantial amount of second order number theory to get up and running: the degree of second order comprehension needed to produce the sequence and validity sets is way beyond 'predicative mathematics' that Feferman *et al.* would claim is needed for any application of mathematics; further it is beyond the current reach of any proof theoretic ordinal analysis (although nearer this border than for predicative mathematics). We should perhaps not be losing sight here of what that particular construction is supposed to achieve: a theory of *first order* truth for $\mathbb{N}$ in the extended language satisfying the desiderata (I)-(III). Unless there is a simpler 'consistency proof' of whatever the theory turns out to be, there will remain the feeling that defining first order truth over the natural numbers should not need the weapons of mass (set-theoretical) construction that are here being implicitly deployed.

Perhaps proponents of Field's approach will argue that this is just one particular construction, and does not, can not, define for us a theory of real world truth; hence we should not worry about the complexity of the construction when applied to just the model of arithmetic. Perhaps strong axioms are needed for defining truth for $\mathbb{N}$ after all? However let us continue in this vein a little while. In Sect. 16.1 Yablo's suggestion of creating fixed points, and Field's suggested replies to some of Yablo's own criticisms of his theory (in Sect 17.5), all involve a successor stage operation of extremely high complexity: instead of taking the minimal Kleene fixed point (or some such) the definitions require at every successor stage a universal quantification over *all* fixed points (satisfying certain conditions). I think such theories are not the way to go. The universal quantification is then a universal one over the whole real continuum. The history of revision theory serves as a lesson here. When people objected to how, say Gupta's or Herzberger's theory handled their favourite sentence, the limit rules were first tweaked, and then finally all control was abandoned. Belnap suggested that in fairness all limit rules are equally good, so let us quantify over them all. The result then is a definition that produces validity sets over $\mathbb{N}$ that are complete Π^1_2 sets, and the same will happen here with such a universal quantification. Moreover once one allows such Ptolemaic validity sets it is then not hard to ask questions about them that run into independence phenomena in ZF set theory. (Again, to reiterate, in this part of the discussion, we are interested *only* in truth for $\mathbb{N}$ in $\mathcal{L}^+$; do we really need the theory of analysis, or even ZF to tell us what the features of this theory are?) I have heard the counter-objection to the last, that "if the philosophical intuitions are there, then we should be allowed to follow those intuitions", and investigate. Well, yes, but what are the intuitions underlying the *liminf*/CONT rule for evaluation of semantic value at limit stages λ?

One might simply wish to say that the the principal model construction of 16.2, say, is just one model construction demonstrating the consistency of some (as yet) unspecified theory. It is admittedly complicated, but one might adopt the attitude that that the final formal theory, however it turns out, will have some much simpler consistency proof. I should like to take a closer look at the details of the construction. Again let us take the ground model $\mathcal{M}$ as $\mathbb{N}$, the standard model of the natural numbers, with its language $\mathcal{L} = \mathcal{L}_{\mathbb{N}}$, and the extension $\mathcal{L}^+$ and extended model M^+. I should like to make some comments in particular on the internal structure of the model, chiefly to try and clarify (if only to an audience of one, myself) Field's notion of *path independent hierarchy*. This notion seems crucial to any claim, or discussion of *revenge immunity*. The discussion here pertains to Section 22.2 ("*Transfinite Iterations made rigorous*"). Recall that we have a determinateness operator $D(A) \equiv A \wedge \neg(A \rightarrow \neg A)$. Field shows that defectiveness of the Liar sentence can be expressed. D is simply a syntactic operation, and can be repeated in an obvious fashion yielding $D^{n+1}(A) \equiv D(D^n A))$, and for stage ω we can cook up $D^{\omega}(A) \equiv \forall n \forall y(y = \ulcorner D^n(A) \urcorner \rightarrow T(y))$. It is clear what to do to define $D^{\omega+1}$ and start the process again. At 'small' limit ordinals similar *ad hoc* syntactic devices can be employed, and in [2] it is remarked that this can be done up to some, unspecified, limit ordinal λ_0. There $\lambda_0 < \omega_{1\,\mathrm{ck}}$ (the first non-recursive limit ordinal) was assumed. In a specific sense the operators D^α become stronger as α increases; there is parallel to this a sequence of strengthened liar sentences of the form $Q^\alpha \leftrightarrow \neg D^\alpha(Tr(\ulcorner Q^\alpha \urcorner))$. However it is possible to lift the restiction on λ_0 being a recursive ordinal less than $\omega_{1\,\mathrm{ck}}$ by the device of using Kleene's $\mathcal{O}$ notation system; using this one obtains wellordered segments of determinateness operators for any recursive ordinal. (This is mentioned in [5] footnote 54.) However in [4] and [5] the intention is to look for maximal possible "paths" of determinateness operators, defined first in bivalent parts of the language (where the Law of Excluded Middle is known to hold) and then even defined in non-bivalent fragments. There is quite a lot of discussion on the dependence of hierarchies on the possible paths chosen, and a difficult discussion on the definition of 'general path-independent hierarchy' and a theorem thereon. We believe that this can be clarified and would like to illustrate with the standard model case of $\mathbb{N}$ and M^+ above. We let Δ_0 be the least acceptable point. Field asserts (p255 Footnote 8) that any remotely adequate way of iterating D must break down a long way before we reach Δ_0. We shall now make the claim that for any $\xi < \Delta_0$ there is nevertheless an iterated hierarchy of length ξ. However there can be no bivalent hierarchy of length Δ_0. Field's discussion is in terms of path independent hierarchies of lengths ξ less than "the least ordinal not hereditarily definable in $\mathcal{L}^+$" which means for such ξ that both it and all its predecessors $\xi' < \xi$ must have $\mathcal{L}^+$-definitions. *Prima facie* in the language of arithmetic we have no obvious way of defining ordinals in $\mathcal{L}^+$ but of course we can define *wellorderings* isomorphic to von Neumann ordinals - and this is just what one does when talking about recursive wellorderings: they are Σ_1 definable over $\langle \mathbb{N}, +, \times, 0, S \rangle$ and can be found for any ordinal less than ω_1^{ck}. Kleene's $\mathcal{O}$ notation gives us another method of talking about ordinals.

The new idea is simply stated: one lets the *sentences of* $\mathcal{L}^+$ *themselves* stand in for ordinal notations. For a sentence A we may define $\rho(A)$ to be the least ordinal ρ (if it exists) so that the semantic value of A is stable from ρ onwards.

We may define in the language $\mathcal{L}^+$ a *prewellordering* $\preceq$ of sentences of stabilizing truth value: we set $P_\preceq(\ulcorner A \urcorner, \ulcorner B \urcorner)$ if and only if $\rho(A) \leq \rho(B)$, where $\ulcorner A \urcorner$ is an integer Gödel code for A. (It has to be shown that we can do this and that $P_\preceq$ is given by an $\mathcal{L}^+$ formula. We could do this just for sentences stabilizing on 1 or 'designated truth values' but we do this here for 0,1 only.) Letting $\|A\|$ be the ultimate semantic value of the sentence

A, in the model M^+, we then show:

Lemma 1 *There are formulae $P_{\preceq}(v_0, v_1)$, $P_{\prec}(v_0, v_1)$ in $\mathcal{L}^+$ so that for any sentences $A, B \in \mathcal{L}^+$, we have*

$$\begin{aligned} \|P_{\preceq}(\ulcorner A \urcorner, \ulcorner B \urcorner)\| &= 1 \ \textit{ iff } \rho(A)\downarrow, \rho(B)\downarrow \textit{ and } \rho(A) \leq \rho(B); \\ &= 0 \ \textit{ iff } \rho(A)\downarrow, \rho(B)\downarrow \textit{ and } \rho(A) > \rho(B); \\ &= \tfrac{1}{2} \ \textit{ otherwise.} \end{aligned}$$

(And similarly for the formula $P_{\prec}$.)

We abbreviate $A \preceq B$ for $P_{\preceq}(\ulcorner A \urcorner, \ulcorner B \urcorner) = 1$ *etc.* Clearly if this works then, if $\|A\| = 1$ (or 0) say, then $\{B : B \prec A\} = \{B : \|P_{\prec}(\ulcorner A \urcorner, \ulcorner B \urcorner)\| = 1\}$ is a prewellordering of order type some ordinal $\xi < \Delta_0$. (Meaning the wellorder of the equivalence classes of the sentences in this set under $\preceq$ has order type ξ. It is less than Δ_0 since, recall, that any sentence that stabilizes must do so by Δ_0 by the latter's definition.) We let Field$(\preceq)$ denote the set of sentences stabilizing on 0 or 1. Moreover:

Lemma 2 *For any $\xi < \Delta_0$ there is $A = A_\xi$ in Field$(\preceq)$ with the order type of $\{B|B \prec A\}$ equalling ξ.*

We now have the wherewithal to define hierarchies of iterated determinateness along initial segments of $\prec$ given by the sets $\{B : B \prec A\}$. We may define for *any* sentence C $D^C(A) \equiv \forall B \prec C \forall y(y = \ulcorner D^B(A) \urcorner \rightarrow T(y))$. For $C \in$ Field$(\preceq)$ this defines a 'genuine' determinateness hierarchy of length $\rho(C)$. However it is not a bivalent matter as to whether a general C is or is not in Field$(\preceq)$. (In other words Field$(\preceq)$ is not a crisp subclass of $\mathbb{N}$.) However if $C \in$ Field$(\preceq)$ then it can be shown that it is a bivalent matter whether a general B is $\prec$-below C or not. Consequently the expression "$\langle D^B(v_0)|B \prec C\rangle$ *forms a determinateness hierarchy*" is not in the classical part of the language $\mathcal{L}^+$ to which the Law of Excluded Middle holds. I believe that this gives a precise formulation to the idea that 'O is an iteration of D is 'fuzzy' ' in this context.

The '*ordinals hereditarily $\mathcal{L}^+$-definable*' are thus here the ordinals $\xi < \Delta_0$, which we define through our use of stabilizing sentences and the ordering $\preceq$. Although the latter has order type precisely Δ_0 (by Lemma 2) there is no sentence C stabilizing precisely at stage Δ_0. Thus the internally defined determinateness hierarchy breaks down, not fuzzily, but precisely, at Δ_0. There is no internally definable maximal hierarchy. *Externally* we see exactly what is going on, and could of course, define a hierarchy of length Δ_0 using the full field of the ordering $\prec$.

Could there nevertheless be a way of defining a hierarchy in the language $\mathcal{L}^+$ using some different formula P', or other device, of length greater than Δ_0? We can rule this out by the following that argues against there being any definable (pre)wellordering of order type $> \Delta_0$ (and *a fortiori* no (pre)wellordered determinateness hierarchy of length $> \Delta_0$.

Lemma 3 *Let $Q(v_0, v_1)$ be a formula of $\mathcal{L}^+$. Define $n \prec_Q m$ if $\|Q(n,m)\| = 1$. Then if $\prec_Q$ is a prewellordering,* $\mathrm{ot}(\prec_Q) \leq \Delta_0$.

There is a formula Q with $Q_\prec$ of two variables, defining an ordering $\prec_Q$ as above, and ordinals ξ with $\Delta_0 < \xi < \Sigma$ (where Σ is the next acceptable point above Δ_0) so that for some m $\{n : |n \prec_Q m|_\xi = 1\}$ is a prewellordering of order type $\delta > \Delta_0$ and

$|D^m(A) \equiv \forall n \prec_Q m \forall y(y = \ulcorner D^n(A) \urcorner \to T(y))$ *forms a determinateness hierarchy*$|_\xi = 1$.

Indeed the formula $P_\prec$ is itself an example of such. However this is only an evaluation at a non-acceptable point ξ, and the semantic value of such when evaluated at Δ_0 or Σ is quite different, as it must be by Lemma 3. (To be clear, in the above formula '$n \prec_Q m$' is also to be evaluated at stage ξ as $Q(n,m)$ is too.) Thus, viewing the construction of the model dynamically, there are longer hierarchies, prewellorders etc, but they are evanescent: they appear for a while in the revision process, but then disappear: Δ_0 is the sum total of all the hereditarily definable ordinals. It is the least 'fuzzy' ordinal in that it is the least ordinal which is not the length of a 'stabilized' or 'bivalently defined' wellordering.

It is not possible to *internally* define a hyper-determinateness operator D^{Δ_0} in a bivalent fashion, in this model. (One can only define non-bivalently defined wellorderings and hierarchies in the evanescent fashion mentioned above.) However, as alluded to above, it is possible to define operators D^{Δ_0} either externally, or perhaps more interestingly, internally inside different *longer models*. By changing the limit rule only, and keeping all else fixed, one can define models M_n^+ ($n < \omega$) with respective least acceptable ordinals $\Delta_n < \Delta_{n+1}$ which satisfy all the other desiderata (I)-(III). Inside M_n^+ the bivalently defined determinateness hierarchies stretch out now to Δ_n (thus we keep (IV) too); so in this sense we can define a hyperdeterminateness operator for M_n^+, but in M_{n+1}^+. However if we are to consider the logical validities as some "effectively generable" subset of the weak validities verified by a model M_0^+ as Field suggests, then we may have also to consider the possibilities of considering the weak validities of the other M_n^+, and this throws another parameter into the pot. (I have mentioned above the desirability of *not* taking an intersection over all these weak validities as n varies as a single G-logic.)

The discussion above has all been predicated under the assumption that the ground model $M = \mathbb{N}$ is the standard model of arithmetic. However we can realise other 'conservative extensions' in Field's sense, over other ground models, such as taking $M = V_\alpha$ for some level of the ramified hierarchy. The language now could contain constants for each $x \in V_\alpha$ and a model M^+ built up in the same fashion. The existence, and theory of, acceptable points and their cycling nature, and the semantic valuations in either the 3 valued algebra, or an appropriate de Morgan algebra can be effected. This envisages a language of the cardinality that of V_α, but the above comments on bivalent or non-bivalent prewellorderings obtain *mutatis mutandis.* The arguments for the existence of acceptable points and their calculation as some $\Delta_0(V_\alpha)$ are parallel to the principal construction. (The latter could be done either with the language containing names for each set in V_α directly, or by using the construction in Sect. 16.2 using satisfaction and 'true of' assignments s.)

(Parenthetically, [12] is supposed to continue waging this discussion by other means. Namely over the whole universe V of sets. There the idea had been that one could use a small fragment of the second order theory ZF^2 to effect such a construction over V, using plural quantification to finesse the requirement to quantify over new objects. Whether this is a coherent viewpoint may be up for discussion, but the point is that the amount of second order ZF^2 needed to construct the model over V is small, and should correspond to the amount of second order number theory to construct M^+ over the ground model $\mathbb{N}$. The 'class sized' wellorderings that arise when considering a language $\mathcal{L}_V$ containing names for all sets, then correspond to those of length less than the first acceptable point. This was supposed to be effected again in a classical world, so we may be believers in *Bivalent Impredicative Pluralese* of [5] p342, but we only need this for a very small fragment of second order theory. Just as Field states (Footnote 8 p109) "the metatheory needed to develop the metatheory of continuum valued semantics requires a rather meagre fragment of set theory that is not plausibly regarded as vague" we (or those who support second order methods), would say the fragment of second order logic needed to develop a G-solution over V is also meagre and might to them also not be regarded as vague. There seems to be also some divergence of viewpoint, with Field stating that the determinateness hierarchy again would have countable length, which would be the case if the language were countable. With $\mathcal{L}_V$ we may define class terms defining wellorders of length much larger than ON - it is simply that these wellorders are not *set-like*: they have initial segments that are not sets but proper classes. However such wellorders are regularly, if implicitly, countenanced by inner model theorists, and can well exceed the first "acceptable point or wellordering which one might name $\Delta_0(V)$.")

I want to turn to the possibilities for *axiomatising* a theory of the kind Field discusses. This is speculative. As already remarked above, the computational complexity of the weak validity set over M^+ is high and there is no completeness theorem for this G-logic. Is there not some recursive theory T in the language $\mathcal{L}^+$ which we can use to axiomatise the desired truth theory we are aiming for? This is a recurring question that has been asked concerning the stable truth set formed by taking a single Herzberger sequence starting from, say a recursive distribution of truth values. The Fieldian validity set over M^+, F, and the Herzberger stability set starting from some, or any, recursive distribution of truth values, H, are recursively isomorphic and thus have the same complexity ([13]). We commented a little in [14] on the difficulties of finding axiom sets for the Herzberger set and the same comments apply to F. That paper showed how one could give an alternative semantics for the Herzberger stability set using infinite *games* (with players playing sentences and orderings in an attempt to show that a particular starting sentence φ was, or was not, stable and thus in or out of H). A very similar game shows that F also has a game semantics representation. One point made in [14] is that these games are not open games (and thus over in finitely many moves for the player with a winning strategy) but of necessity must in general take infinitely many moves. However the fact that H (and so F) form a complete set in a certain so-called *Spector class* allows one to give an open game representation of F, (*i.e.* one that is over in finitely many steps for the player attempting to play into the open set). However this representation is not in the standard first order language but one in an extended language with an additional non-standard quantifier Q. Suggestively (and somewhat beautifully) the natural interpretation of in the Fieldian setting is as a *path-hierarchy quantifier*: $\mathsf{Q}x\varphi(x)$ holds iff for 'path-many' x $\varphi(x)$.

Definition 1 $\mathsf{Q}x\varphi(x)$ *iff* $\{n \in \mathbb{N} | \varphi(n)\} \supseteq \{\ulcorner B \urcorner \mid B \preceq A\}$ *for some sentence* $A \in \mathrm{Field}(\prec)$.

Speculatively this suggests a possible axiomatisation of a theory of truth *together with* determinateness satisfying the laws or properties Field has already given for T and D.

Now, one way this can be concretely done is by using axioms that incorporate the basic CONT axiom for the liminf rule, together with a heavy use of:

Lemma 4 *There is a formula '$V(v_0, v_1, v_2)$' of $\mathcal{L}^+$, so that:* $|A|_{\rho(B)} = i$ *iff* $\|V(A, i, B)\| = 1$.

In other words $V(A, i, B)$ expresses *internally to the model*, that the semantic value of A at stage $\rho(B)$ (if the latter is defined) is $i \in \{0, \frac{1}{2}, 1\}$. Consequently one can list axioms which explicitly write in the whole construction as to how CONT *etc.* are used, to fix the semantic values at stages.

That is building a lot in to the theory: in effect the whole construction of the specific model, and that is unsatisfactory, being far more than a few perspicuous axiom sets about *Tr* and D. However this looks like it would be an axiomatisation of the ultimate truth set F in the model M^+. Ideally we should want a theory of truth and determinateness that does *not* require such a complex semantic construction, and thus we should aim for an axiomatisation that addresses just truth and determinateness theoretic principles. Hence whether these latter comments are of any real philosophical value is arguable.

References

[1] J.C. Beall. *Revenge of the Liar*. Oxford University Press, 2007.

[2] H. Field. A revenge-immune solution to the semantic paradoxes. *Journal of Philosophical Logic*, 32(3):139–177, April 2003.

[3] H. Field. The consistency of the naive theory of properties. *Philosophical Quarterly*, 54:78–104, 2004.

[4] H. Field. *Saving Truth from Paradox*. Oxford University Press, 2008.

[5] H. Field. Solving the paradoxes, escaping revenge. In J.C. Beall, editor, *The Revenge of the Liar*. O.U.P., Oxford, 2008.

[6] A. Gupta and N. Belnap. *The revision theory of truth*. M.I.T. Press, Cambridge, 1993.

[7] V. Halbach and L. Horsten. Axiomatising Kripke's Theory of Truth. *Journal of Symbolic Logic*, 71(1):677–712, september 2006.

[8] H.G. Herzberger. Naive semantics and the Liar paradox. *Journal of Philosophy*, 79:479–497, 1982.

[9] H.G. Herzberger. Notes on naive semantics. *Journal of Philosophical Logic*, 11:61–102, 1982.

[10] S. Kripke. Outline of a theory of truth. *Journal of Philosophy*, 72:690–716, 1975.

[11] H. Leitgeb. On the metatheory of Field's "Solving the Paradoxes, escaping revenge". In J. C. Beall, editor, *Revenge of the Liar*, pages 159–183. C.U.P, 2007.

[12] A. Rayo and P. D. Welch. Field on Revenge. In J. C. Beall, editor, *Revenge of the Liar*, pages 234–249. C.U.P, 2007.

[13] P. D. Welch. Ultimate truth *vis à vis* stable truth. *Review of Symbolic Logic*, 1(1):126–142, June 2008.

[14] P.D. Welch. Games for truth. *Bulletin of Symbolic Logic*, 15(4):410–427, December 2009.