THE REVIEW OF SYMBOLIC LOGIC
Volume 13, Number 4, December 2020

SOME OBSERVATIONS ON TRUTH HIERARCHIES: A CORRECTION

PHILIP D. WELCH
School of Mathematics, University of Bristol

Abstract. A correction is needed to our paper: to the definition contained within the statement of Lemma 1.5 and thus arguments around it in §3.

§1. The need for a correction.

I am much indebted to Chris Scambler for pointing out (in private correspondence, but see now [3]) the existence of errors in the argument of the 'proof' of Lemma 1.5 of the paper [5], and this note tries to set matters aright. The lemma should have expressed the fact that there is an obvious *norm* (a partial injective map into the ordinals) given by assigning stabilization times to sentences in Field's model of ([1], [2]). The thought had been that the Fieldian set of ultimate truths over a starting model of the natural numbers is recursively isomorphic to a complete semidecidable set of integers, in the sense of infinite time Turing machine theory, and thus is a complete set of a particular Spector class. There is thus a *semidecidable norm* that can be put on the set, and as for infinite time Turing machines, this can be taken as the ordinal stage at which an integer in the complete set is seen to stably enter it. Corresponding to this in the Fieldian model, and more simply put, is the ordinal stage, $\rho(A)$, at which a sentence A with ultimate semantic value 1 say, ("$||A|| = 1$") settles down to have that value 1 thereafter. Lemma 1.5 was an attempt to express this norm within the Fieldian theory for our three valued setting, but was incorrectly stated. The consequent 'proof' was then nonsensical.

The interested reader will be presumed to have a copy of [5] to hand.

§2. Changes – determinateness hierarchies.

We change Lemma 1.5 to state the following

LEMMA 1.5. *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*

$$\begin{array}{rcl} \|P_{\preceq}(\ulcorner A \urcorner, \ulcorner B \urcorner)\| & = & 1 \textit{ iff } \rho(A)\downarrow \wedge [\,\rho(B)\uparrow \vee \rho(A) \leq \rho(B)\,]; \\ & \leq & \tfrac{1}{2} \textit{ otherwise.} \end{array}$$

(And similarly *mutatis mutandis* for the formula $P_{\prec}$.) The purpose of the two formulae is to define the relationships between sentences that eventually settle to some fixed semantic value and can therefore be used as a *notation system* for ordinals, much as Kleene's $\mathcal{O}$ is a

Received: October 23, 2018.
2010 *Mathematics Subject Classification*: 03B80, 03B60, 03D55, 03D80.
Key words and phrases: truth theory, hierarchies, revision theory.

doi:10.1017/S1755020319000042

notation system for the recursive ordinals. Using notations we can then define extended Liar sentences Q_B and also iterations of Field's determinateness operator D^B along a prewellordered path given by $P_{\preceq}$. Further analogies obtain: we may consider the sets 'internal' to Kripke's least strong Kleene fixed point model to be the hyperarithmetic sets of integers, namely those in $L_{\omega_1^{ck}}$ where ω_1^{ck} is the supremum of the ordinals thus notated by when sentences arrive in the least fixed point ('internal' because they are definable by sentences with the T-predicate which end up in the fixed point). So now for us, the similarly internal sets are those in L_ζ where ζ is again the supremum of the ordinals notated by sentences eventual stabilization, and is Field's least 'acceptable point'.

(The previous definitions of Lemma 1.5 did not correspond to the canonical expression of a norm; then we tried to shoe-horn what was a Π_2 formula X into expressing it. The version of $P_{\preceq}$ above when it has value 1 correctly expresses a norm. The formula X below now more credibly expresses the Σ_2 notion of stability. We take note of the fact that here stability can mean with stable value $\frac{1}{2}$.)

Proof of Lemma 1.5. Rather than make a list of changes we have rewritten the proof. We have seen (Lemma 3.6) that there is a single arithmetical formula Φ (in the language $\mathcal{L}_{\dot{F}}$ the language of arithmetic augmented by a symbol $\dot{F}$ to be interpreted as F_β etc.) that defines over any $\langle \mathbb{N}, F_\beta\rangle$ for $(\beta < \Sigma)$ a wellorder of type β together with the associated previous F-sets $\langle F_\alpha \mid \alpha < \beta\rangle$. In particular it means that many things that we might express in a first order way about the sequence $\langle F_\gamma \mid \gamma < \beta\rangle$, for example whether a particular sentence A is stably 0, is then translatable into $\mathcal{L}_{\dot{F}}$, that is, or is not, true in $\langle \mathbb{N}, F_\beta\rangle$. We shall use the

FACT. *The proof of Lemma 3.2(ii) above cited as [4] Lemma 2.2, in fact shows that if $\tau_\iota < \Sigma$ is an admissible ordinal that is a limit of such, then $T^2_{\tau_\iota}$ is uniformly in ι (1-1) reducible to F_ι.*

(In [4] this is proven under the stipulation that L_{τ_ι} is a model of Σ_1-Separation, but actually the proof only requires the weaker assumption just stated.) For the rest of this note let $\tau_\iota < \Sigma$ be such an ordinal.

Let $X(x)$ be: "$\exists\alpha\forall\gamma > \alpha |x|_\gamma = |x|_\alpha$." This expresses that $x \in \mathcal{L}^+$ has a stable semantic value. Considering the revision process run in L, this set-theoretical statement then happens to be true vacuously about any sentence $x \in \mathcal{L}^+$ in L_{τ_β} for any successor $\beta = \alpha + 1$, and may or may not be so for $Lim(\beta)$. By the Fact above, there is a (1-1) effective $G : \mathcal{L}^+ \to \mathcal{L}^+$ (effective meaning G is recursive in the gödel codes) so that:

$$\begin{aligned} X(x) \in T^2_{\tau_\iota} \quad &\longleftrightarrow \quad \langle \ulcorner \top \longrightarrow G(x)\urcorner, 1\rangle \in F_\iota \\ &\longleftrightarrow \quad |\top \longrightarrow G(x)|_{\iota,\Omega} = 1 \longleftrightarrow |Tr(\ulcorner \top \longrightarrow G(x)\urcorner)|_{\iota,\Omega} = 1. \end{aligned}$$

The range of G is a recursive set of formulae, so let $\chi_G : \mathbb{N} \to \mathbb{N}$ be the recursive function with $\chi_G(\ulcorner\sigma\urcorner) = \ulcorner \top \longrightarrow G(\sigma)\urcorner$. Then we have an $\mathcal{L}^+$-formula A_X so that

$$A_X(x) \equiv \exists v_0(v_0 = \chi_G(x) \wedge Tr(v_0)) \qquad (1)$$

expressing the stability of $X(k)$ (where "$v_0 = \chi_G(x)$" is replaced by its defining Σ_1-formula). We then have that $||A_X(x)|| = 1 \leftrightarrow \rho(x)\downarrow$.

Note that $\neg X(x)^{L_{\tau_\iota}}$ is the assertion of x's instability below the ι'th stage, and is a Π_2 sentence over L_{τ_ι}. If x is so unstable, then $|\top \longrightarrow G(x)|_\iota \neq 1$. Continuing with this discussion we may define the 'local rank' of a sentence x at the ι'th iteration: that least stage $\alpha_x < \iota$ (if it exists) such that $(\forall\alpha \geq \alpha_x |x|_\alpha = |x|_{\alpha_x})^{L_{\tau_\iota}}$. This then is '$\rho(x)$'$^{L_{\tau_\iota}}$. We shall thus write

$$\text{“}\alpha = \alpha_x \leftrightarrow (\forall \beta \geq \alpha |x|_\beta = |x|_\alpha) \wedge \forall \gamma < \alpha \exists \beta \in (\gamma, \alpha](|x|_\beta \neq |x|_\gamma)\text{”}$$

and say that “α_x exists” if there is some α witnessing the above condition. Let $\Upsilon_{\preceq}(x, y)$ abbreviate the following Σ_2-sentence of $\mathcal{L}_\in$ about x, y:

$$\exists \xi_0 \forall \xi > \xi_0 \left[(\alpha_x \textit{ exists and equals } \alpha_x^{L_{\xi_0}}) \wedge (\alpha_y \textit{ does not exist } \vee \alpha_x \leq \alpha_y) \right]^{L_\xi}.$$

Let $\Upsilon_{\prec}(x, y)$ abbreviate the analogous Σ_2-sentence about x, y for the strict $\prec$ by replacing $\leq$ by $<$ in the above.

Again there is a (1-1) effective $H_{\preceq} : (\mathcal{L}^+)^2 \to \mathcal{L}^+$ such that

$$\Upsilon_{\preceq}(x, y) \in T^2_\iota \longleftrightarrow \langle \ulcorner \top \longrightarrow H_{\preceq}(x, y) \urcorner, 1 \rangle \in F_\iota.$$

Define $\chi_{H_{\preceq}}(x, y)$ as $\ulcorner \top \longrightarrow H_{\preceq}(x, y) \urcorner$ from $H_{\preceq}(x, y)$ just as χ_G was defined from G; then define $P_{\preceq}(x, y)$ from $\chi_{H_{\preceq}}(x, y)$ just as A_X was defined from χ_G at (1) above. Similarly define $P_{\prec}(x, y)$ from $H_{\prec}(x, y)$,

Claim $P_{\preceq}$ and $P_{\prec}$ are as demanded by Lemma 1.5.

Proof: We just check the former.

$$\begin{aligned}
&|P_{\preceq}(x, y)|_\zeta = 1 \leftrightarrow |H_{\preceq}(x, y)|_\zeta = 1 \leftrightarrow \Upsilon_{\preceq}(x, y) \in T^2_\zeta \\
&\leftrightarrow L_\zeta \models \text{“} \exists \xi_0 \forall \xi > \xi_0 \left[(\alpha_x \textit{ exists and equals } \alpha_x^{L_{\xi_0}}) \wedge (\alpha_y \textit{ does not exist } \vee \alpha_x \leq \alpha_y) \right]^{L_\xi} \text{”} \\
&\leftrightarrow L_\zeta \models \text{“} (\rho(x) \downarrow \wedge (\rho(y) \uparrow \vee \rho(y) \geq \rho(x))\text{”} \\
&\leftrightarrow \qquad \rho(x) \downarrow \wedge (\rho(y) \uparrow \vee \rho(y) \geq \rho(x))
\end{aligned}$$

noting (i) that $(\alpha = \alpha_z \text{ exists})^{L_\zeta}$ iff $\rho(z)$ exists and equals α, and (ii) for the penultimate ‘$\leftrightarrow$’ that our formulation allows for unboundedly many $\xi < \zeta$ with α_y existing in L_ξ, without $\rho(y)$ being defined. (In the latter case with $\rho(y) \uparrow$ we shall also have that for unboundedly many $\xi < \zeta$, α_y will not exist in L_ξ, by reflection.) □Lemma 1.5

One should perhaps point out that we may have sentences B, C with $|P_{\preceq}(B, C)|_\iota = 1$ where local versions of both $\rho(B), \rho(C)$ are defined in the ι'th stage without necessarily $B \preceq C$. This could be for stages $\iota \neq \zeta$ where the stabilization is only apparent, being just local to this stage. Indeed we shall see later that there are $C \notin \textit{Field}(\preceq)$ so that for any $B \in \textit{Field}(\preceq)$ there are unboundedly many stages ι in ζ (and so in Σ) where $|P_{\preceq}(B, C)|_\iota = 1$.

We are eschewing the use of levels of L which are models of Σ_1-Separation, so the next proof needs to be adjusted at one or two points. We again write out the first part in some detail.

Proof of Lemma 1.6. It suffices to show that $\zeta_0 =_{\text{df}} \text{ot}(\prec) = \zeta$. Note first that $\zeta_0 \leq \zeta$ since by definition of $\Delta_0 = \zeta$ it is the least acceptable point, i.e., any sentence that is going to stabilize will do so by stage ζ. We show that $\zeta_0 \geq \zeta$. The reader who does not wish to go through the details below, may perhaps be satisfied with the following heuristic argument: we have asserted that T^2_ζ is (1-1) reducible to F_ζ *via* a recursive function G. One may show that the order type of the set theoretical Σ_2-sentences in T^2_ζ, preordered by the

https://doi.org/10.1017/S1755020319000042 Published online by Cambridge University Press

natural ranking of the ordinal stage in the L_α hierarchy (for $\alpha < \zeta$) at which they settle down to their ultimate truth value, is ζ. That prewellorder induces *via G* a prewellorder of stabilizing sentences of $\mathcal{L}^+$ that are coded in F_ζ. Hence ζ_0 will be no less than ζ.

By the reflection property that defines ζ as the least such that there is $\Sigma > \zeta$ with $L_\zeta \prec_{\Sigma_2} L_\Sigma$, one may easily show that ADM $=_{\rm df} \{\alpha < \zeta \mid \alpha$ is admissible$\}$ is unbounded in ζ and has order type ζ. Hence, letting $\mathrm{ADM}^+ = \mathrm{ADM} \cap \mathrm{ADM}^*$ be the set of limit points of ADM, which are themselves admissible, ADM^+ also has order type $\zeta+1$ (again by reflection, as $\zeta \in \mathrm{ADM}^+$). (For the rest of this proof we abbreviate ADM^+ by S^*. We also note without further remark that $\tau_\xi = \xi$ for any $\xi \in S^*$.)

The rest of the proof continues as before almost verbatim, using this S^* and thus applying the Fact above rather than appealing to Lemma 3.2(ii). □ *Lemma 1.6*

(The Corollary 3.7 would need slightly adjusting to take care of liminf's, but as it is not used later we can omit it.)

Scambler also emphasises the observation that to obtain a sentence whose indeterminacy escapes the Fieldian hierarchy, it is not necessary to proceed by diagonalizing to get an ineffable liar (this last move we just did to get it into the liar hierarchy). The sporadic sentences of [5] (which he also calls ineffables) already have this property. He further asks whether there can be such a sentence that is ineffable for all ground models, just as the simple liar is for the strong Kleene fixed point starting from any model. We can show there is a single sentence for all countable models; although for all models this seems possible, it remains open.

§3. Acknowledgment. I should like to reiterate my thanks to Chris Scambler for his helpful comments throughout our discussion of these matters, and now also on these corrections.

BIBLIOGRAPHY

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

[2] Field, H. (2008). *Saving Truth from Paradox*. Oxford University Press.

[3] Scambler, C. (2019). Ineffability and revenge. *Review of Symbolic Logic*, doi:10.1017/S1755020318000473.

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

[5] Welch, P. D. (2014). Some observations on truth hierarchies. *Review of Symbolic Logic*, **7**(1), 1–30.

SCHOOL OF MATHEMATICS
UNIVERSITY OF BRISTOL
BRISTOL BS8 1TW, UK
E-mail: p.welch@bristol.ac.uk