ON GUPTA-BELNAP REVISION THEORIES OF TRUTH, KRIPKEAN FIXED POINTS, AND THE NEXT STABLE SET.

P.D.WELCH

Abstract. We consider various concepts associated with the revision theory of truth of Gupta and Belnap. We categorize the notions definable using their *theory of circular definitions* as those notions universally definable over *the next stable set*. We give a simplified (in terms of definitional complexity) account of *varied revision sequences* - as a *generalised algorithmic theory of truth*. This enables something of a unification with the Kripkean theory of truth using supervaluation schemes.

§**1. Introduction.** The purpose of this note is to state some recent results concerning the theory of revision sequences and circular definitions derived from Gupta and Belnap's Revision Theory of Truth.

This theory of truth has sometimes been regarded as an alternative approach to the Kripkean theory of fixed points *via* monotone inductive operators, in that it is also a semantical attempt to describe how a language may contain its own, necessarily partially defined, truth predicate. Indeed there are several accounts of semantical approaches to this problem and that of the Tarskian Liar which broadly speaking split the semantical theories into these two camps (for example, [17],[16]).

The tenor of the results here is threefold.

(i) *Revision theories of truth are complicated.* Revision theories of truth (one should speak of theories as there are a class of theories based upon various technical choices to be made - just as in the Kripkean "theory" there are alternative choices of jump evaluation scheme) are complicated. They result in truth sets (defined below) of complexity at least the level of Π^1_2 or higher in the projective hierarchy.

(ii) *Definability issues:* Gupta and Belnap have produced an approach to the theory of *circular definitions*. There are some indications in the literature

Received by the editors September 4 2000.

Key words and phrases. revision theory, theories of truth, Kripkean fixed points.

We should like to express our gratitude to the Department of Mathematics at UCBerkeley for its hospitality during the Fall Semester 1999, which enabled the research outlined here to be undertaken.

concerning what kind of concepts are in general definable in this way, using revision theoretical semantical schemes, and some partial results about the scope of such definitions. We give an explicit approach, mirroring the theory of monotone inductive definitions and the theory of the "next admissible set" (*cf.* [2],[15]) that yields a reasonably complete account.

(iii) *A* rapprochement *with Kripke.* We propose an approach (realistic variance) to introducing *variance* into revision sequences that solves many of the puzzles arising in the revision theory of truth of certain intuitively "true" (or stable ...) sets of sentences being poorly classified. In so doing we are able to show that the stable sets arising in such sequences are none other than Kripkean fixed points for the supervaluation jump scheme. We believe that, in fact, one may give a revision theoretic account of truth (incorporating realistic variance) called, somewhat awkwardly, a "generalised algorithmic theory" below, that *prima facie* is very different from the Kripkean one, but which ultimately yields stable truth sets that are also Kripkean fixed points.

The results here are all of a technical nature (although some are of a "soft" variety) and so no proofs will be given. These will appear elsewhere [18]. There is also little discussion of the philosophical issues involved, or proposed. Again we hope to discuss the ramifications of these results elsewhere.

For an introduction to both the Kripkean theory and to revision theories the reader may refer to the accounts of [17],[16], or [14].

§2. Revision Theories of Truth and Definability.

Belnap and Gupta in their book develop a general theory of circular definitions. This arose out of their earlier theory of truth, (*cf.* [3],[7]) which construed the Tarskian Truth Biconditionals as being *definitional* of truth.

Briefly:

Let $\mathcal{L}$ be a first order language, and let $\mathcal{L}^+$ be the language with a possibly infinite set of new predicate symbols $\dot{G}_n(x_1, \ldots x_n)$. For each $\dot{G}_n$ there is a definition from the set of definitions $\mathcal{D}$ of the form

$$\dot{G}_n(x_1, \ldots x_n) =_{df} A_{G_n}(x_1, \ldots x_n).$$

The point is that A_G is a formula of $\mathcal{L}^+$ which may contain occurrences of $\dot{G}_n$, or of any other of the new symbols $\dot{G}_m$. (And which certainly need not be positive occurrences.) A revision process is invoked to say that the extension of the predicate(s) G_n at some moment in time is inserted on the right side, is revised and a new extension is found according to the rules the definitional equations provide. A particular case of the above is where the definitional equations are simply those of the Tarskian Biconditionals for Truth, where a predicate letter $\dot{T}$ alone has been introduced to $\mathcal{L}$ to form $\mathcal{L}'$. For such a process to get off the ground an initial hypothesis $h = h_{\langle G_n \rangle}$ is made for the extensions of G_n. This hypothesis is then revised in discrete stages resulting in a *revision sequence*

$\langle s_\alpha | \alpha < \infty \rangle$, of successive extensions of the predicates, where $s_0 = h$, and ∞ is officially at least the length of all the ordinals. The gap in this brief sketch is what to do at limit stages of this process. A number of views have been expressed (summarised in [8]), as the theory evolved, but in all cases some kind of "bootstrapping policy" or "limit rule", let us call it here Γ, is invoked to tell us how how to handle limits and to define the sequence of extensions s_λ for such limit λ. (We detail some of these below). We have been somewhat vague as to what formally s_α actually is - for arithmetical revision operators over $\mathbb{N}$ we may consider it as a set of integers - but more generally it is the sequence of current extensions at stage α of the definenda G_n.

Thus: in general $s_{\alpha+1} = \delta(s_\alpha)$ for some "revision operator" δ, but here in particular we take $\delta = \delta_{\mathcal{D}}$ for some set of definitions as above, and $s_\lambda = \Gamma(\langle s_\alpha | \alpha < \lambda \rangle)$ for some limit rule Γ. Given such a Γ, Revision Theory gives an account of both validity and definability for (more than one) semantic system based on Γ.

DEFINITION 2.1. (Validity in S^*_Γ) *([8], 5D.1) Let $\mathcal{L}$ be any first order language, M be an $\mathcal{L}$-model, and $\mathcal{L}^+ \supseteq \mathcal{L}$ contain predicate letters for new definienda G_n using equations in $\mathcal{D}$.*

(i) An $\mathcal{L}^+$ sentence σ is valid on $\mathcal{D}$ in M in the system S^*_Γ *(written $M \models^*_{\mathcal{D}} \sigma$) iff, for all initial hypotheses h, and all revision sequences $\vec{s}$ based on $\delta_{\mathcal{D}}$, Γ with $s_0 = h$, σ is* stably true *in ∞. That is, for all sufficiently large $\alpha < \infty$ $\langle M, s_\alpha \rangle \models \sigma$.*

(ii) σ is valid on $\mathcal{D}$ in S^*_Γ *(written $\models^{\mathcal{D}}_{*,\Gamma} \sigma$) iff for all models M of $\mathcal{L}^+$, $M \models^*_{\mathcal{D}} \sigma$.*

Actually a principal system elaborated in [8] is not S^* (with $\Gamma = \Gamma_B$) but a variant of the above. One requires not that a sentence σ be in all s_α from some point on, in every revision sequence, but only that for every revision sequence $\vec{s}$, there should be a finite number, n depending on $\vec{s}$, so that for any limit ordinal λ we should have $\sigma \in s_{\lambda+n}$. (We refer the reader to [8] for the motivating discussion on any of the definitions of this section.) For all practical purposes of this paper, the reader will lose little by considering only the S^* versions, as, for the most part, the proofs of the results given here for the two systems are at most minor variants. If we write S_Γ without any decoration this is to be read as either S^*_Γ or as $S^\#_\Gamma$. If Γ is omitted without qualification, let us agree to take it as Γ_B - the limit rule of [8].

For the reader's benefit we give a formal definition of $S^\#$ below mirroring 2.1, but thereafter shall not have nay reason to refer to the detail. (The definition strictly speaking, is not that of [8] 5D.1 but is a simpler equivalent (*cf* Theorem 5D.14)).

DEFINITION 2.2. *Let $\mathcal{L}, \mathcal{L}^+, M, \mathcal{D}, G_n$ be as in Definition 2.1.*

(i) An $\mathcal{L}^+$ sentence σ is valid on $\mathcal{D}$ in M in the system $S^\#_\Gamma$ *(written $M \models^\#_{\mathcal{D}} \sigma$) iff, for all revision sequences $\vec{s}$, σ is* almost stably true *in ∞. That is,*

$\exists\beta\forall\alpha \geq \beta\exists n < \omega$ *so that* $\forall p \geq n(p < \omega \longrightarrow \langle M, s_{\alpha+p} \rangle \models \sigma)$.

(ii) σ is valid on $\mathcal{D}$ in $S^{\#}_{\Gamma}$ *(written $\models^{\mathcal{D}}_{\#,\Gamma}\ \sigma$) iff for all models M of $\mathcal{L}^+$, $M \models^{\#}_{\mathcal{D}} \sigma$.*

The concomitant notion of *revision theoretic definability* (*cf.* [8] 5D.18) which we shall deal with in this section is as follows.

DEFINITION 2.3. (Definability in S^*_Γ) *Let $\mathcal{L}$, $\mathcal{L}^+ \supseteq \mathcal{L}$, Γ be as above.*

(i) A formula $\varphi(v_0)$ of $\mathcal{L}^+$ (weakly) defines a set $X \subseteq |M|$ in S^*_Γ *if:*

$x \in X$ iff, for all initial hypotheses h, and all revision sequences $\vec{s}$ with $s_0 = h$, $\varphi(x)$ is stably true in ∞. That is, for all sufficiently large α $\langle M, s_\alpha\rangle \models \varphi[x]$.

(ii) If additionally the following holds then we say that φ strongly defines *X:*

$x \notin X$ iff, for all initial hypotheses h, and all revision sequences $\vec{s}$ with $s_0 = h$, then $\varphi(x)$ is stably false in ∞. That is, for all sufficiently large α $\langle M, s_\alpha\rangle \models \neg\varphi[x]$

([8] 5D.18 actually defines the analogous version for $S^{\#}_{\Gamma}$ which we leave here for the reader.)

Antonelli ([1]) gives explicit revision theoretic sets of definitions $\mathcal{D}$ for each of the complete Σ^0_n sets, *i.e.* for each level of the arithmetic hierarchy. Kremer ([12] §8) gives an argument of Gupta showing that this result can be extended to the inductive sets:

(i) Inductively definable subsets of $\mathbb{N}$ are $S^{\#}$ and S^*-definable.

[8] considered at 5D.7 the question of providing an axiomatisation of $S^{\#}$. One can deduce a negative answer to this (well at least directly for S^*) from an earlier result of Burgess [4] who showed that the set of stable truths over $\mathbb{N}$ formed a complete Π^1_2 set. Kremer also solved this negatively as follows:

(ii) For any Π^1_2 set of integers X, there is a finite set of positive definitions $\mathcal{D} = \mathcal{D}_X$ so that X is recursively embeddable into $\models^{\mathcal{D}}_i$ - the set of sentences valid on $\mathcal{D}$ in S_i, where S_i is the semantic theory associated to truth-at-the-first-fixed-point of the positive inductive definitions $\mathcal{D}$.

As he shows that S^* and $S^{\#}$ extend S_i, one concludes that the complexity of $\models^{\mathcal{D}}_{S^*_\Gamma}$ as $\mathcal{D}$ varies, and of $\models^{\mathcal{D}}_{S^{\#}_\Gamma}$, are also at least Π^1_2.

We mention the particular cases of limit rules Γ of revision sequences that arose from discussions on Revision Theories of truth. (We shall henceforth simplify matters by not making great distinction - unless required - between the variants for the semantical scheme $S^{\#}$ or $S^{\#}$.)

Notation For $\vec{s} = \langle s_\alpha \mid \alpha < \infty\rangle$, we write the *stability set* or *pair* as $s_{<\infty} = (s^+_{<\infty}, s^-_{<\infty})$ where $d \in s^{+/-}_{<\infty}$ if, for all sufficiently large $\alpha < \infty$, $d \in s_\alpha$ (respectively $d \notin s_\alpha$). *Local stability pairs* $s_{<\lambda}$ are defined analogously with any limit ordinal λ replacing ∞ .

DEFINITION 2.4. *If $\vec{s} = \langle s_\alpha \mid \alpha < \infty\rangle$ is a sequence, then we say that s* coheres with *$\langle s_\alpha \mid \alpha < \lambda\rangle$ (for any limit $\lambda \leq \infty$) if $s \cap s^-_{<\lambda} = \emptyset$ and $s^+_{<\lambda} \subseteq s$.*

Example (1) Herzberger Limit Rule, Γ_H.

Here Γ_H is single valued and returns as s_λ $s^+_{<\lambda}$, the smallest s coherent with

$s_{<\lambda}$, for all $\lambda < lh(\vec{s})$. In some sense this is a "minimal" policy, only those objects locally "stably in" $s^+_{<\lambda}$ can be taken as in s_λ.
Example (2) Gupta Rule, Γ_G
is also single valued: Γ_G is defined by $\Gamma_G(\vec{s} \restriction \lambda) = s_\lambda = s^+_{<\lambda} \cup (s_0 \backslash s^-_{<\lambda})$. The idea here is that we refer back to our original "hypothesis" $h = s_0$ to fill in for the ambiguous values in $\mathbb{N}\backslash(s^+_{<\lambda} \cup s^-_{<\lambda})$.
Example (3) Belnap Rule, Γ_B.
This is an example of a multi-valued limit operator: s_λ may be chosen as any $s \in \Gamma_B(\vec{s} \restriction \lambda) =_{df} \{s \mid s \text{ coheres with } s_{<\lambda}\}$. The Belnap rule thus places the least possible restriction on the choice at limit stages. The motivation here is to "allow free and full play" of the Tarskian Biconditional definitions (*cf.* [4]) and not to artificially make up some limiting rule.

We consider first as a paradigm example that of arithmetic and the structure of natural numbers $\mathbb{N}$ augmented with a predicate symbol $\dot{s}$. Let $\varphi(v_0)$ be a formula with one free variable in this language. Let δ_φ be defined by

$$\delta_\varphi(s) = \{n \mid \langle N, s\rangle \models \varphi(\dot{n})\}.$$

The question arises (in [13]), what kinds of sets of natural numbers are S^*_Γ or $S^{\#}_\Gamma$ definable for various Γ as φ is allowed to vary?

We may show that the strongly definable sets extend throughout Δ^1_2 regardless of the choice of limit rule $\Gamma_H, \Gamma_G, \Gamma_B$ or of semantical scheme S^*_Γ or $S^{\#}_\Gamma$. In the following we allow any limit rule Γ which returns a value at a limit λ simply defined from the sequence $\vec{s} \restriction \lambda$. We express this by saying that s_λ should be definable in a Δ^1_2-way in a code for the wellordered sequence $\langle s_\alpha | \alpha < \lambda\rangle$. (The set of such codes forms a Π^1_1 set of reals, see for example, [11],§40.)

THEOREM 2.1. *Let S be $S^{\#}_\Gamma$ or S^*_Γ; let Γ be any Δ^1_2-definable (in the codes) limit rule (this includes any of $\Gamma_H, \Gamma_G, \Gamma_B$).*
The class of S_Γ-definable reals is precisely that of the Π^1_2 reals; the class of Δ^1_2-definable reals coincides with the class of S_Γ-, co-S_Γ-definable reals, i.e.*, with the strongly $S_\Gamma$definable reals.*

The upper bound here, that S_Γ-definable reals are all Π^1_2, was noted by Löwe in [13] for $\Gamma = \Gamma_B$. As a corollary, the proof of this yields a previous result of Burgess [4], that for the language for arithmetic with a partially defined T-predicate the truth set of those sentences stably true in all revision sequences (using the Tarskian Biconditionals to revise the extension of T), (the "categorical truths"), using $S^*_{\Gamma_B}$ form a complete Π^1_2 set.

Remark: 1 The S_Γ-, co-S_Γ-definable reals (for $\Gamma \in \Delta^1_2$, again *e.g.* for $\Gamma \in \{\Gamma_H, \Gamma_G, \Gamma_B\}$) are thus the reals of the first transitive *stable set*, $\mathbb{S}_\mathbb{N}$, over the structure $\mathbb{N}$. In fact, by Levy-Shoenfield, they are the reals of the smallest Σ^1_2-correct model of Δ^1_1-Comprehension. The last theorem but more especially its proof naturally leads one to the following analysis of revision theoretic definability over more general structures. Indeed the last theorem is thus really a special case of the one to follow.

Making use of an analogy with admissibility theory and the sets hyperelementary over a structure, we propose the view here of revision theoretic (RT) definability as building up for us the domain of $\mathbb{S}_{\mathcal{M}}$ over M for some first order $\mathcal{M} = \langle M, R_1, \ldots, R_k\rangle$ with sufficient coding apparatus. Just as the inductive/co-inductive sets over an acceptable M yield the hyperelementary sets of the "next admissible set" over M, so the RT-/ co-RT-definable sets yield the domain of the "next stable set over M". The notion of "sufficient coding apparatus" or "acceptability" is that of Moschovakis [15].

DEFINITION 2.5. *Let $\mathcal{M} = \langle M, R, \ldots\rangle$ be any structure. Let $\mathbb{S}_{\mathcal{M}}$ be $L_{\sigma_M}(\langle M, R, \ldots\rangle)$ be the first level of the relativised Gödel L-hierarchy built over M, using elements of M as* urelemente*, so that $\mathbb{S}_{\mathcal{M}} \prec_{\Sigma_1} \langle V, M, R, \ldots\rangle$.*

Thus $\mathbb{S}_{\mathcal{M}}$ is correct about Σ_1 facts true in V, the universe of all sets, of M. Note that as $\mathbb{S}_{\mathcal{M}}$ is an admissible structure, $\Delta_1(\mathbb{S}_{\mathcal{M}})$ subsets of M are in $\mathbb{S}_{\mathcal{M}}$.

THEOREM 2.2. *Let S_Γ be $S^{\#}_\Gamma$ or S^*_Γ; let $\mathcal{M}$ be a countable acceptable structure; let Γ be any $\Delta_1^{(HC,\in)}(\{\mathcal{M}\})$-definable limit rule (this includes any of $\Gamma_H, \Gamma_G, \Gamma_B$).*
The class of S_Γ-definable subsets of M is precisely that of the $\Pi_1(\mathbb{S}_{\mathcal{M}})$ sets; the class of $\Delta_1(\mathbb{S}_{\mathcal{M}})$-definable sets coincides with the class of S_Γ-, co-S_Γ-definable subsets of M, that is, again, with the strongly S_Γ-definable sets.

Remark: 2 $\mathbb{S}_{\mathbb{N}}$ has domain that of L_σ where σ is the first stable ordinal. (For information on the stable ordinals, see for example, [2].) Note that many first order structures $\mathcal{M}$ will have $|\mathbb{S}_{\mathcal{M}}|$ the same. For example $\mathbb{S}_{\mathbb{N}}$ will have the same domain as $\mathbb{S}_{\mathcal{A}}$ where $\mathcal{A}$ is the least β-model of analysis, and exactly the same class of sets of integers are revision theoretically definable over each structure. In our terms L_σ is a very large set. One might add the comment that it is the *process* of revision theoretic definition that constructs these sets: the underlying model plays almost no role.
Remark: 3 There are strengthenings of the last theorem where we weaken the acceptability requirement, to allow for structures $\mathcal{M}$ over which there is a strongly definable *coding scheme*. It is unknown whether weakly definable coding schemes suffice. (*Weakly acceptable* structures suffice for Moschovakis, but we have a quantifier switch here.)
Remark: 4 It is easy to ask questions about such truth sets which are independent of the axioms of ZF. Theorem 2.1 shows there is a natural mutual interpretation of the notion of strongly revision theoretically definable over $\mathbb{N}$, in the sense of 2.3 (ii), with that of "Δ^1_2-definable". The theory of the latter reducibility is known to be independent of ZF (*cf.* Friedman [6]).
Remark: 5 Similar considerations to that of the theorem show that if, for example, L is the language of Arithmetic (or any recursive language that has

acceptable models), then V_L^* (essentially the intersection of the stable truth sets over all models of signature that of the language L) and $V_L^{\#}$ (defined *mutatis mutandis*) have complexity also precisely that of a complete Π^1_2 set. (This answers Problem 32 of [12]).

§**3. Fully varied sequences.** Stronger, and as we shall see, much more complex semantical systems are afforded by Yaqūb sequences [20] and the *fully varied (fv)* sequences suggested by Belnap & Gupta [8], p.168) and discussed in Chapuis [5]. The authors are only considering sequences based on the revision function δ_τ - the evaluation of sentences containing a T-predicate, based on the Tarskian biconditionals. We generalise this to arbitrary operators, before specialising it again to consider any arithmetic revision operation δ_φ. Their motivations are to iron out certain ill-classifications or anomalous behaviour of limit rules. (An example of this is the Gupta puzzle variant of [8] 6C.10[1].) They impose a global restraint on the class of all revision sequences.

DEFINITION 3.1. *A revision sequence (based on a revision rule δ, and say limit rule Γ_B) $\vec{s} = \langle s_\sigma \mid \alpha < \infty \rangle$ is-* fully varied (fv) *if any extension r that is coherent with the whole sequence $\vec{s}$, has actually been applied as a limit rule cofinally in ∞.*

Remark: 6 In general then fv-sequences over countable structures must have length at least $\mathfrak{c} = 2^{\aleph_0}$, that of the continuum - at least *prima facie* - although it is easy to see by a Löwenheim-Skolem argument, that proper initial segments of sequences determine the set of stabilities; moreover for any fv-sequence $\vec{s}$ there is another fv-sequence $\vec{r}$ with the same set of stabilities, that in fact is determined by a countable initial segment of $\vec{r}$.

Note: If $\vec{s}$ is fv, then $\{\alpha < lh(\vec{s}) \mid s_\alpha = s^+_{<\infty}\}$ is cofinal in $lh(\vec{s})$. (Since $s^+_{<\infty}$ is coherent with $s_{<\infty}$!) Then the final set of stabilities of a fully varied sequence is destined to appear cofinally in the whole sequence.

Yaqūb has a different definition of revision sequence to enforce full variability of bootstrapping limit rules. His sequences are (at best) members of $H_{\mathfrak{c}^{++}}$. We omit his somewhat baroque definition. But it is a result of Chapuis and Gupta ([5] Theorem 3.1, proven for the Tarskian revision rule τ but which works in this

[1] We do not wish to discuss these examples at any great length, but to serve as motivation for the definitions here and of §5, this Gupta Puzzle shows that the intuitively correct classification fails to occur in some revision sequences. Consider the situation where persons A and B make the following statements: A says two things: S_1: "It is true that everything B says is true", S_2: "Not everything B says is true", whilst B says only S_3 "At most one thing A says is true". Intuitive reasoning argues that S_1 and S_3 should be allocated the truth value true, whilst S_2 should receive falsehood. Without an insistance on full variance Belnap and Gupta note (p.228, *op.cit.*) that some Belnap sequences do not stabilise on these intuitively argued values, because we always chose at limit stages an unfortunate evaluation that resulted continually in instability. This is a simple (and finite) example of a set of sentences that may be called "ill-classified" under the usual $S^{\#}$ scheme.

more general setting) that the stabilities of any given fv-sequence are exactly the stabilities of a Yaqūb sequence.

We consider the set of integers that are stably in all fv-sequences for revision operators derived from arithmetic definitions over $\mathbb{N}$. (We drop the $*$ or $\#$ decoration, as the distinction becomes superfluous when calculating these sets.) They are thus the weakly definable sets for this notion.

The notion of a set being (weakly or strongly) RT-definable is just that of 2.3, with the notion of revision sequence being strengthened to fully varied revision sequence, and we shall denote this as $\tilde{S}_\Gamma$ or simply $\tilde{S}$.

It looks, again *prima facie*, as if $\widetilde{S}$ is $\Pi_1(H_{\mathfrak{c}^+})$ definable. But it is simpler than that.

THEOREM 3.1. *The class of $\widetilde{S}_\Gamma$-definable reals is precisely that of the Π^1_3 reals; the class of Δ^1_3-definable reals coincides with the class of $\widetilde{S}_\Gamma$-, co-$\widetilde{S}_\Gamma$-definable reals.*

We have not bothered to list the variants obtained by letting Γ be other limit rules. However the very same class of definable sets will also result if we allow functions $\delta \in \Delta^1_2$ besides the arithmetic operators δ_φ.

THEOREM 3.2. *(i) For any operator δ, $\widetilde{S}(\delta)$ is a $\Pi^1_3(\delta)$ set of integers, where*

$$\widetilde{S}(\delta) =_{df} \bigcap\{s^+_{<\infty} \mid \vec{s} = \langle s_\alpha \mid \alpha < \infty\rangle \text{ is a fv revision sequence based on } \delta\}.$$

If we specialise the result to the Tarskian rule δ_τ for partially defined truth predicates we obtain:

COROLLARY 3.3. *Let $\widetilde{V}_{\mathbb{N}}$ be the truth set over the standard model of arithmetic, using $\widetilde{\boldsymbol{T}}$, the theory of truth for fully varied revision sequences. (That is*

$$\widetilde{V}_{\mathbb{N}} = \bigcap\{s^+_{<\infty} \mid \vec{s} = \langle s_\alpha \mid \alpha < \infty\rangle \text{ is a fv revision sequence based on } \delta_\tau\}.)$$

Then $\widetilde{V}_{\mathbb{N}}$ can be construed as a complete Π^1_3 set, and thus is $\widetilde{S}_\Gamma$-definable.

Again there are variants given by considering models other than $\mathbb{N}$ of arithmetic here.

§**4. Categoricity.** In [8] a theory is developed of how the notion "categorical in a language" can be treated in a similar fashion to truth. They wish to argue that, by doing so they can fend off the spectre of *Strong Liar Paradoxes*.

The authors consider an augmented language to that (here) of arithmetic, $\mathcal{L}'$, containing a new predicate symbol $\dot{K}$, in addition to $\dot{T}$ for truth, to be interpreted as the current hypothesis concerning the *categorical sentences* (that is: stably true over $\mathbb{N}$ using all revision sequences)[2]. The basic model of arithmetic

[2]Actually [8] call the categorical sentences those that receive the same truth value stably in all revision sequences. As σ is stably true in all revision sequences iff $\neg\sigma$ is stably false in all such, it makes little difference to the analysis here if we concentrate just on those stably true everywhere.

is enlarged to a model $\mathbb{N}^+ =_{df} \langle \mathbb{N}, T, K \rangle$ with the displayed predicates being the obvious interpretation. They wish to consider revisions of a hypothesis K, concerning now what will ultimately be the set of categorical sentences, in exactly the same way that revisions were used to create new approximations to truth. One thus takes a hypothesis concerning the categorical sentences, call it $K = k_0$, and uses this extension in the expanded model above, and finds the stably true sentences relative to the new model in the new language. One thus keeps the extension of $\dot{K}$ fixed as we perform the revision process on extensions of $\dot{T}$ until we have the stably true sentences with this language. The latter yields a new set of sentences as a revised hypothesis for $\psi_{\mathbb{N}}(k_0) = k_1$.

DEFINITION 4.1. *$\psi_{\mathbb{N}}(K) = \{n | n$ is the gn of a sentence of $\mathcal{L}'$ that is stably true under the revision process S^* for $\dot{T}$, over the expanded model $\mathbb{N}^+$ with $\dot{K}_{\mathbb{N}^+} = K\}$.*

They remark ([8], p.231) that "in a sense the semantics for $\dot{K}$ is at a higher level than that for $\dot{T}$." It involves the whole revision process for $\dot{T}$, including the concomitant quantification over all starting hypotheses for $\dot{T}$, being considered as a single successor step in the revision process for $\dot{K}$. It is perhaps thus unsurprising that the complexity of the resulting *stably categorical* set (those sentences that occur on a final segment of every On length of revisions under $\psi_{\mathbb{N}}$, for every possible choice of starting hypothesis K) of sentences, or that of the *almost stably categorical* set - that obtained by using the scheme $S^{\#}$, with Belnap's limit rule (the preferred definition of [8] 6D.9) is considerable.

THEOREM 4.1. *(i) The stably (and almost stably) categorical (over $\mathbb{N}$ and using any Δ^1_2 limit rule Γ) set of sentences form a Π^1_3 set.*
(ii) In Gödel's constructible universe L, the set obtained using the semantical scheme $S^{\#}$, is a complete Π^1_3 set.

As remarked above, membership questions about sets of integers at this level of complexity are, in general, not absolute between models of set theory. Note the above calculation is based on the original revision theory of [8]; for fully varied revision theory, the complexity is yet higher, as it will be for the stronger notions of "n-categorical", needed to fend off stronger liar paradoxes. We believe that (ii) of the theorem is also true for S^*, as well as for these semantical schemes with the Herzberger rule, however using the latter imposes severe constraints, and we leave these matters as open questions. The point to be made here is to make precise (at least in one situation) their remark above and ascertain at which higher level the semantics for $\dot{K}$ is, in fact, taking place.

§5. Realistically Varied Sequences.

In this section we make some observations. The motivation is that of seeking for a definition of a revision process that does three things:

(i) reduces the mathematical complexity of the fully varied stable truth set;

(ii) yields some representation of the set (of pairs) of stable truth sets in revision sequences, or at least gives some structure to the class of such stability sets;
(iii) solves (as far as possible) the problem of simple sets of sentences that are intuitively felt to be of a certain category (always stably true/false/unstable *etc.*) but which are ill-classified under the current schemes.

That we should aim to reduce the complexity of fully varied sequences and the accompanying truth set to something that is at least ZF-absolute should be a fair desideratum. One observation on (iii), is that it is not necessary to globally quantify over all sequences and ensure that *all* possible coherent limit rules are used cofinally in every sequence. Just require that simple functions have to be used, as follows.

Idea: we only need to enforce variability in a simple class of limit rules in order for the examples that have been produced in the literature as "ill-classified", to "come out right"; and that *realistic variance* ensures this. Such an example of ill-classification occurs in [C] to provide a counter-example to a revision theoretic system of Yaqūb's, and this would also come out "right" using our definition below; similarly for the subtler variants of the Gupta Puzzle type etc. (*e.g.* [8] 6C.10). [3] Hence any example used as an "objection" to this realistically varied revision theory as being classified as sometime undesirably unstable, or whatever, will have to consist of (at least) a non-recursive set B of sentences (or of some non-recursive sequence of t/f assignments to the sentences of B). It is hard to imagine someone claiming to have sufficient intuition about such a set of sentences B, and the revision processes involved to claim that B has been improperly served by this form of revision process. The test of this notion is then to see if there are such simply defined "ill-classified" sets under realistic variance. (This is the import of the "Challenge Problem" below.)

DEFINITION 5.1. *A revision sequence $\vec{s}$ is* realistically varied *if for all limit $\lambda < \infty$, we let $s_{<\lambda} = (s^+_{<\lambda}, s^-_{<\lambda})$ be the local pair of stability sets at λ, then s_λ is chosen as a coherent extension of $s_{<\lambda}$ in the following fashion:*
(i) Either s_λ is recursive in $s_{<\lambda}$ or in some s_α for an $\alpha < \lambda$, and s_λ has not been used as a limit rule cofinally in λ;
(ii) Or, if at stage λ there is no s_λ that satisfies clause (i), then s_λ may be chosen arbitrarily.

The maxim here then is "use the simple ones first" when it comes to formulating bootstrapping policies. So, to paraphrase, a realistically varied revision sequence is one in which we always first try to set s_λ as something recursive in $s_{<\lambda}$, or in some previous s_α, that we have not already used unboundedly often

[3] Realistic variance argues that intuitive arguments will be, at their most sophisticated, about recursive, or as treated here, hyperarithmetic sets of sentences. Our definition ensures that every recursive choice of cohering t/f assignments is used at limit ordinals cofinally in ∞, and so of course will the finite assignment needed to get this example to stabilise.

below λ. One may show that any t coherent with $s_{<\infty}$ *and* recursive in it, has been used unboundedly in a realistically varied [4] $\vec{s}$. Although the definition is complicated to state in words, it does actually yield a a structural reward (i), and a mathematical simplification (ii):

THEOREM 5.1. *(i) If $\vec{s}$ is a realistically varied revision sequence, then $s_{<\infty}$ forms a Kripkean fixed point for the supervaluation operator.*
(ii) The categorical truth set (over $\mathbb{N}$) of sentences stably true in all realistically varied revision sequences, is complete Π^1_2.

By the supervaluation operator we mean that jump operation j_{vF} that acts on disjoint pairs of sets of sentences (A^t, A^f) (partial sets considered as those true or false at a particular stage) defined as $j_{vf}(A^t, A^f) = (B^t, B^f)$ where $B^t = \bigcap\{V_A | A^t \subseteq A \wedge A \cap A^f = \emptyset\}$ and, using truth and falsehood defined in the partial structure, $V_A = \{\ulcorner\varphi\urcorner | \langle\mathbb{N}, \ldots, A^t, A^f\rangle \models \varphi\}$; $B^f = \bigcap\{\mathbb{N}\backslash V_A | A^t \subseteq A \wedge A \cap A^f = \emptyset\}$.

At this level, the results above on the strongly definable sets over a model M being those of the next stable set $\mathbb{S}_M$ apply.

As mentioned above one test of this theory is to see how hard it is to solve the following:

Challenge Problem Find a set of sentences B that is intuitively of a certain category under some starting hypotheses, but, for example, that is badly classified as "sometimes unstable", according to realistic variance.

§6. An algorithmic theory of truth: stable sets as certain Kripkean fixed points.

The general thrust of these results is that the machinery of revision theory is complicated. It results in truth sets that are either Π^1_2 or yet more complex. The notion of stable categoricity (even assuming a notion of stable truth that is not based on full variance) is also Π^1_3. An approach suggested by realistic variance is that if we focus attention on a single revision process starting from a given hypothesis, then we arrive at a supervaluation fixed point. In particular we can regard such a revision theory as being a generalisation of the Kripkean supervaluation fixed point approach. In the Kripkean theory we may focus attention on certain fixed points (the minimal fixed point, certain intrinsic fixed points, notably the maximal one etc.) rather than try and "take an average" over all such processes. We attribute meaning to a "stable Kripkean set" that is, to a fixed point. Similarly we here attribute meaning to each stability set of each suitable revision sequence. Under the Belnap and Gupta approach no particular meaning is assigned to the set of stabilities occurring in any one revision sequence: it is one more set to feed into the averaging process. We here adopt

[4] Indeed this latter sentence (with hyperarithmetic replacing recursive), could serve as an alternative defining requirement for realistic variance in what follows. It may be that the formal definition above may be too restrictive for some purposes. The point of stating it in this fashion is to emphasise its *non*-globalarity.

the view that the revision process is seeking information based on our initial hypothesis h. The information we seek is itself the set of stabilities of our revision process.

We may thus regard each revision process as a process illustrating an approach to solving the problem of the extent of a language's ability to express truth in itself (as one would the Kripkean approach). In that case it would seem entirely reasonable to restrict the limit rule to resources no more complicated than the process has produced so far. In particular we do not wish to import into the process information which is "remote from" our starting hypothesis, or more complicated than what we are doing (this being one of the sources of complexity of the theory of standard revision theory.).

Let us suppose that in the theory of realistic variance (Definition 5.1) the choices of limit extensions s_λ have been done in some fashion that shows that s_λ has been chosen in some reasonably uniform manner in λ from the preceding sequence. Several examples spring to mind. Let us say, being generous, that s_λ is $\Delta_1(\vec{s} \upharpoonright \lambda)$ definable uniformly in λ in some weak set theory, say KP, (Kripke-Platek which we take to include the Axiom of Infinity). (Surely *primitive recursive*, will more than suffice for any reasonable theory? The Γ_G and Γ_H both conform to this, but we are adding to these the requirement of realistic variance.) Call such a sequence a *(generalised) algorithmically varied* sequence. Given a hypothesis $h = s_0$ to the extension of the truth predicate one then has

THEOREM 6.1. *If $\vec{s}$ is algorithmically varied with starting hypothesis $h = s_0$, then:*
(i) The stability set $s = (s^+_{<\infty}, s^-_{<\infty})$ is a Kripkean fixed point under j_{vF};
(ii) s is recursively isomorphic to the complete eventually writable infinite time Turing machine set of integers relative to h, $\tilde{h}$ ([19] Def. 2.7); equivalently to the complete arithmetical quasi-inductive set relative to h (cf *[4] 13.1).*

As the function $h \to \tilde{h}$ is Δ^1_2 we keep within the bounds of absoluteness between ZF-models. One may show that algorithmically varied sequences are "fully varied" in the sense of Section 3, but where we ensure only that any r that coheres with the whole sequence $\vec{s}$ and is such that r is recursive in any s_α, has been used cofinally. For such sequences we may calculate the length of the "stabilization ordinal" $\sigma(\vec{s})$ - the ordinal by which the revision process starts to cycle repeatedly.

DEFINITION 6.1. *Let $\vec{s} = \langle s_\alpha | \alpha < \infty \rangle$ be an algorithmically varied revision sequence. Let the* stabilization ordinal*, $\sigma(\vec{s})$, be the least σ so that $\forall \alpha \geq \sigma \exists \beta > \alpha \; s_\alpha = s_\beta$.*

The list of equivalences in the theorem below illustrates an interesting convergence of a variety of ideas and concepts. The identity of the ordinals defined in (i) and (iv) is the relativised result [4], 14.1, due to Burgess.

THEOREM 6.2. *Under the hypotheses of the last theorem $\sigma(\vec{s})$ is equivalently:*
(i) The least ordinal $\zeta = \zeta^h$ so that $L_\zeta[h]$ has a transitive Σ_2 end extension;
(ii) The supremum of the infinite time Turing machine "eventually writable" ordinals using as oracle h;
(iii) The starting point of the "Herzberger Grand Loop" ([9]) based on initial hypothesis h.
(iv) The closure ordinal of arithmetical-in-h quasi-inductive definitions.

Of course, from the viewpoint we are adopting it makes no sense to "average out" such truth sets by taking an intersection over all starting hypotheses: we should just arrive back at the same level of complexity: a Π^1_2 complete categorical truth set (albeit with now improved classificatory properties for sets of sentences).

Note: One can still work the theory of "circular" definitions using this approach: the point again is that the extension of a definition is calculated anew from each starting hypothesis as to its extension. Again we do not take an intersection over all starting hypotheses. With this approach:

THEOREM 6.3. *a) The algorithmically varied strongly definable sets of natural numbers from an hypothesis h are then, equivalently: (i) the sets of integers in $L_{\zeta^h}[h]$; (ii) the set of reals eventually writable by an infinite time Turing machine from input h.*
b) $\tilde{h}$ is a complete algorithmically varied weakly definable set.

To each countable model M with, say, an inductive coding scheme, of a language, here the "companion model" would be an analogous structure $\mathbb{Z}_M$ - the "next Σ_2-extendible"-set over M. Again there is an analogous definability theorem to that of Theorem 2.2, with definable subsets of $|M|$. If one wanted one could even construe these as "eventually writable" for some generalised computation over the structure M (much as can be done for ordinary computations over suitable structures - see, for example Hinman's article in [10].)

In a sense to name this (or the realistically varied theory of the previous section) a "generalisation" of the Kripkean supervaluation theory is a misnomer, since not all j_{vF} fixed points occur as algorithmically varied stability sets: the class of such stability sets is a proper subset of the class of such fixed points. But we may view algorithmic revision processes as "stretched out" or elongated processes of attempting to reach certain supervaluation fixed points.

If one desired to adopt the Gupta and Belnap tactic for dealing with Strengthened Liar paradoxes in this context, one could also define the notion of stable categoricity here, just as in Section 3, by adding a predicate to the language and finding repeatedly stability sets relative to such a "hypothesis" in this extended notion, and cycle these stability sets as the successive hypotheses.

THEOREM 6.4. *The stably categorical set over* $\mathbb{N}$*, relative to a starting hypothesis* h*, and using algorithmically varied revision sequences, is* (1-1) *equivalent to the complete* Σ_2 *theory of* $L_{Z^h}[h]$ *where the latter is the smallest transitive model containing* h*, closed under arithmetical quasi-inductive definitions, with a transitive* Σ_2*-end extension.*

If there is any point in stating this rather technical sounding theorem, it is that stable categoricity - whatever that means - now is no longer a non-absolute notion.

Perhaps more germane however, is that the whole theory is simpler in this sense: the Kripkean theory of fixed points, (using either supervaluations or Kleene 3 valued schemes) uses say $KP+$"there exists a transitive model of KP" in the metatheory to find at least one fixed point. However the Belnap and Gupta theory requires a very substantial part of ZF in the metatheory to define the set of stable truths of arithmetic. By way of contrast, the generalised algorithmic theory of truth outlined above, including the notion of (finite orders of) categoricity can all be developed within $KP + \Sigma_2$-Separation.

REFERENCES

[1] G.A. ANTONELLI, *A revision-theoretic analysis of the arithmetical hierarchy*, ***Notre Dame Journal of Formal Logic***, vol. 35 (1994), pp. 204–208.

[2] JON BARWISE, ***Admissible sets and structures***, Perspectives in Mathematical Logic, Springer Verlag, 1975.

[3] NUEL BELNAP, *Gupta's rule of revision theory of truth*, ***Journal of Philosophical Logic***, vol. 11 (1982), pp. 103–116.

[4] JOHN P. BURGESS, *The truth is never simple*, ***Journal for Symbolic Logic***, vol. 51 (1986), no. 3, pp. 663–681.

[5] A. CHAPUIS, *Alternate revision theories of truth*, ***Journal of Philosophical Logic***, vol. 25 (1996), pp. 399–423.

[6] H. FRIEDMAN, *Minimality in the* Δ^1_2*-degrees*, ***Fundamenta Mathematicae***, vol. 81 (1974), pp. 183–192.

[7] A. GUPTA, *Truth and paradox*, ***Journal of Philosophical Logic***, vol. 11 (1982), pp. 1–60.

[8] A. GUPTA and N. BELNAP, ***The revision theory of truth***, M.I.T. Press, Cambridge, 1993.

[9] H. HERZBERGER, *Notes on naive semantics*, ***Journal of Philosophical Logic***, vol. 11 (1982), pp. 61–102.

[10] P. G. HINMAN, *Recursion on abstract structures*, ***The handbook of computability theory*** (E. Griffor, editor), Studies in Logic series, vol. 140, North-Holland, Amsterdam, 1999.

[11] T. JECH, ***Set theory***, Pure and Applied Mathematics, Academic Press, New York, 1978.

[12] P. KREMER, *The Gupta-Belnap systems* $S^{\#}$ *and* S^{*} *are not axiomatisable*, ***Notre Dame Journal of Formal Logic***, vol. 34 (1993), pp. 583–596.

[13] B. LÖWE, *Revision sequences and computers with an infinite amount of time*, ***Journal of Logic and Computation***, vol. 11 (2001), pp. 25–40.

[14] V. MCGEE, ***Truth, vagueness, and paradox: An essay on the logic of truth***, Hackett, 1991.

[15] Y. MOSCHOVAKIS, ***Elementary induction on abstract structures***, Studies in Logic series, North-Holland, Amsterdam, 1974.

[16] M. SHEARD, *A guide to truth predicates in the modern era*, ***Journal for Symbolic Logic***, vol. 59 (1994), no. 3, pp. 1032–1054.

[17] A. VISSER, *Semantics and the liar paradox*, ***Handbook of philosophical logic*** (D.Gabbay and F.Guenther, editors), Reidel publishing Co., Dordrecht, 1989, pp. 617–706.

[18] P. D. WELCH, *On revision operators*.

[19] ———, *Eventually infinite time turing degrees: infinite time decidable reals*, ***Journal for Symbolic Logic***, vol. 65 (2000), no. 3, pp. 1193–1203.

[20] A. YAQŪB, ***The liar speaks the truth. a defense of the revision theory of truth***, O.U.P, New York, 1993.

DEPT. OF MATHEMATICS,
UNIVERSITY OF BRISTOL.
BRISTOL BS8 1TW, ENGLAND.
and
GRADUATE SCHOOL OF SCIENCE & TECHNOLOGY,
KOBE UNIVERSITY,
ROKKO-DAI, NADA-KU
KOBE 657, JAPAN.
E-mail: welch@kobe-u.ac.jp
Current address: Institut für Formale Logik, Währinger Str. 25, A-1090 Wien, Austria.