

Rethinking Revision

P.D. Welch
School of Mathematics,
University of Bristol,
Bristol, BS8 1TW, England

October 27th 2016

Abstract

We sketch a broadening of the Gupta-Belnap notion of a *circular* or *revision theoretic definition* into that of a more generalized form incorporating ideas of Kleene's *generalized* or *higher type recursion*.

This thereby connects the philosophically motivated, and derived, notion of a circular definition with an older form of definition by recursion using functionals, that is functions of functions, as oracles. We note that Gupta and Belnap's notion of '*categorical in L*' can be formulated in at least one of these schemes.

1 Introduction

This article is partially of the nature of a survey, partially some exposition, but works towards a context in which it is natural to make a suggestion for a generalisation of Gupta and Belnap's ([7]) theory of circular definitions to a broader theory. This takes ideas from Kleene's theory of higher type recursion ([14],[15]) to allow for circular definitions that make calls to "sub-revisions" for further input to the revision process. We offer a purely qualitative account of this, eschewing formal definitions which might be intimidatingly lengthy (which can happen in higher type recursion theory) in favour of a short informal description. Thus this article is not about truth *per se*, but about definability. In the next section we outline the path taken from researchers working on pure theories of truth, to the theory of circular definitions of the monograph [7]. This, whilst very brief, has some historical flavour to it but has no pretense to completeness. In Section 3 we discuss Burgess's reformulation of Herzbergerian revision sequences as a 'theory of arithmetic quasi-inductive definitions'. This is thus an early example of a shift away from Truth alone to thinking of revisions as in some way building sets by some sort of inductive process (as Herzberger himself noted).

The emphasis of this article, once past Section 2, is entirely on Herzbergerian revision sequences, namely those that use a \liminf process to form extensions of a predicate to get past limit stages in the evolution of a revision sequence. Being a mathematically precise, and determined, rule it makes each such sequence susceptible to analysis in a way that a Belnap sequence with its unruly behaviour at limit stages does not. This author was quite unaware of the theories of truth that used

revision whilst working on Hamkins's and Kidder's model of *infinite time Turing machines* ([8], Section 4), but it turns out that such a machine can be thought of as an example of a Herzbergerian revision sequence, since it too uses a liminf rule to update its tape's cell values at limit stages of time. Moreover, as a converse, an ittm can be programmed to produce a Herzberger revision sequence on its output, together with a final stability set of that revision. They are thus formally equivalent, and the same countable ordinals appear in the analysis of their behaviour, of looping, halting, or whatever. There are now several notions using this kind of quasi-inductive behaviour and hence often the sets generated are, up to recursive isomorphism, the same (Theorem 1).

Any theory of inductive definability in whatever form must take account of Moschovakis's beautiful theory of *Spector Classes* (Section 5). There is a modest amount of exposition here of this important theory. Sections 6 and 7 are devoted to some new proposals for generalising circular definitions. The comments on ittm's will require the reader to be somewhat familiar with such machines to be entirely meaningful. No prior knowledge of Kleene's higher type recursion is needed however. We conclude with some questions, all relating to this new kind of definability, except the final one which relates to Gupta and Belnap's notion of 'definably categorical' (Sect 6.D *op. cit.*).

A central reference is Gupta's [6] where the theory as one of truth is explained; however we shall move on to the monograph [7] which treats of circular definitions more generally. The first part of the book indeed considers the problems of the semantic approaches to truth, however (p.103):

"In our view hierarchies, whether they are constructed using fixedpoints or otherwise, do not provide a promising solution to the descriptive problem". (They seem to think that such a theory reduces to a theory of "levels" which is "just as difficult".) *"No matter how one constructs the hierarchy, one has to argue that certain sentences containing global uses of 'true' do not mean what they seem to mean."*

In Chapter 4 a sketch is given of motivations for the Revision Theory to follow; we quote liberally from their book to also motivate the ideas here. *"The key to the proper resolution of the problem of truth and paradox lies, in our view, in the theory of definitions. The behavior of the concept of truth, both ordinary and pathological, is strikingly similar to that exhibited by certain kinds of definitions."* (An example is given that allows one to "prove" the doctrine of Monism; they point out that it is similar to the moves in the Epimenides the Cretan paradox.) After giving further examples of how circular definitions behave like self-referential truth assertions, they set out (p117): *"Our aim in this essay is to argue that these suggestions are more than mere possibilities, that they are close to actuality. The logics of circular definitions and truth, in our view, illuminate and support one another."* Henceforth (p119 - here and later always with their emphases) *"A circular definition, though it may not determine the extension of the definiendum, does provide a rule that can be used to calculate what the extension should be once we make a hypothesis concerning the existence of the definiendum"*.

"The meaning a circular definition ascribes to its definiendum, we wish to suggest, should be viewed as having a hypothetical character." (p119). However (p121) what will remain is *"the main problem before us is to state precisely how the transition from the merely hypothetical to the categorical is to be made."* (In other words how to extract from all the various hypotheses one could give to the definiendum, those facts of the matter as to which objects always fall into the eventual definiens).

Once the theory of circular definitions is founded it can be used (p142): *"The theory of truth sketched above is a consequence of combining a general theory of definitions and Tarski's suggestion that the biconditionals be viewed as partial definitions. Tarski's suggestion brings out clearly the circu-*

larity in the notion of truth”.

In order to fix some notation we sketch the main features of the General Theory of Circular Definitions. Each definiendum G has a unique definition of the form:

$$(1) G(x_1, \dots, x_n) =_{\text{df}} A_G(x_1, \dots, x_n)$$

where A_G is a formula of the language (that may contain other definienda including G itself). If we restrict ourselves to countable languages (which we shall) then a more general scheme is to have a countable sequence of definienda $G_i(x_1, \dots, x_{n_i})$ each given by a similar scheme with a formula A_{G_i} for each G_i :

$$(2) G_i(x_1, \dots, x_{n_i}) =_{\text{df}} A_{G_i}(x_1, \dots, x_{n_i}) \quad \text{for } i \in \mathbb{N}.$$

In more model theoretic looking terms this is:

$$G'_i = \{(x_1, \dots, x_{n_i}) \mid (\mathcal{M}, (G_i)_{i \in \mathbb{N}}) \models A_{G_i}[x_1, \dots, x_{n_i}]\}$$

where G'_i is the ‘revised’ version of G_i at the next ‘stage’. Whilst it would appear, *prima facie* that there was a great deal of complexity, or greater generality possible with such ‘interlocking’ or mutually dependent defining clauses, in general this is not the case: multiple simultaneous revision-theoretic definitions can be replaced, or mimicked by a single definition. We shall take as a paradigm the usual structure of $\mathcal{M} = \mathbb{N}$ the natural numbers (where we have suppressed for convenience the other predicates and function symbols, thus \mathbb{N} is ‘really’ $\langle \mathbb{N}, 0, ', +, \times, \dots \rangle$). We consider then arrays of circular definitions as at (2). If we have an array that we can write down, we shall assume therefore that we have an effective method for listing the definitions of (2); thus the list is *effectively computable*. This is a starting assumption. We shall also assume a starting distribution of extensions for the G_i : some $g_i^0 \subseteq \mathbb{N}^{n_i}$, as initial hypotheses for each $i \in \mathbb{N}$. The revisions $g_i^0, g_i^1, g_i^2, \dots, g_i^k, \dots$ for $k < \omega$ are unproblematic: they are given by (2). The problems start at limit cases: how to define g_i^ω ?

2 Limit rules

In the Herzbergerian theory of truth ([9],[10]) we take a structure \mathcal{M} (for simplicity we assume that every element of the domain of \mathcal{M} is named by a term of its language - as is the case for \mathbb{N}) and we have by transfinite recursion on the ordinals H_α sets:

$$H_{\alpha+1} = \{\ulcorner \varphi \urcorner \mid (\mathcal{M}, H_\alpha) \models \varphi\}.$$

where we assume as usual $\ulcorner \cdot \urcorner: \text{Sent} \rightarrow \text{dom}(\mathcal{M})$ is a coding of the sentences of $\mathcal{L}_{\mathcal{M}, \dot{T}}$ - the language of \mathcal{M} augmented by a unary symbol \dot{T} , by objects in $\text{dom}(\mathcal{M})$. A starting hypothesis H_0 is assumed. Herzberger himself never said that he was formulating a theory of truth: he wanted a “non-directive approach”; “Rather than try to eliminate those paradoxes, I want to consider the experiment of positively encouraging them to arise and watching them work their own way out.” ([9], p.479) He formulated a ‘Liminf’ policy at limit stages: a sentence would be evaluated as ‘true’ at limit stage λ , if and only if there was some stage $\alpha < \lambda$ so that for all later $\beta < \lambda$ it was in H_β . Whilst one might argue that this was some kind of parsimony in keeping ‘trues’ to a minimum, it builds a fundamental asymmetry into the process: the ‘falses’ (or ‘not trues’) are swelled by all the sentences which have switched cofinally in λ . He argued that there would be eventually ‘looping’ of

the extensions of H_α : some points $\sigma_0 < \sigma_1$ with $H_{\sigma_0} = H_{\sigma_1}$, and indeed there would be periodicity in such loops which would run through the class of all the ordinals On. Herzberger again: “*The idea is to stand back and let the paradoxes reveal their inner principles*” (*op.cit.* p.479).

This is not the place to catalogue the subsequent discussions and papers by Gupta, Belnap and others in their adaptations of Herzberger’s original idea. The thrust of those discussions appears to be that limitations of the limit rules proposed always treated some sentence in an ill-advised or undesirable manner. The ‘Belnap Rule’ that at a limit stage anything could be substituted as long as it was consistent with the semantic values that were stable up to the point in question, then appears as the inevitable and final limiting rule: there is no rule. It should however be pointed out that both Gupta and Belnap thought that any attempt to move to the final truths or falsities of their systems should involve considering all possible starting distributions or hypotheses. The truths that emerged would be those stable in all possible sequences from all possible starting hypotheses. This trial by fire would end up with the truth sets being of very high complexity: *e.g.* with $\mathcal{M} = \mathbb{N}$ such sets, albeit ostensibly about *first order* truth, would only be *second order* definable in analysis with an $\forall\exists$ alternating of second order quantifiers (see, for example, [2], 12.3). This is an order of complexity that overwhelms the strength of almost any theorem of mathematics. (In this author’s view, this already vitiates such theories as theories of truth over \mathbb{N} with a simple T -predicate.) Further efforts were made to introduce more sifting through of truth sets, by requiring whole enumerations of the continuum be repeated cofinally often through the ordinals, this somewhat Ptolomaic approach perhaps reaching its apogee in [26].

Moreover, it would turn out that such truth sets would, up to a pencil and paper algorithm, turn out to be the same whether obtained from a Herzberger, a Gupta, or a Belnap rule, if one is adapting the procedure of intersecting over all starting hypotheses. (Moreover for a Belnap Rule - [1] - with its unlimited freedom at limit stages, one can replace the quantification over the real continuum of starting hypotheses that results in truth sets of such high complexity just alluded to, and instead one could always start with a single, empty say, hypothesis as extension; the application of the limit ‘Rule’ alone would then embed that real universal quantification in another way. However, as the referee has kindly pointed out, this embedding of universal quantification into the limit stages will not be possible for simple models where sentences only have quote-names in the domain of the model, but where coding and self-reference is not possible. This article really only considers models extending that of the natural numbers with revision theoretic definitions that make use of Gödel coding - and hence allow the possibility of self-reference.) The phenomenon is discussed in [23], [16], indeed in a number of other places, and by a variety of authors.

However the emphasis of the 1993 book [7] is shifted. Whilst the introductory sections are speaking of truth and how the Tarskian biconditionals are *definitional* of truth, and deliver the *signification* of truth, when the actual descriptions of the technical apparatus commence they are through the medium of a *theory of circular definitions*. The definition of truth will then be just one example of a circular definition. The thrust of the argument is then in defining, and investigating their notion of circular definition as I have hastily sketched it above.

As opposed to the theory of truth, the theory of circular definitions seems entirely unexceptionable: Belnap and Gupta give a full definition of such. Then, afterwards, the application to truth is considered with various alternative versions of notions of validity *etc.* discussed. It is solely on the theory of circular definitions that we want to focus in this article.

3 Circular definitions

The case of circular definitions using the Herzberger Liminf rule is amenable to a close mathematical investigation of its structure and outcomes. For the rest of this article we shall restrict ourselves to this style of Herzberger sequence. (Thus ‘circular definition’ or, synonymously, ‘revision theoretic definition’ will be interpreted *via* revision sequences based on the Liminf rule. Usually we consider just a single inductive process emanating from a single starting ‘hypothesis’, ‘point’ or ‘set’. There will be no ‘averaging out’ over all starting hypotheses.) In his 1985 article [2], Burgess takes a notion of *arithmetical quasi-inductive definition* (*a.q.i.*). An a.q.i. set $Y \subseteq \mathbb{N}$ is given by having an arithmetic formula $\Phi(v_0)$, taking a starting hypothesis $Y_0 = \emptyset$, and revising according to scheme (1) over $\mathcal{M} = \mathbb{N}$, with the Liminf rule at limit stages. The emphasis is then on the single revision sequence, rather than ‘averaging’ over the continuum many such with differing Y_0 . Tractably, he gave an exact characterisation of the a.q.i. sets of integers, as those definable at a certain level of complexity at a certain level of the Gödel constructible hierarchy. As this feature will loom large in the following discussion we dwell on this here. The connection with the constructible hierarchy should not be surprising: each successor stage operation of $Y_\alpha \rightarrow Y_{\alpha+1}$ is an arithmetic one; the starting hypothesis, \emptyset , is a constructible set (!) as is the structure $\mathcal{M} = \mathbb{N}$. As arithmetic operations are *absolute* between V and L , running the revision sequence inside L yields exactly the same sequence $\langle Y_\alpha \mid \alpha \in \text{On} \rangle$ as outside. But the liminf rule is expressible in a Σ_2 -fashion. Consequently the sequence $\langle Y_\alpha \mid \alpha < \lambda \rangle$ is definable by a Σ_2 -recursion set-theoretically inside the structure $\langle L_\lambda, \in \rangle$. Hence the membership of say n in Y_λ is given by a Σ_2 expression over $\langle L_\lambda, \in \rangle$, and is dependent only on Φ and Y_0 . The easy part of Burgess’s argument is to consider what happens if we then continue until some ordinal ζ_0 where for some larger $\zeta_1 > \zeta_0$ we have that $L_{\zeta_0} <_{\Sigma_2} L_{\zeta_1}$. ($\mathcal{M} <_{\Sigma_2} \mathcal{N}$ means that any Σ_2 sentences about objects in $|\mathcal{M}|$ which is true in \mathcal{N} is also true in \mathcal{M} .) But this means the membership question of whether n is in Y_{ζ_0} has the same outcome as in Y_{ζ_1} . In short $Y_{\zeta_0} = Y_{\zeta_1}$. We thus have arrived at an hypothesis that is destined to ‘loop’ with a fixed periodicity (in fact of period $\leq \zeta_1$). Then Y_{ζ_0} is deemed to be the a.q.i. set that this process throws up. We may then write $Y_\infty = Y_{\zeta_0}$. (The harder part of the argument is to say that any set $Z \subseteq \mathbb{N}$ that is Σ_2 -definable over L_{ζ_0} is also definable by an a.q.i. scheme of this kind.)

That a.q.i. schemes in this manner run to the ordinal ζ_0 is indicative of their strength. (The assertion that every a.q.i. scheme over $\mathcal{M} = \mathbb{N}$ reaches a loop can be proven in $\Pi_3^1\text{-CA}_0$ but not in $\Pi_2^1\text{-CA}_0$: one needs these strong systems to prove that sufficiently long ordinals exist along which one may run the quasi-inductive progressions.) The a.q.i. set Y_∞ is generated by a process over a long stretch of the countable ordinals, far longer than the often considered *recursive ordinals* with supremum the Church-Kleene ω_1^{ck} . One may I suppose, take the idea that positive, or ‘certain’, membership in Y_∞ is afforded by this process of determining *via* ‘eventual membership’, whilst all other elements are excluded.

In any case, it is a scheme of definability that involves processes with ordinals to a greater degree than the monotonic processes that usually occur in, say theories of truth, or indeed philosophical logic generally. At the risk of repetition it is precisely this involvement of ‘ordinal analysis’ at limit stages, that introduces an element that is far from the original motivation of finding ‘truths’ over the structure of the natural numbers - indeed disqualifying that outcome set H_∞ from having much to do with truth over \mathbb{N} .

However we can just regard the theory of revision theoretic circular definitions as proposed by

[7] as one further method of defining interesting classes of formulae/sets, without suggesting that this gives a formal method to deciding what, *e.g.* ‘truths’ are over \mathbb{N} , or what a given scheme such as at (1) delivers. It should be pointed out again, that one indication of the overwhelming nature of the ordinal construction, is that for Herzberger revision sequences, a whole slew of different sets will yield essentially the same outcome: whether H_0 is empty, or is any recursive set of truth values (or even hyperarithmetic or yet more complicated) the resultant sets H_∞ will all be recursively isomorphic: there will be again pencil and paper algorithms for converting one to the other: all differences of the initial hypothesis will have been dwarfed by the complexity of the definitions’ ordinal machinery.

However the theory of circular definitions still provides a method for defining sets of objects from a structure by some form of ‘quasi-inductive’ definition. The non-monotonicity of the construction and the dependence on relatively long countable ordinals has probably meant that this is a system that does not crop up frequently (or as some would say: ‘naturally’) elsewhere. There are two avenues to explore here: relations to infinite time Turing machine computation - which turned out, independently, to have exactly the same strength as Herzberger revision; and the methodology with which we shall be concerned to sketch an exposition of here: an enlarged notion of higher type circular definition akin to Kleene’s recursion in higher types.

4 Infinite Time Turing Machines

To this author’s knowledge, Benedikt Löwe seems to have been the first to point out (in correspondence) the formal similarities between the construction of a single Herzberger sequence and that involved in the *Infinite Time Turing Machines* (ittm’s) of Hamkins and Kidder ([8]). In the papers ([18],[17]) this was investigated. The original ittm model of [8] used a limsup rule to update a Turing machine tape’s cells, but in terms of computational power this is identical to a liminf rule, and so this difference is ignored. The formalisms are indeed parallel: one can write an ordinary Turing machine program P_H so that on input the hypothesis H_0 , the machine will behave just like a Herzberger sequence writing successively the values H_α to its output tape. The machine will loop itself (this is an essential feature of such machines just as it was for Herzberger), and so for $H_0 = \emptyset$ the machine loops with $H_\infty = H_{\zeta_0} = H_{\zeta_1}$, for the very same ζ_0, ζ_1 occurring above in Burgess’s proof.¹

In general such machines (on input some set of integers in L_{ζ_0}) when not halting will loop in the same fashion with output tapes repeating at the same times ζ_0, ζ_1 for whatever the program, just for the same reasons as for H -sequences.

The equivalence of a string of such sets defined in differing manners, show that mathematically the same processes have been involved. One can establish the following recursive isomorphisms between sets of integers. Here D is the set of *determinately true* sentences in Field’s [3]; $\tilde{0}$ is the Σ_2 -truth set of L_{ζ_0} , and U_∞ is a *complete* or *universal* a.q.i. set. Finally a *jump* set: $J = \{e \mid \text{the } e\text{’th ittm program } P_e(e) \text{ has a fixed output}\}$.

¹Indeed these features of ittm’s had been independently discovered in [22], whilst being quite unaware of the connections to the earlier Herzberger revision theory that Löwe pointed out, and in ignorance of Burgess’s work.

Theorem 1 [24] *The following sets of integers are recursively isomorphic:*

- a) H_∞ ; b) U_∞ ; c) D ; d) $\tilde{0}$; e) J .

(Burgess's arguments in [2], without explicitly stating it, essentially show the equivalence of a), b), and d).) The appearance of D on the list is all down to the fact that Field uses a liminf rule starting from a recursive distribution of initial semantic values. At successor stages he uses a completely different approach: instead of an arithmetic operation, or a simple truth taking operation, he uses a Kripkean Strong Kleene jump at each stage, which is a big leap in complexity. But as indicated above the liminf rule used at limit stages irons out all subtleties of the successor step. One should say, *mathematical* subtleties, since his final set D , eminently does what he requires: it is a set of sentences with precisely the set of *logical* subtleties he is aiming for: this shows the consistency of his approach. Since his stance is purely instrumentalist, and is using the revision sequence purely to demonstrate the existence of a set with the right desiderata, the revision theoretic means of doing it, is, for him, not the point.

However the point of highlighting the theorem is that we have several quite different ideas, two truth theoretic, one set theoretic, and one from generalised notions of computability, all of which converge on the same sets, and all are examples of revision theoretic definitions. So perhaps this kind of quasi-inductive definition is a little more 'natural' than at first thought.

5 Moschovakis's theory of Spector Classes

The mathematical logician may be prompted to ask how revision theoretic definitions, or the "a.q.i." in Burgess's nomenclature, fit in with the well-developed theory of *Spector Classes*. This is a beautiful structure theory of definable relations over structures \mathcal{M} elucidated by Moschovakis in the late 1960's and 1970's. (See his monograph [19] but also the later [20]; the paper [13] also gives short definitions and gives results relating to generalised quantifiers.)

Definition 1 *Let $\mathcal{M} = (M, R_1, \dots, R_n)$ be an infinite relational structure. A class of relations Γ on M is called a Spector Class on \mathcal{M} , if:*

- (i) Γ is closed under $\wedge, \vee, \exists, \forall$;
- (ii) Γ contains all relations that are first order definable over \mathcal{M} using parameters from M ;
- (iii) Γ is \mathcal{M} -parameterized;
- (iv) \mathcal{M} contains a coding scheme;
- (v) (PWO) Γ has the prewellordering property.

We have given the full definition in order to talk about some aspects in more detail. We need to give some explanation: (i) and (ii) are straightforward basic closure conditions. (iii) amounts to saying that for each n there is a *universal* set $U^n \in \Gamma \cap M^{n+1}$ so that for any other $R \in \Gamma \cap M^n$ there is some $a \in M$ so that

$$(3) \quad \langle r_1, \dots, r_n \rangle \in R \iff \langle r_1, \dots, r_n, a \rangle \in U^n.$$

(If we think of $\mathcal{M} = \mathbb{N}$ and Γ as the collection of Π_1^1 -relations over \mathbb{N} , then these are \mathbb{N} -parameterized: there is a universal $U^n \in \Pi_1^1$ so that any n -ary Π_1^1 -relation R is the fibre over some $a \in \mathbb{N}$ of U^n in the manner of (3).)

A coding scheme is somewhat lengthy to define and write out and we shall not do it, but essentially it amounts to a definable ‘pairing function’ over M , sending $M \times M$ to M ; by ordinary definability there are thus functions $M^n \rightarrow M$: Γ must contain a suitable class of such functions.

Lastly the important property (v): for any, say, relation $R \in \Gamma$ we may *prewellorder* R using some ordering $\leq \in \Gamma$ (a *prewellorder* only fails to be a wellorder because we may have $a \leq b \wedge b \leq a$ without $a = b$; it thus wellorders a collection of equivalence classes of elements of M). Associated to a prewellordering \leq then is a wellorder of these *equivalence classes* and hence a rank function or *norm* $\sigma : R \rightarrow \text{On}$. Actually the prewellordering property requires something slightly stronger than requiring that $\leq \in \Gamma$, it *almost* requires that both \leq and $\not\leq$ are in Γ , but it is not quite as strong as that. It actually says of R and σ that there are relations $S, T \in \Gamma$ so that (letting $\neg T$ be the complement of T):

$$(4) \quad y \in R \rightarrow [(x \in R \wedge \sigma(x) \leq \sigma(y)) \leftrightarrow S(x, y) \leftrightarrow \neg T(x, y)].$$

Why are Spector classes important? Clearly the definition collects together some basic properties needed to allow some basic reasoning about classes of relations. The PWO condition is of a different kind to the the others. What it enables is the following: If one has relations R_1, \dots, R_n in a Spector class Γ over a structure \mathcal{M} , and P is any relation over \mathcal{M} inductive in the \vec{R}_i then too $P \in \Gamma$. A Spector class associated with a suitable structure \mathcal{M} is thus *closed under inductive definability over that \mathcal{M}* . In general there is a rich structure theory of Spector classes going beyond these few statements here.

All the Kleene pointclasses of Π_n^1, Σ_n^1 for any $n \geq 1$ satisfy all the clauses, except perhaps for the prewellordering property. The smallest provable Spector class over \mathbb{N} is the class of relations Π_1^1 - so inductive - over \mathbb{N} . This is the canonical example of such a Spector class. From this it can be shown that the Σ_2^1 classes form a Spector class. However beyond that, it is a matter of strengthening the axioms of set theory (usually by determinacy hypotheses) to establish this for further classes such as Π_3^1 and above. (And even just for Π_1^1 and Σ_2^1 , isolating the Spector class properties they enjoy allows many more results to be proved about them - see [20] 4C.)

Since Π_1^1 forms a Spector class we already have an instance of this in theories of truth: the minimal fixed point in the various schemes proposed by Kripke. Suppose we take the Strong Kleene scheme as an example. The PWO property of the definition above requires that the set of truths in the minimal fixed point, (V^+, V^-) say, (which is a Π_1^1 (pair of) sets) be prewellordered in a Π_1^1 fashion. This it is: for the natural ordering on sentences $\varphi \in V^+$ given by ‘ $\sigma(\alpha)$ is the least ordinal α such that $\varphi \in V_{\alpha+1} \setminus V_\alpha$ ’ is map from V^+ onto ω_1^{ck} , one can find $S, T \Pi_1^1$ relations to satisfy (4).

Further properties of Spector classes imply that one can prove versions of the Recursion Theorem. One may ask: is a Spector class always given by some form of inductive process? A theorem of Harrington shows that any Spector class over a structure \mathcal{M} can always be viewed as an inductively given class, if one allows the inductive definition to be in a language with a *generalised quantifier* (such a quantifier Q is any set $\emptyset \neq Q \subseteq \mathcal{P}(M)$ which is monotone, that is $A \subseteq B \wedge A \in Q \rightarrow B \in Q$; such a quantifier has a dual $\neg Q$). Let us denote the usual language for \mathcal{M} enhanced by the quantifier Q and its dual, as $\mathcal{L}_{\mathcal{M}, Q}$. We then have:

Theorem 2 (Harrington - Representation Theorem, [13] 3.2) *If $\mathcal{M} = (M, R_1, \dots, R_n)$ is a structure and \mathbb{S} a Spector class over \mathcal{M} then there is a quantifier Q over M so that \mathbb{S} is precisely the class of relations inductive over \mathcal{M} in the language $\mathcal{L}_{\mathcal{M}, Q}$.*

Turning to the specific matter in hand, we have that:

Theorem 3 *The relations a.q.i. over \mathbb{N} form a Spector class on \mathbb{N} .*

Proof: The relations on \mathbb{N} which are the outputs of ittm's can be shown to form a Spector class. As each a.q.i. can be mimicked on an ittm, and *vice versa*, these outputs coincide with the a.q.i.-definable relations of \mathbb{N} , and the result follows. Q.E.D.

The essential part of the fact that the a.q.i.-definable relations form a Spector class, has to be then that there is an a.q.i. prewellordering on any $X \subseteq \mathbb{N}$ which is a.q.i. Just as for Kripkean minimal fixed points that prewellordering is simply explained: define a norm by setting, for an $n \in X$, $\sigma(n)$ to be that ordinal α such that from stage α onwards n is in the α 'th iteration of the a.q.i. definition. Then (it can be shown) σ maps the stable truths *onto* ζ_0 . Some consequences are then deducible for revision theoretically definable sets over \mathbb{N} . The following is a 'Boundedness Lemma' for a.q.i. definitions. It can be regarded as an example of a general theorem about Spector classes due to Moschovakis (see [19] 3C.1).

Lemma 1 (Boundedness Lemma for a.q.i. definitions)

- (i) *Any wellorder α that is coded by some set $X \subseteq \mathbb{N}$, if both X and $\mathbb{N} \setminus X$ are a.q.i. then $\alpha < \zeta_0$.*
- (ii) *But every ordinal $\alpha < \zeta_0$ is so coded by an a.q.i.-definable X whose compliment is also a.q.i.*

Ultimately some form of this Boundedness Lemma is used in the main argument of [25] which refutes Field's claim that there is a revenge-immune solution to the semantic paradoxes (as worked towards in [3], [4]): the argument of [25] establishes that the potentially definable length of Field's *determinateness hierarchies* is again this ordinal ζ_0 , but that diagonalisation still enables us to transcend this.

Lemma 2 'Uniform Definability' ([25]) *There is a single uniform method of arithmetically defining the whole sequence $\langle H_\gamma \mid \gamma < \beta \rangle$ from H_β for any $\beta < \Sigma$. This method is uniform in the sense that it is independent of β .*

From the lemma one may derive the fact that we can express "being stably true" within the language:

Corollary 1 *There is a formula $\Phi(v_0, v_1)$ so that $\Phi(\ulcorner \psi \urcorner, \bar{n})$ expresses " $\ulcorner \psi(v_0) \urcorner$ is a code of a formula and $\psi[\bar{n}/v_0]$ is stably true" in the language of arithmetic, so that if ψ a.q.i. defines the set $A \subseteq \mathbb{N}$, then $k \in A$ iff $\Phi(\ulcorner \psi \urcorner, \bar{k})$ itself is in H_∞ .*

6 Revision theoretic definitions in higher type: relativized definability

The previous sections have sketched out some of the historical development of the analysis of one particular kind of revision theoretic definition: that of the Herzberger style, liminf, single revision sequence, with the monadic relation ultimately defined being the eventual stable collection of objects in the extension of the defined symbol. However if we take revision theoretic definitions seriously we should also think about how such definitions can be strengthened or extended, or how

they might fit in either with notions coming from the Spector class theory of definable relations, or, for example, the classical theory of higher type recursions due to Kleene, or indeed any other theories of definability developed by mathematical logicians.

Theorem 1 above encourages us to think of, at least in its Herzberger version, circular definition as potentially a kind of recursion, or quasi-inductive process, albeit of a transfinite kind. If we consider ordinary recursions or *computations* à la Turing, there is the notion of an *oracle*: a set $A \subseteq \mathbb{N}$ and an adaptation of the computable process with a functional call to ask the question ? Is $k \in A$? and to receive a 1/0, or Yes/No answer; this is envisaged as taking one step in the computation's programme which has an enlarged syntax and grammar to accommodate such oracle queries. The resulting theory of *relativised computability*, going back to Turing and Post, is a highly developed one (see for example [21] for a contemporary account). One can effect the same for ittm's. However note here that A , being merely a subset of \mathbb{N} , could actually have been written out in full on the starting input tape, and a programme consulting an oracle for A is tantamount to another related programme searching through an input tape on which A is written out in full. Suppose one envisaged a Herzberger revision sequence over $\mathcal{M} = \mathbb{N}$, with the ability at a stage to quiz a set A of this kind. Then we could regard this as a circular definition over the expanded structure (\mathcal{M}, A) instead, and thus have no need for this kind of extended circular definition. The circular definition over the expanded structure thus replicates the move of writing the data to the tape of the ittm.

So whilst this can be developed as a theory of *relativised circular definability* we have not captured any significantly new process. A more substantial notion would be to consider instead as oracle set some $\mathcal{A} \subseteq 2^{\mathbb{N}}$. Now we may think of relativised circular definability as occurring over the two sorted structure $\mathcal{M} = (\mathcal{A}, \mathbb{N})$. We continue to think of the domain of \mathcal{M} as still the natural numbers, and all of these are named in the language. It is now that the circular definition, meaning the actual formula defining by repetition our extension, contains a symbol $\dot{\mathcal{A}}$, and may quiz \mathcal{A} as to whether the current extension of \dot{G} is, or is not, in \mathcal{A} . In this way, significant information can be introduced into the quasi-inductive process. As the reader by now may suspect, this is mirroring an analogous notion of relativised ittm-computation, where an ittm may query the oracle set \mathcal{A} ($\subseteq 2^{\mathbb{N}}$), as to whether, say, the current output tape contains an element of \mathcal{A} .

We may then consider the class of relations over \mathbb{N} that are \mathcal{A} -revision theoretically definable, for any given \mathcal{A} :

Lemma 3 *Let $\mathcal{A} \subseteq 2^{\mathbb{N}}$. The class of relations over \mathbb{N} that are \mathcal{A} -revision-theoretically definable form a Spector class.*

Whilst ittm's are intimately tied up with computing on Cantor space, that is $2^{\mathbb{N}}$, revision theoretic definitions are not. We correspondingly have something more general:

Lemma 4 *Let \mathcal{M} be an acceptable structure. Let $\mathcal{A} \subseteq 2^{\text{dom}(\mathcal{M})}$. The class of relations over \mathcal{M} that are \mathcal{A} -revision-theoretically definable form a Spector class.*

(We abstain from defining properly what an acceptable structure is: the reader will have enough of an idea, if one imagines a structure which contains something like a copy of the natural numbers, with definable pairing functions. The Lemma bears the same relation to the predecessor as similar results about acceptable structures with coding schemes do to the basic case of the natural number structure. The reader may find how this is done for acceptable structures in [19].) One may then

define ordinals which correspond to ζ_0 and ζ_1 above; these may be identified, and play the same analogous role. We do not explore this here. But there is one particular $\mathcal{A} \subseteq 2^{\mathbb{N}}$ we should like to consider.

$$\mathcal{E} = \{S \subseteq 2^{\mathbb{N}} \mid \exists H \exists \Phi \in \mathcal{L}_{\dot{H}}[S \text{ is the stability set of the revision theoretic definition starting with } H = H_0 \text{ and defining formula } \Phi]\}.$$

Thus \mathcal{E} is the set of all possible stability sets of (ordinary) circular definitions with any possible starting hypothesis. However now note that for circular definability relativized to \mathcal{E} , more will be definable using such relativized definitions: for, e.g. with $H_0 = \emptyset$, a relativized definition in $\mathcal{L}_{\dot{H}}$ can be written that eventually will ‘recognize’ H_∞ , the stability set (it will successively quiz \mathcal{E} and eventually receive a 1 back that the extension is indeed that stability set). The definition may then invoke a clause to go off and do something else. Thus the ability to recognise simple stability sets, will lead to a larger class of circular-definable relations. These will also form a Spector class. The ‘Spector ordinal’ corresponding to ζ_0 can then be investigated (and will be larger than the original).

7 Generalised revision theoretic definitions in higher type: recursion *à la* Kleene

Kleene’s equational calculus of recursion in type 2 (or more) objects ([14],[15] are the original papers, but rather for expository accounts see [5], [11], [12]) lends itself to motivating a corresponding generalisation of revision theoretic definability which is much wider and further reaching. Instead of the ‘simple’ \mathcal{A} -revision theoretically definable relations, where the oracle set \mathcal{A} is quizzed at a single step in the revision process, we can demand more. We envisage a query as being a ‘subroutine’ call, not just to give a simple yes/no answer about some membership questions (although in effect it will do that), but rather dynamically as a call to answer a query by acting out another subordinate revision theoretic definition demanded by the ‘control’ or ‘master’ revision process. To illustrate: suppose as above we have a Herzberger-style revision sequence based on a defining formula $\Phi(k, \dot{H})$. Suppose we have a recursive enumeration $\langle \Phi_e(\nu_0, \dot{H}) \mid e \in \mathbb{N} \rangle$ of all such formulae. The formula Φ may contain some clause so that at some stage a query is triggered about some e , some k , and the current extension of H at this stage of the revision process, as to whether k would be in the stability set of the revision sequence starting with H , using formula Φ_e . (The e and k need not be fixed in advance, but could be determined by the revision along the way. We abbreviate by ‘ $\langle e, k, H \rangle$ ’ such a query triple.) We thus may imagine not just receiving out of the ether a yes/no response, but a sub-revision takes place that performs the sequence of revisions, and inspects the result, and passes back the appropriate answer.

So far, little would appear to be different from the \mathcal{E} -revision theoretic definitions just defined. The crucial difference is that we regard the revision theoretic definition based on $\langle e, k, H \rangle$ just described, as also itself capable during its revision stages of making further queries of the same form $\langle e', k', H' \rangle$, thereby initiating further sub-revision-theoretic definitions. (This is mimicking the higher type recursive element of Kleene’s generalised recursion.) Each of these latter sub-revisions themselves may initiate further such, and so on. The outcome is that a successful generalised revision of this form can be represented by a downward growing tree from the top-most node ν_0 as root, labelled with $\langle \Phi, H_0 \rangle$ (see Fig. 1.)

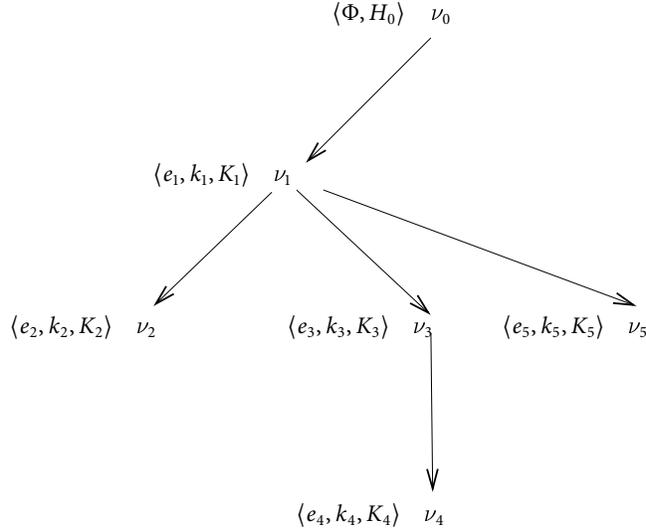


Figure 1: Tree of sub-revision calls

At the first query $\langle e_1, k_1, K_1 \rangle$ a node ν_1 is drawn connected below the top-most node labelled with this triple as the current sub-query. If the revision Φ_{e_1} starting with K_1 in turn instigates a query $\langle e_2, k_2, K_2 \rangle$, this latter triple labels a node ν_2 of the tree connected below ν_1 . If this query at ν_2 is answered without any further sub-revision calls, then an answer is passed up to ν_1 , (meaning that the revision $\langle e_2, k_2, K_2 \rangle$ continues to some looping stage, and then it is apparent whether k_2 is in the stability set or not; this information is passed up to the node directly above - in this case ν_1 - that called this subrevision; the call to the node ν_2 is then over) and the revision $\langle e_1, k_1, K_1 \rangle$ now continues having received its answer. If the continuing revision at ν_1 makes a further query $\langle e_3, k_3, K_3 \rangle$ at a later stage in its own running, then node ν_3 can be represented as being written connected by a branch below ν_1 again, but placed to the right of ν_2 . This ν_3 is at a level “at depth 2” below ν_0 . In Fig. 1. $\langle e_3, k_3, K_3 \rangle$ makes a call at some stage for some $\langle e_4, k_4, K_4 \rangle$, before it stabilizes, and so on. A ‘successful’ or *convergent* revision of this kind then is represented by a well founded finite path tree $T(\Phi, H_0)$ with nodes labelled below the top of the form $\langle e_\tau, k_\tau, K_\tau \rangle$ for $\tau < \theta$. As the revision proceeds each of the nodes is visited in turn (according to the increasing index $\tau < \theta$) in a left-most path search fashion.

Can a revision of this kind fail to produce a stable set? Indeed it can: it would be an exercise to write a self-referential revision theoretic definition which calls itself so to speak. The resulting tree would then have an infinite descending branch - that is would be ill-founded. In this case the outcome of the revision would be undefined, we should say the revision was *divergent*, and no answer or stability set would result.

Can we characterise which initial $\langle \Phi_e, H \rangle$ revisions are convergent? We collect the convergent revisions together by looking for a least fixed point of a functional Δ to be defined in the following manner. We set:

$\Delta(X) :=$
 $\{\langle \langle \Phi_e, k, H \rangle, i \rangle \mid \text{the revision } \langle \Phi_e, H \rangle \text{ is convergent and makes calls only to subrevisions}$
 $\langle \langle \Phi_{e'}, k', H' \rangle, i' \rangle \in X, \text{ with resulting response } i = 1/0 \text{ iff } k \text{ is, or is not, in the}$
 $\text{resulting stability set of } \langle \Phi_e, H \rangle\}.$

We then set $\Delta^0 = \emptyset$; $\Delta^{\alpha+1} = \Delta(\Delta^\alpha)$; $\text{Lim}(\lambda) \longrightarrow \Delta^{<\lambda} = \bigcup_{\alpha < \lambda} \Delta^\alpha \wedge \Delta^\lambda = \Delta(\Delta^{<\lambda})$.

Clearly Δ is a monotone operator and will have a least fixed point Δ^∞ . We then have a (type 2) functional S (for ‘stability’) by setting $S = \Delta^\infty$. The convergent revisions $\langle \Phi_e, H \rangle$ are then precisely those so that for some k $\langle \Phi_e, k, H \rangle \in \text{dom}(S)$.

There are many questions that can be asked about this generalised revision-theoretic definability. We have ‘answered’ the question of which revisions are convergent, by means of the fixed point of Δ , but this gives little concrete information on its own.

Q1 *What is the relation of the original Herzbergerian revision sequences to this proposal?*

Q2 *Is there some further way to characterise the convergent $\langle \Phi_e, H \rangle$?*

More questions will occur to the reader:

Q3 *If $\langle \Phi_e, H \rangle$ is convergent, how many steps of revision does it take to loop?*

Q4 *How complex are the resulting stability sets? What kind of sets can we regard them as defining for us?*

More particular than Q4 is:

Q5 *How strong a theory is needed to define these general revision procedures? In particular for, say, recursive H , which ordinals can be defined by the generalised revision $\langle \Phi_e, H \rangle$?*

We finish with one more question that does return to the roots of revision: the notion of expressing ‘categorical in \mathcal{L} ’ within \mathcal{L} . The relevant starting point here is [7]’s Section 6D. Here ‘categorical in \mathcal{L} ’ means being stably true in every (pertinent for the discussion) revision sequence. Again let us fix the discussion as referring to Herzberger revision sequences over $\mathcal{M} = \mathbb{N}$. The problem of categoricity is that simple assumption of its definability over all revision sequences leads to *strengthened liar paradoxes*. As an example: suppose that the concept “ φ is categorical in \mathcal{L} ” were definable over all revision sequences by the formula $\Psi(v_0)$ of \mathcal{L} . This means:

- (1) $\Psi(\ulcorner \sigma \urcorner)$ is categorical in \mathcal{L} iff σ is categorical in \mathcal{L} .
- (2) $\neg\Psi(\ulcorner \sigma \urcorner)$ is categorical in \mathcal{L} iff σ is not categorical in \mathcal{L} .

By diagonalisation find a sentence σ so that $\sigma \longleftrightarrow \neg\Psi(\ulcorner \sigma \urcorner) \vee \neg T(\ulcorner \sigma \urcorner)$. Such is our strengthened liar sentence. If σ is categorical in this sense then so is $T(\ulcorner \sigma \urcorner)$. But then so is $\neg\Psi(\ulcorner \sigma \urcorner)$ which is absurd. The case of $\neg\sigma$ being categorical, and the case of neither being categorical also easily lead to contradictions.

The question addressed in 6D is whether being ‘categorical in \mathcal{L} ’ can be expressed with the given language or whether one needs to ascend to a metalanguage (p229). Belnap and Gupta argue that

diagnosis of the strengthened liar is that the notion of categoricity is just as circular as that of truth, and should be handled by the same methods as ‘true in \mathcal{L} ’: namely as a rule of revision *within the language*. We thus choose as a *hypothesis* for being categorical some collection of sentences K_0 . We consider the enlarged structure $\mathcal{M} = (\mathbb{N}, K_0)$ and in the extended language $\mathcal{L}_{\mathcal{M}}$ build revision sequences over that. The stable truths that survive this process then form a new hypothesis K_1 for a repeat of the construction. At limit stages one does, well, whatever one does at limit stages, but in this case we naturally take \liminf ’s to mirror that process for truth. Gupta and Belnap then give examples (pp.233-234) of how ‘Strengthened Liars’ and ‘Strengthened Truth Tellers’ are handled in this framework. The thrust of their argument is that certain semantic concepts derived from their approach, such as this form of categorical truth, can be handled with their methods and that there is no need to ascend to a meta-language. But for us here we should like to see how this can be handled with these extended ideas of circular definition. The way this has been described here I think allows us to see that this could be easily done using the relativised circular definability of Section 6: one allows the circular definition to consult \mathcal{E} : we let \mathcal{K} be the stability set of the circularly defined sequence of extensions of $\langle K_\nu | \nu \in \text{On} \rangle$. Then it is easy to see that \mathcal{K} is \mathcal{E} -revision theoretically definable.

It would seem that using generalized revision theoretic definability just sketched ought to allow one to also describe a notion of robust categoricity. so we ask:

Q6 Can one use the generalized revision theoretic definability of this section to define a notion of categoricity similar to, but perhaps stronger than, that of \mathcal{K} ?

References

- [1] N. Belnap. Gupta’s rule of revision theory of truth. *Journal of Philosophical Logic*, 11:103–116, 1982. [4](#)
- [2] J.P. Burgess. The truth is never simple. *Journal of Symbolic Logic*, 51(3):663–681, 1986. [4](#), [5](#), [7](#), [16](#)
- [3] H. Field. A revenge-immune solution to the semantic paradoxes. *Journal of Philosophical Logic*, 32(3):139–177, April 2003. [6](#), [9](#)
- [4] H. Field. Solving the paradoxes, escaping revenge. In J.C. Beall, editor, *The Revenge of the Liar*. O.U.P., Oxford, 2008. [9](#)
- [5] R.O. Gandy. Generalized recursive functionals of finite type and hierarchies of functionals. *Annales de la Faculté des Sciences de l’Université de Clermont-Ferrand*, 35:5–24, 1967. [11](#)
- [6] A. Gupta. Truth and paradox. *Journal of Philosophical Logic*, 11:1–60, 1981. [2](#)
- [7] A. Gupta and N. Belnap. *The revision theory of truth*. M.I.T. Press, Cambridge, 1993. [1](#), [2](#), [4](#), [6](#), [13](#)
- [8] J.D. Hamkins and A. Lewis. Infinite time Turing machines. *Journal of Symbolic Logic*, 65(2):567–604, 2000. [2](#), [6](#)
- [9] H.G. Herzberger. Naive semantics and the Liar paradox. *Journal of Philosophy*, 79:479–497, 1982. [3](#)
- [10] H.G. Herzberger. Notes on naive semantics. *Journal of Philosophical Logic*, 11:61–102, 1982. [3](#)
- [11] P. Hinman. *Recursion-Theoretic Hierarchies*. Ω Series in Mathematical Logic. Springer, Berlin, 1978. [11](#)
- [12] A. Kechris and Y.N. Moschovakis. Recursion in higher types. In Barwise, editor, *Handbook of Mathematical Logic*, Studies in Logic and Foundations of Mathematics, chapter C6, pages 681–738. North-Holland, Amsterdam, 1977. [11](#)

- [13] A.S. Kechris. On Spector classes. In A.S. Kechris and Y.N. Moschovakis, editors, *Cabal Seminar 76-77*, volume 689 of *Lecture Notes in Mathematics Series*, pages 245–278. Springer, 1978. [7](#), [8](#)
- [14] S.C. Kleene. Quantification of number-theoretic predicates. *Compositio Mathematicae*, 14:23–40, 1959. [1](#), [11](#)
- [15] S.C. Kleene. Recursive quantifiers and functionals of finite type II. *Transactions of the American Mathematical Society*, 108:106–142, 1963. [1](#), [11](#)
- [16] K.-U. Kühnberger, B. Löwe, M. Möllerfeld, and P.D. Welch. Comparing inductive and circular definitions: parameters, complexity and games. *Studia Logica*, 81:79–98, 2005. [4](#)
- [17] B. Löwe. Revision sequences and computers with an infinite amount of time. *Journal of Logic and Computation*, 11:25–40, 2001. [6](#)
- [18] B. Löwe and P.D. Welch. Set-theoretic absoluteness and the revision theory of truth. *Studia Logica*, 68(1):21 – 41, June 2001. [6](#)
- [19] Y.N. Moschovakis. *Elementary Induction on Abstract structures*, volume 77 of *Studies in Logic series*. North-Holland, Amsterdam, 1974. [7](#), [9](#), [10](#)
- [20] Y.N. Moschovakis. *Descriptive Set Theory*. Studies in Logic series. North-Holland, Amsterdam, 2009. [7](#), [8](#)
- [21] R.I. Soare. *Recursively enumerable sets and degrees*. Perspectives in Mathematical Logic. Springer, 1987. [10](#)
- [22] P.D. Welch. Eventually Infinite Time Turing degrees: infinite time decidable reals. *Journal of Symbolic Logic*, 65(3):1193–1203, 2000. [6](#)
- [23] P.D. Welch. On revision operators. *Journal of Symbolic Logic*, 68(3):689–711, 2003. [4](#)
- [24] P.D. Welch. Ultimate truth *vis à vis* stable truth. *Review of Symbolic Logic*, 1(1):126–142, June 2008. [7](#)
- [25] P.D. Welch. Some observations on truth hierarchies. *Review of Symbolic Logic*, 7(1):1–30, March 2014. [9](#), [17](#)
- [26] A. Yaqu̇ub. *The Liar Speaks the Truth. A Defense of the Revision Theory of Truth*. O.U.P, New York, 1993. [4](#)

8 Appendix

Proof of Lemma 3 Let $\mathcal{A} \subseteq 2^{\mathbb{N}}$. The class of relations over \mathbb{N} that are \mathcal{A} -revision-theoretically definable form a Spector class.

We first return to considerations of why the a.q.i. relations form a Spector class (Theorem 3 above). The proof of the lemma (which the reader may omit) would start by mentioning Burgess's argument ([2], Lemma 14.1) that the a.q.i. subsets of \mathbb{N} are those (1-1) reducible to those relations that are Σ_2 -definable over L_ζ , where ζ is the first point where a repeating extension of any Herzberger revision theoretic sequence occurs. The class of such relations defined in such a manner is well-known to form a Spector class, but to check the latter one has to check the conditions in Definition 1 above. The main points are to argue first that there is for each n , a *universal* a.q.i. relation U^n of $n + 1$ variables that parametrizes *via* its first coordinate the n -ary a.q.i. relations. This follows since there is a universal Σ_2 formula for Σ_2 -satisfaction in the language of set theory over such an L_ζ . The most significant feature of a class of sets Γ being a Spector class, is the Prewellordering property (v) on the list above: that for any relation (say $R \subset \mathbb{N}$ to take the unary case) in Γ there should be relations S, T in Γ that satisfy (4) above for the norm σ derived from a prewellordering.

For us here, a natural norm on such an R (which is a.q.i. as a binary relation) is given by the ordering that for $n, m \in R$ we set $n \leq m \leftrightarrow n$ is stably in R in the revision sequence from no later a stage than that of m .

One has to check that there are corresponding relations S, T which are also Σ_2 -definable and so themselves a.q.i., but this follows as 'Formula $\lceil \sigma \rceil$ stabilizes' can itself be expressed in a Σ_2 -fashion over L_ζ . Or one may argue this directly from Corollary 1 to the Uniform Definability Lemma. In more detail we may set

$$S(x, y) \leftrightarrow 'y \in R' \text{ stabilizes} \wedge 'x \in R' \neq F \text{ at any stage after } 'y \in R' \text{ stabilizes}$$

and

$$T(x, y) \leftrightarrow 'y \in R' \text{ stabilizes} \wedge 'x \in R' \neq F \text{ at some stage after } 'y \in R' \text{ stabilizes.}$$

The Lemma itself follows from relativising all such concepts to the oracle \mathcal{A} . The ordinal ζ is replaced by $\zeta(\mathcal{A})$ the least ordinal $\bar{\zeta}$ so that $L_{\bar{\zeta}}[\mathcal{A}]$ has a Σ_2 -elementary end-extension. This ordinal $\zeta(\mathcal{A})$ is then the least ordinal so that any \mathcal{A} -revision-theoretically definable relation has stabilized in its revision sequence by stage $\zeta(\mathcal{A})$. The \mathcal{A} -a.q.i. relations are then Σ_2 -definable over $L_{\zeta(\mathcal{A})}[\mathcal{A}]$, and one proceeds to argue that they form a Spector class as before. Q.E.D.

For $h \subseteq \mathbb{N}$ a recursive or hyperarithmetical starting distribution of truth values we let the Herzberger sequence commencing with $H_0^h = h$ be: $\langle H_\alpha^h \mid \alpha \in \text{On} \rangle$, with eventual stability set H_∞^h . If $h = \emptyset$, where all sentences are initially assigned falsity, then it is omitted. The following lemma sketches why starting with different hypotheses leads to stability the sets that, "up to isomorphism" are the 'same'. Sameness here is not identity, but in computational terms they are equivalent: there is a pencil and paper algorithm that given a number from one sets computes a corresponding value in the other, and *vice versa*.

Lemma 5 H_∞^h is recursively isomorphic to the stability set obtained by taking $h = \emptyset$. That is $H_\infty^h \equiv_1 H_\infty$.

Proving that $H_\infty^h \equiv_1 H_\infty$ holds directly appears difficult. So we go *via* Σ_2 -truth sets of the Gödel L -hierarchy.

Proof: Let $\langle H_\nu \mid \nu \in \text{On} \rangle$ be the Herzberger revision sequence commencing with $H_0 = \emptyset$. Let $(\bar{\zeta}, \bar{\Sigma})$ be the lexicographic least pair with $L_{\bar{\zeta}} <_{\Sigma_2} L_{\bar{\Sigma}}$. By absoluteness of the definition of $\langle H_\nu \rangle_{\nu \in \text{On}}$ we shall have that $(\langle H_\nu \rangle)^{L_{\bar{\Sigma}}} = \langle H_\nu \mid \nu < \bar{\Sigma} \rangle$. Since $H_{\bar{\zeta}}(H_{\bar{\Sigma}})$ is defined by the Σ_2 liminf rule, it is thus Σ_2 -definable over $L_{\bar{\zeta}}(L_{\bar{\Sigma}})$. We use the notation ' $n \in T_\alpha$ ' to denote that n is (a code of) a sentence in the Σ_2 -truth set, or theory, of L_α . We shall then have that $n \in H_{\bar{\zeta}}$, if it holds, is one Σ_2 fact that holds about n over $L_{\bar{\zeta}}$ and so contributes to $T_{\bar{\zeta}}$. There is thus a 1-1 recursive function $g: \mathbb{N} \rightarrow \mathbb{N}$ so that $n \in H_{\bar{\zeta}} \leftrightarrow g(n) \in T_{\bar{\zeta}}$.

By the given Σ_2 -elementarity $H_{\bar{\zeta}} = H_{\bar{\Sigma}}$ and so this pair of ordinals gives the first pair of repeating extensions of the T -symbol in the revision sequence: it is the beginning of Herzberger's "Grand Loop". Thus $H_{\bar{\zeta}} = H_\infty$ and we have then $H_\infty \leq_1 T_{\bar{\zeta}}$ (where \leq_1 stands for "is 1-1 reducible to").

However we can show that $T_{\bar{\zeta}} \leq_1 H_\infty$. This is shown in [25] in the proof of Lemma 2.6. More precisely, it is shown there that there is a (1-1) recursive function \bar{g} , so that for any limit ordinal $\lambda < \bar{\Sigma}$ that $\widehat{T}_\lambda = \bar{g}^{-1} \circ H_\lambda$. Here \widehat{T}_λ is the Liminf of the theories T_α for $\alpha < \lambda$: $\ulcorner \sigma \urcorner \in \widehat{T}_\lambda \leftrightarrow \exists \alpha < \lambda \forall \beta > \alpha (\beta < \lambda \rightarrow \ulcorner \sigma \urcorner \in T_\beta)$.

But for L_λ a model of Kripke-Platek set theory KP , it can be easily shown that this liminf theory is just the Σ_2 -theory T_λ . [For the 'harder' direction: as $L_\lambda \models KP$ it satisfies Π_2 -Reflection, that is for any Π_2 sentence τ , if $L_\lambda \models \tau$ then there are unboundedly many $\alpha < \lambda$ with $L_\alpha \models \tau$. Hence $\ulcorner \tau \urcorner \in$ could not be in \widehat{T}_λ .] As $L_{\bar{\zeta}}$ has a Σ_2 -end extension, it can be reasoned that it is such a model of KP , and thus $T_{\bar{\zeta}} = \bar{g}^{-1} \circ H_{\bar{\zeta}}$. As $H_{\bar{\zeta}} = H_\infty$ we get $T_{\bar{\zeta}} \leq_1 H_{\bar{\zeta}} = H_\infty$. Together with the last paragraph this yields $T_{\bar{\zeta}} \equiv_1 H_\infty$, that is recursive isomorphism, between these two sets.

All the above relatives uniformly to any starting hypothesis $H_0 = h$ for any h . We have then (with obvious adaptations of notation) that $T_\zeta^h \equiv_1 H_\infty^h$ with appropriate ordinals ζ^h, Σ^h .

In particular if h is recursive, or indeed hyperarithmetic, we shall have that $L_{\zeta^h}[h]$ is just L_ζ (and similarly for the Σ case). It can be shown that such h are outright Σ_1 -definable over L_ζ (or L_Σ) without the use of any further parameters. But this means that the Σ_2 -Theory of $L_{\zeta^h}[h]$, that is T_ζ^h , will be recursively obtainable via a (1-1) reduction procedure from T_ζ alone. Thus we shall have: $H_\infty^h \equiv_1 T_\zeta^h \leq_1 T_\zeta \equiv_1 H_\infty$. But conversely $T_\zeta \leq_1 T_\zeta^h$ because the 'pure' L -part of the Σ_2 -theory T_ζ^h is easy to extract: it is just those Σ_2 -sentences that do not mention the predicate for h anywhere. Thus $H_\infty \leq_1 H_\infty^h$, and we have the final recursive isomorphism: $H_\infty^h \equiv_1 H_\infty$. Q.E.D.