THE BULLETIN OF SYMBOLIC LOGIC
Volume 15, Number 4, Dec. 2009

GAMES FOR TRUTH

P. D. WELCH

Die Wahrheit liegt weder in der unendlichen Annährung an einer objektiv Gegebenes noch in der Mitte, sondern rundherum wie ein Sack, der mit jeder neuen Meinung, die man hineinstopft, seine Form ändert, aber immer fester wird.

R. Musil

Abstract. We represent truth sets for a variety of the well known semantic theories of truth as those sets consisting of all sentences for which a player has a winning strategy in an infinite two person game. The classifications of the games considered here are simple, those over the natural model of arithmetic being all within the arithmetical class of Σ^0_3.

§1. Introduction. These remarks examine the possibilities for representing the various commonly proposed solutions to the paradox of the liar, as *games* between players. We are dealing here with *semantical solutions* or approaches to those paradoxes. The most well known of those is due to Kripke [10]. In that paper he suggested building up values to a *partially defined* truth predicate, using (amongst other suggested methods) the three valued Strong Kleene truth tables. The monotonicity of the procedure implied that starting from the empty set the *least fixed point set* could be built up as an extension of the partially defined truth predicate, interpreted over some countable model $\mathcal{M}$. Thus, if we let the language of the model be $\mathcal{L} = \mathcal{L}_{\mathcal{M}}$, we may extend it to the language $\mathcal{L}^+$ by adding a predicate symbol $\dot{T}$ which is then partially interpreted as $T = (T^+, T^-)$, yielding an extension T^+ of true sentences and a so-called anti-extension T^- of false sentences. We then give, iteratively, successive progressive extensions $T_\alpha = (T^+_\alpha, T^-_\alpha)$ each containing all previous ones. A fixed point $T^\infty = (T^+_\infty, T^-_\infty)$ must then be reached when no further applications of the truth table rules

Received September 2, 2009.
Key words and phrases. Theory of truth, games, determinacy.
The author would like to thank the Templeton Foundation and the EPSRC of the UK for grants contributing to this research, and the Kurt Gödel Research Institute, Vienna, for its hospitality during the writing of this paper. He would like to thank the members of the Luxemburger Zirkel for their continuing criticism and support.

1079-8986/09/1504-0002/$2.80

add any further sentences of the language $\mathcal{L}^+$ to the extension or anti-extension. (See Burgess [1] or the monograph of McGee [12] for a complete description.) This monotone operation results in complete inductive sets of sentences via the use of any of the Kleene schemes and of (variations on) the supervaluation jump operator when starting with the standard model of the natural numbers $\mathbb{N}$ as $\mathcal{M}$.

At a similar period to Kripke, Hans Herzberger produced in [6, 7] a construction for studying the behaviour of diagonal liar, or liar-like, sentences that involved a*totally* interpreted truth predicate $\dot{T}$, but where the iterative stages did not build up any truth set in an accumulative fashion, but *revised* the extension at stage T_α into a new extension at $T_{\alpha+1}$. The intention was to provide a tool for the analysis of sentence behaviour along such a sequence of revisions. (Herzberger explained in the opening of [6] that he was not thinking of providing a full-blown theory of truth with these mechanisms, but intended them as diagnostic tools.) We might then start with an 'empty' or null assignment to the predicate at the initial stage: $T_0 = \varnothing$, corresponding in fact to assigning *false* to all sentences of $\mathcal{L}^+$; 'iterate' by applying the standard Tarskian truths, and obtaining a sequence of predicates in this fashion. Thus σ is in $T_{\alpha+1}$ if and only if $\langle \mathcal{M}, T_\alpha \rangle \models \sigma$; the procedure was supposed to be carried through at limit stages by using a truth-parsimonious 'liminf' rule where by for limit ordinals λ:

$$T_\lambda = \mathrm{Liminf}_{\alpha \to \lambda} T_\alpha =_{\mathrm{df}} \bigcup_{\beta < \lambda} \bigcap_{\beta < \alpha < \lambda} T_\alpha.$$

The revision semantics then discusses three kinds of sentences: those σ that are *stably true* (*stably false*)—meaning that $\exists\alpha\forall\beta(\alpha < \beta \longrightarrow \sigma \in T_\beta)$ ($\exists\alpha\forall\beta(\alpha < \beta \longrightarrow \sigma \notin T_\beta)$ respectively) and those that are *paradoxical* namely $\forall\alpha\exists\beta(\alpha < \beta \wedge \sigma \in T_{\beta+1} \backslash T_\beta)$. As is easily argued there is a "*stability point*": a countable ordinal $\zeta = \zeta(\mathcal{M})$ so that $\sigma \in T_\zeta \longleftrightarrow \sigma$ is stably true; and hence a *stability set* $T_\infty = T_\zeta$.

One may try to understand these processes through different means. One may try and *axiomatise* a theory dealing with a truth predicate to attempt to capture the semantic intention. The theory KF of Feferman is an axiomatisation of Kripke's strong Kleenean scheme. (See again [12]). The naturalness of the axiom set gives a successful description or fit to Kripke's original semantical theory. (However Halbach and Horsten have recently argued that since the Kripkean theory is one of partial interpretation of a $\dot{T}$-predicate, then this should be better performed in a *partial logic* which they do in [5]. KF was originally developed in a classical framework). In [2] Cantini develops a theory which he calls VF to try and axiomatise a Kripkean approach using supervaluations. He has also asked whether there is any axiomatisation possible for the stable truth set T_∞ arising from Herzberger's theory. We are somewhat doubtful of the possibilities for this here. Part of this doubt comes from the mathematics, and derives firstly from the strength of the statement that any revision sequence stabilizes (this requires a not insubstantial piece of second order number theory to establish), and secondly from the reflecting properties of the particular level of the Gödel hierarchy where the stability set occurs. The more serious doubt is philosophical, and comes from the lack of any motivation based in truth-theoretic principles for the choice of this (or any other current) limit rule. We turn to this issue in the final Section.

However a second means to try and understand the various semantical approaches is through the medium of *games*, meaning the notion of *two person perfect information* (*Gale–Stewart*) *games*. These games are a familiar tool to descriptive set theorists. Martin has shown [11, p. 417] how to illuminatingly represent the Strong Kleene fixed point over a countable model as the set of sentences for which a player has a strategy in an open game. He shows:

A Kleenean fixed point game (Martin). *There is a game G_φ for the least fixed point $T_\infty = (T^+_\infty, T^-_\infty)$ of the strong Kleenean scheme so thatPlayer I has a winning strategy in G_φ if and only if $\varphi \in T^+_\infty$; Player II has a winning strategy if and only if $\varphi \in T^-_\infty$; such winning strategies, if they exist, result in games of finite length; if neither player has a winning strategy, then, $\varphi \notin T^+_\infty \cup T^-_\infty$, and play may continue for infinitely many steps; neither player wins the game and thus it can be declared a draw.*

The appeal of these games is partly that the moves in the game reflect the compositionality inherent in the Strong Kleene Truth Tables and hence ultimately in the fixed points that arise. They thus allow a new epistemological slant on the semantically defined fixed point.

The task here is to provide games for the supervaluational fixed point theory (which we define below but for further discussion see [12]), but principally for the stable theory of a Herzberger sequence (whose definition we gave above).

We have not seen a similar account of a game devised specially for supervaluation fixed points—although such is certainly possible on general grounds, as the least supervaluational fixed point over $\mathbb{N}$ is a complete Π^1_1-set, and as such is representable by an open game formula (by a result of Svenonius for $\mathcal{M} = \mathbb{N}$, and by Moschovakis [14] for more general countable acceptable structures). Such games may well be known to others. They can not be expected to be "compositional" in the same way as for Strong Kleene, given the non-compositional nature of the supervaluation process. The formulation of the game here is thus more redolent of this process which involves looking at all possible 'completions' of the theory's extension built so far.

In the third section, we give a game, G^{sv}_φ, mirroring the statement above for Martin's game in terms of players' winning strategies and sentences in the fixed point. A variation on this game is shown to give a similar characterisation of the *grounded dependency* fixed point of Leitgeb [8].

In Section 4 we address the main problem here of providing a game representation of the Herzberger stability set. Such games can no longer be represented as *open* games, (where winning runs of play for either of the players are given by an $\exists$ formula, or Σ^0_1 in the arithmetical hierarchy if we are considering $\mathcal{M}$ as the standard model of arithmetic); we discuss the matter further there, but on general grounds the payoff sets for such games can not be even $\exists\forall$, although they are $\exists\forall\exists$. In terms of the usual Levy hierarchy of first order formulae, these are thus best possible results. Unlike the other games discussed here, play is not effectively over after a finite number of moves if one of the players has a winning strategy: play must in general continue for infinitely many moves whether there is a winner or it is a draw. We have then a further way of looking at, or measuring, the size/complexity of the most basic sequence, the null sequence of the Herzbergerian revision theory: it is at the level of that of determinacy of Σ^0_3 games. That is lucky: determinacy for such games is at least provable in second order number theory. Had it been Σ^0_4 then such determinacy is not provable in full second order number theory (by results of Friedman [4] and Martin). It should be stressed that the game given in detail here, although stated in terms appropriate to Herzberger sequences, has variants for certain quasi-inductive definitions over appropriate countable ground models $\mathcal{M}$. The pertinant feature is the stabilisation of values by a liminf rule (together with a certain "universality" of the inductive process). In particular it can be adapted for the notion of *ultimate truth* of Field [3]. There a three valued semantics is given using a model also constructed by such a quasi-induction (albeit not an arithmetical one). We comment on this game below at Remark 3.

The final section also makes some remarks, and tries to draw some conclusions, about the possibility of finding open game representations of T_ζ in a logic with generalised quantifiers, which, speculatively, could be germane to an axiomatisation of either of the theories of Herzberger or Field.

§2. A Kleenean fixed point game. Given a countable first order structure $\mathcal{M}$ with a suitable first order language $\mathcal{L}_{\mathcal{M}}$, we may extend the language to $\mathcal{L}^+$ to contain an additional unary predicate symbol $\dot{T}$. We shall assume that every element u of the domain of $\mathcal{M}$ has a name $\dot{u}$ in the language $\mathcal{L}_{\mathcal{M}}$. In $\mathcal{L}^+$ consider the least fixed point of the strong Kleene three valued logic built up from the empty extension for $\dot{T}$: starting with $T_0 = (T_0^+, T_0^-)$ with both $T_0^{+/-} = \varnothing$ this least fixed point $T_\infty = (T_\infty^+, T_\infty^-)$ is thought of as built up in ordinals stages $T_\alpha = (T_\alpha^+, T_\alpha^-)$

We describe the game due to Martin [11] that produces for $\mathcal{M}$, a representation of those sets of sentences that are true or false in $\mathcal{M}$ with the fixed point partially interpreting $\dot{T}$. Let φ be a sentence of $\mathcal{L}^+$. Players *I* and *II* then make moves which simply add a single sentence from $\mathcal{L}^+$ to a list. Play does not strictly alternate between players with every move, but the player currently listing elements from the language continues to do so, until the rules below oblige them to halt and the other player to take over.

The *Rules* are as follows: Player I starts and the first move must be the sentence φ.

If Player J (where J is one of I, II) in the last move added τ to the list, then, depending on τ, the Rules stipulate if:

$\tau = (\chi \vee \psi)$ then in the next move J moves again, playing either χ or ψ; *or* if

$\tau = \exists v \psi(v)$ then in the next move J moves again, and must play $\psi[v \backslash \dot{b}]$ for some constant $\dot{b}$ in the language $\mathcal{L}_{\mathcal{M}}$; *or* if

$\tau = \neg\psi$ then in the next move the other player must play ψ; *or* if

$\tau = \dot{T}(\ulcorner\psi\urcorner)$ then in the next move J must play ψ; *or* if

τ is some atomic sentence, then play halts.

The *Winning Condition* if this final case occurs, is that J wins if τ is true in $\mathcal{M}$. If not the other player wins. However if this final case does not occur then the play will continue for infinitely many moves, and then the game is declared a draw.

Notice that play only swaps to the other player if a negation is played. It is not hard to see that the rules conform to the entries in the Strong Kleene truth tables, and thus the game fulfils its task. Martin also points out some variations on the winning conditions, that allow one to have a theory similar in spirit to Yablo's theory of [17]. The very simple nature of these games, reflecting as they do the structure of the logic, makes a very attractive characterisation of this minimal fixed point. We shall see that even for the least supervaluational fixed point, although the complexity of such (in analytical terms) is the same as that for the Strong Kleene fixed point, the game itself requires a more substantial description.

§3. A supervaluational fixed point game. Again let $\varphi \in \mathcal{L}^+$ for some countable model $\mathcal{M}$ with $\mathcal{L}^+$ just as above. For ease of exposition we assume that $\mathcal{M}$ contains a name $\ulcorner\sigma\urcorner$ for each sentence $\sigma \in \mathcal{L}^+$.
There is a game G^{sv}_φ for the least fixed point $T_\infty = (T^+_\infty, T^-_\infty)$ of the supervaluational scheme so thatPlayer I has a winning strategy in G^{sv}_φ if and only if $\varphi \in T^+_\infty$; Player II has a winning strategy if and only if $\varphi \in T^-_\infty$; such winning strategies, if they exist, result in games of finite length; if neither player has a winning strategy, then, $\varphi \notin T^+_\infty \cup T^-_\infty$, and play may continue for infinitely many steps, and then the game is declared a draw.

We shall first define a simpler two person perfect information game G^+_φ played between Player I who we shall enliven with the name *Ulrich*, and Player *II*, whom we shall call *Agathe*, now moving in strict *rounds*, the k'th round will consist of the plays i_k, m_k respectively from $\mathcal{L}^+$. This game will provide one half of the above game $G^{\mathrm{sv}}_{\varphi}$. (It is simpler, and less cluttered with notation, to slice the game up into two halves and prove the requisite Lemma 1 below for this; then the game $G^{\mathrm{sv}}_{\varphi}$ can be assembled from the game G^{+}_{φ} and an easily understood dual game.)

Before giving a more formal definition, we give the original motivation for G^{+}_{φ}.

If she plays well in the game, Agathe is attempting to play a sequence Σ of sentences $(m_0 = \neg\varphi, m_1, \dots)$ of the language $\mathcal{L}^{+}$ such that, if Ψ is the extension of $\dot{T}$ so listed, ($i.e.$, $\Psi = \{\ulcorner\sigma\urcorner | \exists i\ m_i = \text{“}\ulcorner\sigma\urcorner \in \dot{T}\text{”}\}$) then Σ is the complete $\mathcal{L}^{+}$-theory of $\langle \mathcal{M}, \Psi \rangle$.

We arrange this as follows: we may think of the game as a series of queries by Ulrich as to whether sentences $m_k = v$ are, or are not, in the Σ that she is constructing (including of course the sentences as to whether “$\ulcorner\sigma\urcorner \in \dot{T}$”); Agathe will be giving, in effect, yes or no answers to these queries, when she responds by agreeing with, or contradicting each of his regular moves in turn; this, sometimes together with some additional information. During the course of the game Ulrich may issue a *challenge* to Agathe's earlier replies about the extension of her Ψ. He may either attempt to point out that her extension Σ is inconsistent, or he may directly challenge an assertion by Agathe that some σ is, or is not, in Ψ. In this latter case, the emphasis of the game G^{+}_{φ} is shifted and a subgame G^{+}_{σ} is initiated.

Now a more formal description:

$$\begin{array}{lllll} \textit{Ulrich I}: & i_0 & & i_1 & \dots \\ \textit{Agathe II}: & & m_0 & & m_1 \dots \end{array}$$

Rules for II in G^{+}_{φ}: Each m_j must be a sentence from $\mathcal{L}^{+}$; m_0 must be $\neg\varphi$.

Rules for I in G^{+}_{φ}: Each move i_j must be one of three types: *either* a sentence from $\mathcal{L}^{+}$ *or* a finite list of sentences from $\mathcal{L}^{+}$ *or* a pair (Flag, σ) where *Flag* is any object not in $\mathcal{L}^{+}$, and σ is a sentence of $\mathcal{L}^{+}$. The initial move i_0 must be the pair (Flag, φ).

This completes the rules for moves in the game. If a *Rule* is broken the game halts immediately, and the player breaking the rule loses. We shall refer to the three types of move of I as being either a *regular move*, a *consistency challenge*, or a *game challenge*. (Thus the initial move i_0 is a game challenge, and initates the whole game. It is of this peculiar form just for convenience, so that the whole game resembles later parts that are similarly initiated by later game challenges.)

Winning Conditions for the game at round k: We may assume then no one has broken a *Rule* at any previous round.

Winning Conditions for Player I at Round k: (i) Suppose Player I plays a regular move “i_k”. If i_k is a sentence σ not of the form $\exists v_i \psi(v_i)$ of $\mathcal{L}^{+}$ and if Player II's reply move m_k is neither “σ” nor “$\neg\sigma$” then II loses immediately.

If however, σ is of the form $\exists v_i \psi(v_i)$ and m_k is not of the form $\psi[\dot{b}]$ nor of the form $\neg\sigma$, then again *II* loses immediately.

(This can be interpreted as *II* asserting "yes, $\sigma \in \Sigma$" or, "no, $\neg\sigma \in \Sigma$" depending on the response m_k she plays. If however σ is of the form "$\exists v_0 \psi(v_0)$" and she wishes to answer "yes", then she must also choose some object $b \in \mathcal{M}$ to witness this, and play "$\psi[\dot{b}]$".)

(ii) If $i_k = \sigma$ is an atomic sentence of $\mathcal{L}_{\mathcal{M}}$, then her answering move m_k of σ or $\neg\sigma$ must be true in $\mathcal{M}$, or else she forfeits the game immediately.

(iii) Suppose player *I* makes a *consistency challenge* to Player *II* by playing $i_k = \vec{\tau}$ sentences. Suppose $k' < k$ is the last round in which Ulrich issued a game challenge (k' is always defined but may be zero). If $\vec{\tau}$ consists of a correct proof of a contradiction using only as hypotheses moves that *II* has played since (and including) round k' then he wins outright.

(iv) Suppose he makes a game challenge $i_k = (\mathit{Flag}, \sigma)$. If *II* responds with m_k different from $\neg\sigma$ then again he wins immediately.

This completes the winning conditions for I at a round k.

Winning Conditions for II at round k: *Either* Ulrich makes a consistency challenge at round k which fails to prove Agathe's moves since the last game challenge are inconsistent *or* he makes an *improper* game challenge. These are of the form $i_k = (\mathit{Flag}, \sigma)$ (respectively $(\mathit{Flag}, \neg\sigma)$) where Agathe has *not* played $\ulcorner\sigma\urcorner \notin \dot{T}$ (respectively $\ulcorner\sigma\urcorner \in \dot{T}$) since the last game challenge. In either case as soon as Ulrich makes a failing challenge of either of these types, the game is over with Agathe declared the winner.

This completes the ways a game is won at a finite round. If the game continues for infinitely many moves then *II*, Agathe, wins.

We explain the game challenge moves which are those of the form $i_k = (\mathit{Flag}, \sigma)$ (for $k > 0$). If Ulrich makes a proper game challenge, then the game does not yet terminate with a winner, but continues just as if i_k was the first play in a game defined in exactly the same way as G^+_φ but with the original φ instead replaced by σ, if "$\ulcorner\sigma\urcorner \notin \dot{T}$ " was the earlier play of *II* challenged, and replaced by $\neg\sigma$ if it was "$\ulcorner\sigma\urcorner \in \dot{T}$ " that was challenged. We may imagine that play in G^+_φ concentrating on φ is abandoned, and instead play starts afresh in G^+_σ (or in $G^+_{\neg\sigma}$ if " $\ulcorner\sigma\urcorner \in \dot{T}$ " was the sentence challenged). It should be noted that consistency challenges by I can only be made up from hypotheses that *II* has made since the last game challenge. Hence the moves of *II* prior to the last game challenge are no longer relevant, and we may indeed think of the initial part of the run of play up to the last game challenge as being, in effect, discarded. We shall thus think, and talk in this case (with some terminological abuse), of G^+_σ as a "subgame" of G^+_φ, which Ulrich has caused to be initiated through a game challenge in the course of playing G^+_φ, but in any case it is defined from σ (or $\neg\sigma$) just as G^+_φ was defined from φ.

To recapitulate: Ulrich wins G^+_φ if and only if (i) Agathe makes a mistake on the basic rules of the game, or (ii) he makes a successful consistency challenge to the sentences Agathe is playing. Thus, for Ulrich to win, the overall game must be finite in length. Agathe thus wins precisely when: either (i) Ulrich breaks a basic rule; or (ii) he makes a false accusation that her list is inconsistent, or he makes an improper game challenge; or (iii) she manages to make infinitely many moves in G^+_φ (or one of the above subgames G^+_σ); or lastly (iv) through the whole course of play he makes infinitely many game challenges (thus initiating infinitely many subgames).

We recall the definition of the least supervaluation fixed point (T^+_∞, T^-_∞), (we ignore corner quote marks for readability).

$T^{+/-}_0 = \emptyset$; if $Lim(\lambda)$ then $T^{+/-}_\lambda = \bigcup_{\alpha<\lambda} T^{+/-}_\alpha$, and at successor steps one takes:

$$T^+_{\alpha+1} = \bigcap_{\{T\,:\, T\cap T^-_\alpha=\emptyset, T^+_\alpha\subseteq T\}} \{\varphi \in \mathcal{L}^+ \mid \langle \mathcal{M}, T\rangle \models \varphi\},$$

$$T^-_{\alpha+1} = \{\neg\varphi \in \mathcal{L}^+ \mid \varphi \in T^+_\alpha\}.$$

$T_\infty = \langle T^+_\infty, T^-_\infty\rangle$ is then given by the least α so that $\langle T^+_\alpha, T^-_\alpha\rangle = \langle T^+_{\alpha+1}, T^-_{\alpha+1}\rangle$.

The intersection defining $T^+_{\alpha+1}$ is taken over all T which are said to be *compatible* with $\langle T^+_\alpha, T^-_\alpha\rangle$. (Other variants are possible here.)

LEMMA 1. *I has a winning strategy in* $G^+_\varphi \iff \varphi \in T^+_\infty$ *where* $T_\infty = \langle T^+_\infty, T^-_\infty\rangle$ *is the least Kripkean supervaluation fixed point.*

PROOF. ($\Longleftarrow$) Suppose $\varphi \in T^+_{\alpha_0+1}\backslash T^+_{\alpha_0}$ (we'll say that "rk$(\varphi) = \alpha_0$"). Using the nomenclature from the game motivation above, Agathe attempts to play out a consistent $\Sigma = \Sigma(\varphi)$, and $\Psi = \Psi(\varphi)$, an included extension for $\dot{T}$, so that $\langle \mathcal{M}, \Psi\rangle \models \Sigma$. Additionally, as an example, if φ is of the form $\exists v_0 \psi(v_0)$ then for every constant $\dot{b}$ in $\mathcal{L}^+$, she must be ensuring that $\langle \mathcal{M}, \Psi\rangle \models \neg\psi[\dot{b}]$. In every case, if she is to try and win, Ψ cannot be compatible with $(T^+_{\alpha_0}, T^-_{\alpha_0})$ as $\varphi \in T^+_{\alpha_0+1}$ whilst her first move m_0 asserts that $\neg\varphi \in \Sigma$. Hence by that incompatibility, either:

(i) For some $\sigma_1 \in T^+_{\alpha_0}$, $\sigma_1 \notin \Psi$, or:
(ii) For some $\sigma_1 \in T^-_{\alpha_0}$, $\sigma_1 \in \Psi$.

As part of his strategy, I makes sure that he asks every possible sentence query of the form "$\tau \in \dot{T}$" during his course of play. (Indeed in this way I can ensure the completeness of the pertinent Σ of any infinite run of any subgame. This querying by I is not an essential feature of the game: we could instead have varied the basic rules and simply required II to list facts true or false from a priorly fixed enumeration of all sentences.) At some point then, for some such σ_1 as above, when queried on $?\sigma_1 \in \dot{T}?$, (that is when I plays "$\sigma_1 \in \dot{T}$" as i_k if Case (i) (or "$\sigma_1 \notin \dot{T}$" as i_k if Case (ii) respectively) holds, II plays as m_k "$\sigma_1 \notin \dot{T}$" (or "$\sigma_1 \in \dot{T}$" respectively). I on his next move can immediately call a game challenge ($i_{k+1} = (\mathit{Flag}, \sigma_1)$ (or $(\mathit{Flag}, \neg\sigma_1)$ respectively) in the appropriate numerology) to this statement, and the subgame $G^+_{\sigma_1}$ (or $G^+_{\neg\sigma_1}$ respectively) is initiated. The point is that $\sigma_1 \in T^+_{\alpha_0}$ (or $\sigma_1 \in T^-_{\alpha_0}$). In either case (with the obvious extension of notation) $\alpha_1 =_{\mathrm{df}} \mathrm{rk}(\sigma_1) < \alpha_0$, and the subgame now being initiated is being played on a formula of lower rank. The set $\Psi(\sigma_1)$ (or $\Psi(\neg\sigma_1)$) that *II* now tries to play out is incompatible with $(T^+_{\alpha_1}, T^-_{\alpha_1})$, and thus at some stage in this sub-game I may issue another game challenge on the assertion that some "$\sigma_2 \in / \notin \dot{T}$", and the resulting $\alpha_2 =_{\mathrm{df}} \mathrm{rk}(\sigma_2) < \alpha_1$. Proceeding in this way, as long as *II* does not lose through inconsistency, I eventually challenges with some σ_k with $\mathrm{rk}(\sigma_k) = 0$. However if $\sigma_k \in T^+_1$ then *every* Ψ is compatible with $(T^+_0, T^-_0) = (\varnothing, \varnothing)$ and *II* in her play of $G^+_{\sigma_k}$, simply cannot produce an extension Ψ for $\dot{T}$ and a *consistent* set of sentences Σ so that $\langle \mathcal{M}, \Psi \rangle \models \Sigma$, whilst $\langle \mathcal{M}, \Psi \rangle \models \neg\sigma_k$. (Note that the sets Σ being produced, if consistent, are true in the model $\langle \mathcal{M}, \Psi \rangle$, as we have required that whenever "$\exists v_0 \psi(v_0)$" is queried with affirmative response from *II* she must also provide evidence of the form $\psi[\dot{b}]$ for some constant $\dot{b}$ naming an object in the domain of $\mathcal{M}$.) She thus will end up losing $G^+_{\sigma_k}$ by breaking a basic rule.

($\Longrightarrow$) Suppose $\varphi \notin T^+_\infty$. We describe how Agathe can win. By hypothesis there is $\Psi = \Psi(\varphi)$ and with $\Psi \supseteq T^+_\infty, \Psi \cap T^-_\infty = \varnothing$, with $\langle \mathcal{M}, \Psi \rangle \models \neg\varphi$. Whenever she is queried Agathe consults the above model and gives the appropriate reply. She thus will not lose G^+_φ on consistency grounds, nor by breaking any other basic rule. If she is challenged on her assertion "$\sigma_1 \notin \dot{T}$" then $\sigma_1 \notin T^+_\infty$. The subgame initiated is $G^+_{\sigma_1}$, but as $\sigma_1 \notin T^+_\infty$ she is no worse off than before and can play just as well here using an appropriate $\Psi = \Psi(\sigma_1)$, and the new model $\langle \mathcal{M}, \Psi(\sigma_1) \rangle \models \neg\sigma_1$. If she is challenged on "$\sigma_1 \in \dot{T}$" then the subgame $G^+_{\neg\sigma_1}$ is initiated; but $\neg\sigma_1 \notin T^+_\infty$, and she can continue in the same way. Clearly she can keep this up no matter how many challenges that Ulrich issues, and she will ultimately win. ⊣

For the full game G^{sv}_φ we double up the roles we have described: now both players must produce sequence of sentences $\Sigma(I)$, $\Psi(I)$ and $\Sigma(\mathit{II})$, $\Psi(\mathit{II})$ with I now trying to ensure the truth of φ in $\langle \mathcal{M}, \Psi(I) \rangle$ and *II* as before its falsity in $\langle \mathcal{M}, \Psi(\mathit{II}) \rangle$. *II* now can also make appropriate queries of I's sentence moves. If neither have a winning strategy, then this is because $\varphi \notin T^+_\infty \cup T^-_\infty$, and we have the above properties of the full game. We wish to spare the reader the formal definition here, one can think of the full game as being played on two boards: one for G^+_σ, as above, and on the second board $G^+_{\neg\sigma}$, but on this second board the roles of the players I and *II* are switched. For G^{sv}_φ, play is amalgamated by having one round played on the first board, then a round on the second, and so on, back and forth.

A dependency game. We now informally sketch the variant of the game which characterises Leitgeb's (*dependency*) *grounded sentences*

$$\Phi_\infty = \bigcup_{\alpha<\omega_1^{ck}} \Phi_\alpha.$$

(We give but a very brief review of this notion here and refer the reader to Leitgeb [8] for a full description). Here we set $D(\Psi)$ to be the set of sentences that *depend on* Ψ, where φ *depends on* Ψ if $\forall\Psi_0(\langle\mathcal{M},\Psi_0\rangle \models \varphi \leftrightarrow \langle\mathcal{M},\Psi\cap\Psi_0\rangle \models \varphi)$. It is shown that D is a monotone operator, and hence we may define by induction $\Phi_0 = \varnothing$, $\Phi_{\alpha+1} = D(\Phi_\alpha)$, and taking unions at limit λ to form Φ_λ. The fixed point Φ_∞ (which is $\Phi_{\omega_1^{ck}}$, for the standard model of arithmetic $\mathcal{M} = \langle\mathbb{N},+,\times,0,',\dots\rangle$, and which is a complete Π^1_1 set) is then the set of sentences whose truth value depends on *non-semantic states of affairs*.

The similarity to the notion of supervaluation scheme is apparent and so we may form a similar game to G^+_φ above for sentences φ appropriate for $\mathcal{M}$. G^*_φ is played in a similar fashion, Agathe must produce now two extensions Ψ_0,Ψ_1 and complete sets of sentences Σ_0,Σ_1 about the models $\langle\mathcal{M},\Psi_0\rangle$, $\langle\mathcal{M},\Psi_1\rangle$ respectively. (Say she attends to queries about Σ_0 in even numbered rounds, and to Σ_1 on odd numbered ones). She is trying to ensure that $\langle\mathcal{M},\Psi_0\rangle \models \neg\varphi$ whilst $\langle\mathcal{M},\Psi_1\rangle \models \varphi$. Thus when queried about φ must answer accordingly. The basic rules and winning conditions are the same *mutatis mutandis* and *I* may still challenge on grounds of consistency, with the same outcomes. If however *II* has earlier asserted "$\sigma \in \Psi_0$" and "$\sigma \notin \Psi_1$" (or *vice versa*) then *I* may issue a game challenge, and the subgame G^*_σ is initiated (again the subgame is the same as the game: she tries to produce two extensions $\Psi_0(\sigma),\Psi_1(\sigma)$ with $\langle\mathcal{M},\Psi_0(\sigma)\rangle \models \neg\sigma$, whilst $\langle\mathcal{M},\Psi_1(\sigma)\rangle \models \sigma$). As before, if neither player messes up their basic rules, then if the overall game lasts for infinitely many stages *II* wins.

LEMMA 2. *I has a winning strategy in $G^*_\varphi \iff \varphi \in \Phi_\infty$.*

PROOF. ($\Longleftarrow$) Suppose now $\varphi \in \Phi_{\alpha+1}\backslash\Phi_\alpha$; we set $\mathrm{rk}^*(\varphi) = \alpha$. By the definition of dependency, if $\langle\mathcal{M},\Psi_0\rangle \models \neg\varphi$ whilst $\langle\mathcal{M},\Psi_1\rangle \models \varphi$, then $\Psi_0\cap\Phi_\alpha \neq \Psi_1\cap\Phi_\alpha$. Hence if *II* is trying to produce such Ψ_0,Ψ_1 for some $\sigma_1 \in \Phi_\alpha$ she must answer "$\sigma_1 \notin \Psi_0$" and "$\sigma_1 \in \Psi_1$" (or *vice versa*) when queried by Ulrich, and the latter may now issue a challenge and the game proceeds to the subgame $G^*_{\sigma_1}$. Now of course $\alpha_1 =_{\mathrm{df}} \mathrm{rk}^*(\sigma_1) < \alpha_0$, and if no one messes up their basic rules, we arrive as before at the situation of playing in $G^*_{\sigma_k}$ where $\mathrm{rk}^*(\sigma_k) = 0$. But here, as $\Phi_0 = \varnothing$, for every Ψ,Ψ', $\langle\mathcal{M},\Psi\rangle \models \sigma_k \longleftrightarrow \langle\mathcal{M},\Psi'\rangle \models \sigma_k$ and so Agathe will lose at this point.

($\Longrightarrow$) Suppose $\varphi \notin \Phi_\infty$. Then there exists Ψ_0 with $\langle\mathcal{M},\Psi_0\rangle \models \varphi \longleftrightarrow \langle\mathcal{M},\Psi_0\cap\Phi_\infty\rangle \models \neg\varphi$. Let $\Psi_1 = \Psi_0\cap\Phi_\infty$. She may then play out the complete theories of the two models $\langle\mathcal{M},\Psi_0\rangle$, $\langle\mathcal{M},\Psi_1\rangle$. If she is challenged at some point on her assertions amounting to $\sigma \in \Psi_0 \backslash \Psi_1$, then even when the subgame G^*_σ is initiated we have $\sigma \notin \Phi_\infty$ and she is no worse off than before, and can continue in the same fashion using some Ψ_2 with $\langle \mathcal{M}, \Psi_2 \rangle \models \sigma \longleftrightarrow \langle \mathcal{M}, \Psi_2 \cap \Phi_\infty \rangle \models \neg\sigma$. If in G^*_φ (or some initiated subgame) Ulrich makes no challenge, she can play out to the end without breaking any basic rule.

§4. A game for Herzbergerian revision sequences. We now see how one can formulate a game for a Herzbergerian style revision sequence. We shall assume a further condition on our countable model $\mathcal{M}$, that it should be *acceptable* in the sense of Moschovakis (this seems to be needed in order for the proof of the technical lemma below to go through). Acceptability is, *inter alia*, a requirement that we can define over the structure $\mathcal{M}$ an appropriate *coding scheme* for finite sequences of elements from the domain of $\mathcal{M}$. (We do not need these details here as we shall not be proving the Lemma.) Nothing will be lost if the reader keeps in mind the standard model $\mathbb{N}$ as $\mathcal{M}$. In the sequel we shall assume the sentences of $\mathcal{L}^+$, $Sent_{\mathcal{L}^+}$, are disjoint from $\mathbb{N}$. We let $\langle -, - \rangle : \mathbb{N}^2 \leftrightarrow \mathbb{N}$ be any fixed recursive bijection.

As outlined in the Introduction we let T_α be the extension of the $\dot{T}$-predicate at the α'th stage of a Herzberger Revision sequence starting out with $T_0 = \varnothing$. Then, σ is in the revised extension at stage $\zeta = \zeta(\mathcal{M})$ iff σ is stably true; similarly at such a point we have $\neg\sigma \in T_\zeta \longleftrightarrow \neg\sigma \in T_{\mathrm{On}}$, which expresses that σ is stably false iff $\neg\sigma$ is in the extension T_ζ. Thus the overall status of sentences in the revision sequence is mirrored precisely at this stage $\zeta(\mathcal{M})$. ([1] identified the ordinal $\zeta(\mathbb{N})$ and this can be generalised.) From $\zeta(\mathcal{M})$ onwards the whole process is cycling through a fixed sequence of extensions. The next ordinal where the extension T_ζ reoccurs is called $\Sigma = \Sigma(\mathcal{M})$. The game presentation relies on the following new result on such sequences:

LEMMA 3. *If $\beta < \gamma < \Sigma$, then in the Herzberger revision sequence $T_\gamma \nsubseteq T_\beta$.*

The proof of this is technical and indirect. In order to show this, one first demonstrates how the whole Herzberger sequence over $\mathcal{M}$ up to stage γ can be uniformly reconstructed from knowledge of the γ'th truth set T_γ alone. Moreover one must additionally show that it is possible to do this in a sufficiently uniform manner to ensure the non-decreasing nature of the truth sets as stated in the Lemma's conclusion. It will appear elsewhere. Using the above Lemma we can give the following:

The stability game. *There is a game $\widetilde{G}^H_\varphi$ for the least stability point $T_\infty = T_\zeta$ of the Herzbergerian revision sequence starting out with $T_0 = \varnothing$ so that Player I has a winning strategy in $\widetilde{G}^H_\varphi$ if and only if φ is stably true*, ie, $\varphi \in T_\infty$; *Player II has a winning strategy if and only if φ is* stably false, *that is when $\neg\varphi \in T_\infty$; if neither player has a winning strategy, then φ is paradoxical, and then the game is declared a draw.*

The (*One-sided*) *Stability Game* G^H_φ then is again one 'half' of the required game in which as above, for a sentence φ, I, as Ulrich, tries to show that φ is stably true. II, Agathe, tries simply to defeat this, without being obliged to demonstrate that φ is stably false. We describe then G^H_φ. Lemma 4 below then will demonstrate the outcomes of this game.

Aim for II: To play her moves $(n_0, n_1, \dots)$ in order to code a revision sequence along a linear ordering in which φ is not stably true.

Aim for I: To try and pick out, if possible, an infinite descending chain through II's linear ordering, thereby demonstrating that she has not produced a genuine revision sequence. To help him do this, he is allowed to ask certain queries of II about her final revision sequence, that she must honestly answer and keep to.

It will turn out that if II is to win, remarkably I can, by asking the right questions, force II to play a wellorder, and thus a true revision sequence—this even though there is no mention of wellorders in the game's definition. (Indeed the payoff set when the game is coded over integers, is not Π^1_1—which it would have to be at best, were the game's conditions to explicitly require wellorders—but is merely Σ^0_3.)

The game is played as usual with I, II strictly alternating moves. We denote I's moves as $(m_0, m_1, \dots)$ and II's as $(n_0, n_1, \dots)$. The moves (m_k, n_k) are the moves in *round* k.

The Rules for II: For any k: $n_{2k} \in \mathbb{N}$; $n_{2k+1} \in Sent_{\mathcal{L}^+}$.

The Rules for I: (i) m_i may be a "Pass"; (ii) m_i may be an element of $\mathbb{N}$; (iii) m_i can be a triple $\langle a, b, \sigma \rangle$ with $a, b \in \mathbb{N}$ and $\sigma \in Sent_{\mathcal{L}^+}$. This latter kind will be called a *query* move for II.

This completes the basic *Rules* for the players, and if a player breaks one of them the game is immediately halted and is then forfeit to the other player.

The reason for naming the third kind of move that I makes as a 'query' is that it will have the following interpretation for II:

$$\text{Is it true that } a, b \in \text{Field}(\prec) \wedge \forall c(a \prec c \prec b \longrightarrow \sigma \in T_c)? \qquad (*)$$

Agathe will be required to give honest answers to any such query.

The Winning Conditions for I: Ulrich will be declared the winner if Agathe fails in any of the tasks (a)–(e) below.

(a) Her even moves (n_{2k}) must code a *discrete linear order* $\prec$ of $\mathbb{N}$: where we set $a \prec b \leftrightarrow \exists k(n_{2k} = \langle a, b \rangle)$. It is required that $\prec$ have end points o, s, and a further distinguished point $z \in [o, s]_\prec$. We let Field$(\prec)$ denote $\{a \mid \exists b(a \prec b \vee b \prec a)\}$.

(b) Using the odd round moves, (n_{2k+1}), she must play out a *complete theory* $T_a \subseteq \text{Sent}_{\mathcal{L}^+}$ to each $a \in \text{Field}(\prec)$. (The details of how this and other coding matters are priorly fixed up is immaterial; it suffices to do this in some effective manner: let us say that $a \in \text{Field}(\prec) \to T_a = \{\sigma \mid \exists u(n_{3^a.5^u} = \sigma)\}$).

(c) Both the following must hold: if t' is the $\prec$-successor of t, then $T_{t'} = \text{Th}(\langle \mathcal{M}, T_t \rangle)$. If t is a $\prec$-limit, then $T_t = \liminf_{t' \prec t} T_{t'}$.

(d) Both z, s are limit points of the ordering $\prec$; $T_z = T_s$ and $\varphi \notin T_z$.

(e) "*Committment Condition" for II*: If m_i is a query such as $(*)$, then she must play an answer at round $7i$ with n_{7i} a "1" (read: "Yes") or a "0" (read: "No"). Further, *II* must honour each such answering play she makes, as a committment to her truth sets in the final revision sequence she will build along $\prec$.

List making Condition for I: If m_k is neither a pass nor a query, but $m_k = \langle r, a, b \rangle$ then there must be $2j < k$ with $n_{2j} = \langle a, b \rangle$; and if on the last round k' with $m_{k'}$ which was a move for Ulrich of the form $\langle r, d, c \rangle$ (if any) then $d = b$. If these conditions are fulfilled we shall say that "m_k is a correct entry."

• *Interpretation*: We think of the integer moves m_k, that are of the form $\langle r, c, d \rangle$ as being entries on the r'th list L_r of attempts by I to list an infinite descending chain through the underlying Field$(\prec)$ of Agathe's revision sequence. We shall allow him however infinitely many attempts at writing such a list L_r (as r varies), and the above mechanism simply organises such list-writing moves $m_k \in \mathbb{N}$. By making such a move m_k I has simply adverted to the fact that he has extended the list L_r below $b \prec c$ with an element a satisfying $a \prec b \prec c$ which *II* has already announced will be $\prec$-below b. He is not forcing her into making any committments: *II* has already revealed this much of the ordering.

Winning Conditions for I (concluded):
I wins iff
Either II breaks any of her rules, or leaves one of the conditions (a)–(e) *unfulfilled, or* $\exists r$[*I makes infinitely many correct entries on list* L_r].

The first two disjuncts in the above can be shown to amount to an $\exists\forall\exists$ condition on the set of all runs of play. (All the conditions (a)–(e) apart from (c) are $\forall\exists$ conditions. The requirement that if t is a $\prec$-limit, then $T_t = \liminf_{t' \prec t} T_{t'}$ is *prima facie* $\forall\exists\forall$, but in fact the requirement can be enforced in any case through the committment condition. Hence it may be dropped and (c) then becomes an $\forall\exists$ condition too.) The entry in square brackets is an $\forall\exists$ condition on the run of play. The third disjunct thus adds an $\exists\forall\exists$ condition, to be a win for I. G^H_φ is thus overall an $\exists\forall\exists$ game.

LEMMA 4. *Any sentence σ_0 is stably true under the Herzberger revision sequence with starting distribution of truth values $T_0 = \varnothing$, that is $\sigma_0 \in T_\infty$, if and only if I has a winning strategy in $G^H_{\sigma_0}$.*

PROOF. ($\Leftarrow$) Suppose $\sigma_0 \notin T_\infty$. Then in the "true revision sequence" of length $\Sigma + 1$, we see $\sigma_0 \notin T_\infty = T_\Sigma = T_\zeta$. *II* ignores I and plays out a set of integers on her even moves coding a wellordering of length $\Sigma + 1$, and attaching to appropriate integers a in the field of that wellordering truth sets $T_{\|a\|}$ (where $\|a\|$ is the rank of a in the wellorder) and obeying all the Basic Rules, answering I's questions truthfully, and so keeping, necessarily, to all committments and fulfilling all conditions (a)–(e). This way she wins: I can find no illfounded chains in $\prec$. Hence I can have no winning strategy.

($\Rightarrow$) Suppose $\sigma_0 \in T_\infty$. Could II nevertheless have a way of playing $G^H_{\sigma_0}$ and winning? We show how I can defeat any strategy of II. First note that her ordering $\prec$ cannot be a wellorder (because it would then in effect be a true revision sequence with a repeating pair of stability points, and we'd see "$\sigma_0 \in T_\infty$" in II's sequence—this alone would cause her to lose). So must $\prec$ be illfounded.

Let WFP($\prec$) denote the wellordered part of the linear ordering $\prec$ that II putatively plays out, and let β be its order type. In fact let us identify β with WFP($\prec$). Then: (1) Lim(β); (2) $\Sigma \notin$ WFP($\prec$) again for the reasons just mentioned.

A priori Ulrich has absolutely no idea what ordinal $\beta \leq \Sigma$ is being discussed here. He only knows (1) and (2). However let us suppose that he makes the *working assumption* that some fixed β truly will turn out to be (isomorphic to) WFP($\prec$). We shall see how, if this assumption were to be correct, that he could win. The method used relies on the following technical lemma:

LEMMA 5. *If a is a point of* Field($\prec$), *with $a \notin$* WFP($\prec$) *then $T_a \nsubseteq T_\beta$.*

This lemma in fact is the "non-wellfounded" version of the result about such revision sequences (which we shall not use directly in this argument) mentioned at the outset.

So, if we assume Lemma 5 proven, I additionally does know one more fact: he knows (3) that $T_s \nsubseteq T_\beta$. He may therefore wait until Agathe reveals some $\upsilon \in T_s \backslash T_\beta$ at some stage in time—which she must do sooner or later, as she must tell any basic fact about her sequence at some point since she has to play T_s as complete theories, if she is to win. He is then in business. He starts making queries of the form

Is $s' \prec s$ and $\forall t(s' \prec t \prec s \rightarrow \upsilon \in T_t)$?

Once he gets an affirmative answer to one such query (which again he must do, at some finite stage of the game, if $\upsilon \in T_s$ and II fulfils her conditions—by the 'liminf' nature of T_s) then he knows $s' \notin$ WFP($\prec$). (Since s' cannot be below β, for otherwise we'd have $\upsilon \in T_\beta$ (by the 'liminf' nature of T_β this time). He could then make an m_k listing move indicating $s' \prec s$, and start off a descending $\prec$-chain.

However we can repeat this: as $T_{s'} \nsubseteq T_\beta$, by the Lemma, he can, by consulting T_β again, wait until a similar situation as the one outlined above occurs again with a "Yes" committment, and another m_k move can be made indicating $s'' \prec s'$. Continuing in this way he can list an infinite $\prec$-descending chain.

The above discussion was based on the (probable) fiction that β is to be WFP($\prec$). However *if* this were to be the case, then I could defeat II. If we now let F: $\mathbb{N}\backslash\{0\} \leftrightarrow \mathrm{Lim}\cap(\Sigma+1)$ be a (1–1) enumeration of the limit ordinals below $\Sigma + 1$, then we have an enumeration of all possible working assumptions for I; we may thus let I make moves as listing entries on L_r with the working assumption that $\beta = F(r)$ is the wellordered part of the ordering $\prec$. By constantly shifting his attention infinitely often back and forth between all such working assumptions, I may play using them all, and thereby consulting all the T_β, within the infinite amount of time available during the one game. So if $\prec$ is not a wellorder, and Agathe fulfils all her conditions, there will be at least one list, namely L_r where $F(r)$ in the event turns out to be the wellfounded part of $\prec$, on which he can make infinitely many entries and thus win. $\dashv$ (Lemma 4)

The full game $\widetilde{G}^H_\varphi$ is the variant where, besides the above moves, I also is obliged to play out a linear ordering mirroring precisely the kind of moves II made, with attached complete theories, etc., excepting that I's putative revision sequence must demonstrate that φ is in fact stable. Agathe is obliged in her sequence to show that $\neg\varphi$ is stable. Simultaneously, as I did before, she has the opportunity to write countably many lists of potential descending chains through I's linear order. If φ is truly paradoxical then neither person has a winning strategy. Again this can be realised by amalgamating simultaneous play on two boards, on the first is played G^H_φ as above, and on the second $G^H_{\neg\varphi}$—with the roles of I, II in the latter interchanged.

Remark 1. That I be allowed to draw up infinitely many independent lists can be shown to be necessary. It might be thought that there could be an equivalent variant of the game with I only drawing up a single list, and that the use of infinitely many lists was a symptom of our inability to design the right game (thereby squandering a quantifier). However there is no such game—for which see the next remark. In essence (but still a rough approximation) the relationship between the different truth sets T_β for $\beta < \Sigma$ is so complex that one cannot continuously define a single game with one list whilst still enforcing that II produce a wellordered sequence; one has to consider the "infinitely many working assumptions at once" device as above.

Remark 2. If we start with the standard model $\mathcal{M} = \mathbb{N}$ then the complete runs of play of both players can be coded together as a sequence of integers y, and be classified within the arithmetical hierarchy. We then see that the conditions (a)–(e) on Agathe amount (with some care) to a Π^0_2 condition on y. The payoff set of sequences for Ulrich is then a Σ^0_3 set. It is a non-trivial fact that there is no simpler game (for example, with a Σ^0_2 payoff set) for a Herzberger sequence. Thus in terms of the classificatory Levy hierarchy by formulae, the games G^H_σ here are already at the lowest possible level of complexity.

Briefly we discuss the reasons for this. (The reader uninterested in these issues may simply avoid this paragraph.) Solovay (unpublished by him, but see [9, Thm. 2.5.2]) showed that winning stategies for all Σ^0_2 games occur at a level of the Gödel constructible hierarchy of constructible L-rank much less than ζ, the first repeat point of the Herzberger sequence, that is where $T_\zeta = T_\infty$. If we were able to find Σ^0_2 games to represent the Herzberger stability set T_ζ, then that set could be found at a constructible level much lower than ζ. This is impossible as the Herzbergerian T_ζ only appears first in $L_{\zeta+1}$. (A similar argument using a result of [13] shows that no Δ^0_3 game will do either as these also have strategies that occur well below ζ.) On the other hand [16] shows that there are levels of the L hierarchy beyond ζ whose truth sets can be represented using Σ^0_3 games. It thus has to be at this level of Σ^0_3 payoff sets that one should look for suitable representing games.

Remark 3. Variant games, G^F_φ, are definable for Field's sets of ultimate truths, or falsehoods, or of intermediate value, from [3]. Here for a given φ the players I, and II will be trying to present linear evaluation sequences demonstrating (in the notation from [3]) that $||\varphi|| = 1$ or 0, and if neither has a winning strategy, then we should conclude that $||\varphi|| = \frac{1}{2}$. For this to work we need the appropriate version of the technical Lemma 3. Again here this requires showing that from a knowledge of the semantic values at a stage α less than or equal to the first acceptable point Δ_0 (which is our ζ), the whole sequence of prior semantic evaluations can be recovered. This should perhaps be in any case plausible to readers of [15] since there it is shown how to recover codes for levels of the L_α hierarchy and their complete theories, from a set of semantic values at a stage. At these low levels of the L hierarchy, the complete sequence of codes of previous levels L_β for $\beta < \alpha$ is obtainable from the theory for L_α (at least for mildly closed α). Moreover Field's semantic model construction can be run inside the constructible universe and hence those theories of levels L_α contain *a fortiori* all the previous semantic evaluation sets. One may then prove the Lemma in the following form: if $F_\alpha =_{df} \{\sigma\colon |\sigma|_\alpha = 1\}$, then for $\alpha < \beta < \Delta$, $F_\beta \nsubseteq F_\alpha$. Then the template of the above game G^H_φ can be used to give a similar G^F_φ.

§5. **Implications and conclusions.** What do these games show us about the defined truth sets? For the simpler open games, and especially Martin's Strong Kleene game where the game directly follows the compositionality of the truth forming rules, the games give an alternative epistemological description of the truth sets that do not refer to 'ordinals', 'fixed points' and the like. A strategy for an open game that is over (at least for one player) in finitely many steps is something graspable (*pace* the fact that the strategy itself is nevertheless an infinite object).

What of the strategies for the games involved in the stability game? These, as we have remarked, are necessarily of infinite length, and must have runs of play in prescribed Σ^0_3 sets. Indeed their complexity reflects the complexity of the stable truth sets, and indeed of the whole enterprise of the revision theory of truth. However, we may draw something from this representation *via* descriptive set theory. The quasi-inductive notion that is implicit in the Herzberger revision sequence (and which Burgess made explicit in [1, p. 679]) yields a notion of sets of integers forming a *Spector class* and we may apply a very general theorem due to Harrington, see [9, Thm. 3.2], that represents such classes of integers using open games (and thus over in a finite time for Player I if he has the winning strategy). These, however, are not in the usual first order logic, but in such augmented with non-standard quantifiers. When one inspects the quantifiers the theorem gives for a general Spector class, on this particular case, one sees that the quantifier would 'measure' gödel numbers of sets of sentences in terms of the prewellordering $\prec$ of their 'settling down' or 'becoming stable'. What that indicates is that: (a) one can have an open, and so 'Σ^0_1', game formula representation of the stability set, albeit in terms of a logic of non-standard quantifiers; and further suggests (b) the possibility of axiomatising the stable theory of truth, perhaps not well in a language $\mathcal{L}^+_{\mathcal{M}}$ containing a $\dot{T}$ symbol for truth alone, but in a language augmented with a predicate symbol incorporating somehow the notion of $\prec$. Without some kind of additional modality, or intension, we are rather doubtful of any possibility for meaningfully axiomatising (*qua* a theory of truth alone) the Herzberger style revision theory (or any of the other versions). This is because of the disjunction between the successor rules, which clearly have truth theoretic content, and the limit rules, which as proposed in the literature, have none. Whether there is any meaningful modality that corresponds to $\prec$ remains to be seen, but there are points of contact. As remarked in the conclusion of [15] the revision theoretic machinery gives an interpretation to $T(\varphi) \wedge \varphi$ as something of the form 'having been true and true at this stage now'.

Similarly Field defines in his model [3] a hierarchy of '*determinately true operators*'. This model has an additional binary operator '$\longrightarrow$' added to the language as a sort of extended conditional. This, at least at the bottom level where $D(A)$ is taken as abbreviating $A \wedge (\top \longrightarrow A)$ has a similar effect of taking into account the truth value of a sentence A at previous stage(s). He defines not one but a sequence of determinacy operators D^α for $\alpha < \lambda_0$ (for λ_0 some unspecified countable ordinal) and shows that these have nice properties when evaluated in his model construction. As mentioned at Remark 3 above, from the semantic values at stage α it is possible to recover the whole evaluation sequence of the model up to stage α. This allows us, instead of finding an axiomatisation of the validities of this model, rather to define the determinacy operators *via* an internal characterisation of this particular model. These results will hopefully appear elsewhere.

REFERENCES

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

[2] A. Cantini, *A theory of formal truth arithmetically equivalent to ID_1*, ***The Journal for Symbolic Logic***, vol. 55 (1990), no. 1, pp. 244–259.

[3] H. Field, *A revenge-immune solution to the semantic paradoxes*, ***Journal of Philosophical Logic***, vol. 32 (2003), no. 3, pp. 139–177.

[4] H. Friedman, *Higher set theory and mathematical practice*, ***Annals of Mathematical Logic***, vol. 2 (1970), no. 3, pp. 325–327.

[5] V. Halbach and L. Horsten, *Axiomatising Kripke's Theory of Truth*, ***The Journal for Symbolic Logic***, vol. 71 (2006), no. 1, pp. 677–712.

[6] H.G. Herzberger, *Naive semantics and the Liar paradox*, ***Journal of Philosophy***, vol. 79 (1982), pp. 479–497.

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

[8] H.Leitgeb, *What truth depends on*, ***Journal of Philosophical Logic***, vol. 34 (2005), pp. 155–192.

[9] A. S. Kechris, *On Spector classes*, ***Cabal seminar 76–77*** (A. S. Kechris and Y. N. Moschovakis, editors), Lecture Notes in Mathematics Series, vol. 689, Springer, 1978, pp. 245–278.

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

[11] D. A. Martin, *Revision and its rivals*, ***Philosophical Issues***, vol. 8 (1997), pp. 407–418.

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

[13] M. O. MedSalem and K. Tanaka, *Δ^0_3-determinacy, comprehension and induction*, ***The Journal for Symbolic Logic***, vol. 72 (2007), no. 2, pp. 452–462.

[14] Y. Moschovakis, ***Elementary induction on abstract structures***, Studies in Logic and the Foundations of Mathematics, vol. 77, North-Holland, Amsterdam, 1974.

[15] P. D. Welch, *Ultimate truth* vis à vis *stable truth*, ***Review of Symbolic Logic***, vol. 1 (2008), no. 1, pp. 126–142.

[16] P. D. Welch, *Weak systems of determinacy and arithmetical quasi-inductive definitions*, submitted, `arXiv:0905.4412`.

[17] S. Yablo, *Truth and reflection*, ***Journal of Philosophical Logic***, vol. 14 (1985), pp. 297–349.

SCHOOL OF MATHEMATICS
UNIVERSITY OF BRISTOL
BRISTOL, BS8 1TW, UK
E-mail: p.welch@bristol.ac.uk
URL: www.maths.bristol.ac.uk/~mapdw