Chapter 8

Revenge, I

8.1. Forms of revenge

It is hard enough to find a satisfying response to the paradoxes, but the phenomenon of revenge can make it seem impossible. In the simplest manifestation of revenge - call it *direct revenge* - the pathological sentence or expression re-emerges intact from the attempt to treat it. Any account of paradox will surely characterize the denoting expression C, or the predicate P, or the liar sentence L as pathological in some way or other. For example, we saw that in one treatment of his paradox, Richard suggests that an analogue of C fails to denote;[1] Martin and Maddy suggest that the extension of a Russell predicate, analogous to P, falls into a membership gap, failing to belong to its own extension or anti-extension;[2] and L is characterized by Kripke as ungrounded.[3] But direct revenge seems to show that the withholding of successful reference to C, or a determinate extension to P, or a truth value to L, leads only to the reinstatement of a reference, or an extension, or a truth-value, and the apparent return of paradox. If C doesn't denote, then the sum of the numbers denoted by expressions on the board is π+6, so C does denote – it denotes π+6. And similarly for P and L.

Direct revenge, then, seems to make life very difficult: we surely must conclude that these paradox-producing expressions are pathological in some way or other. But if we do, that seems only to encourage their immediate recovery and restore their power to produce paradox. It seems that we cannot call them pathological on pain of paradox! But C, P and E and their ilk are not pathological, what are they?

Moreover, revenge can seem to iterate. If we declare that C fails to denote, then it seems to follow that C denotes π+6. But then the sum of the numbers denoted by expressions on the board is π+6+(π+6). And so it seems that C does not denote, and denotes π+6, and also denotes π+6+(π+6), and so on, indefinitely. And the Russell predicate P fails to have an extension, has a determinate extension determined only by the predicate 'moon of the Earth', and also has an extension determined both by the predicate P and the predicate 'moon of the Earth'. And if L fails to have a truth value, then, since it says of itself that it isn't true, L is true; but then what it says is the case, so L isn't true. But that's what L says, so it is true – and so on, endlessly back and forth.

Direct revenge is generated by the very sentences and expressions that we were trying to treat in the first place. But revenge can take another form - call it *second-order revenge*. Often a solution to paradox will introduce new notions – for example, gaps (in truth, reference, or predicate-application), levels of a hierarchy, groundedness, determinateness, stability, context. Second-order revenge takes these new notions and constructs new paradoxes for old.[4] Theories of truth, for example, face new challenges presented by sentences that say of themselves that they are false or gappy, or not true at any level of the hierarchy, or ungrounded, or not determinately true, or not stably true, or not true in any context.

The connection between direct and second order-revenge is a delicate matter. The notions that generate direct revenge – denotation, extension, truth - are the initial targets of an attempt to solve the paradoxes. Those that generate second-order revenge appear to be more specialized semantic notions, ingredients of a semantic theory that deals with paradox. Yet theorists are likely to present these notions as themselves natural and intuitive – the solution should not be artificial, unconnected to our ordinary semantic intuitions. But the more natural these notions, the more they should be regarded as an initial target. For example, a gap theorist is likely to appeal to the naturalness of the notion of a truth-gap. And if truth-gaps are part of our ordinary repertoire, then so is the disjunctive notion of being *false or gappy*, along with the coextensive notion of being *not true*, on one natural reading of negation. Here, second-order revenge collapses into the first-order revenge generated by L – and then so much the worse for the gap theorist, if the theory cannot deal with even the initial target.

Where there is no such collapse, second-order revenge presents a distinct challenge to a semantic theory. Suppose the newly introduced concepts, though natural enough, are not part of our immediate repertoire, and so are inappropriate initial targets. But since they do give rise to paradox, the theory is limited – even if it can deal with the initial targets, it cannot deal with these new ones. This is a significant failure: on pain of paradox, the semantic theory cannot accommodate natural enough semantic concepts. Second-order revenge seems to present an unpalatable choice, between contradiction on the one hand, and a significant expressive incompleteness on the other. Second–order revenge threatens to show that however successfully a theory deals with its initial targets, it cannot deal adequately with the general phenomenon of semantic paradox.

In this chapter, I examine the phenomena of direct and second-order revenge as it arises for Kripke's theory, for neo-Kripkean theories (especially Field's), and for paraconsistent theories (especially Priest's). In the next chapter, I turn to revenge and contextual theories, paying particular attention to the singularity theory.

8.2 Kripke's Theory of Truth

We start with a classical first-order language L rich enough to express its own syntax, and expand it to a language £ by adding a 1-place predicate T(x) which will be *partially* defined. Associated with £ is an interpretation function I, and a set V of values {1, ½, 0}, where we can think of 1 as truth, 0 as falsity, and ½ as undefined. In the familiar way, the interpretation function I assigns to each name an element of D, to each n-ary function-symbol a function from D^n to D, and to each n-ary predicate letter a function from D^n to V. The truth-functional connectives and quantifiers are handled by the strong Kleene valuation scheme. First the connectives: where v(A) is the semantic value of A, v(~A) = 1-v(A); v((AvB) = max(v(A),v(B)); v(A&B)=min(v(A),v(B)). Notice that negation here is what is sometimes called *choice negation* – in particular, the negation of a sentence which is undefined is itself undefined. The conditional and biconditional are defined in the usual way in terms of the other connectives. For the quantifiers: v(∃xA(x)) = max{v(A(t/x)) | for all t in D}, and v(∀xA(x)) = min{v(A(t/x) | for all t in D}.

An interpretation of the 1-place predicate T(x) is a function that takes each element of D to exactly one of the values 1, ½, 0. Let the function *f* be an interpretation of T(x). Then the *extension* S_1 of T(x) is the set of elements of D that *f* takes to 1; and the *anti-extension* S_2 of T(x) is the set of elements of D that *f* takes to 0. S_1 and S_2 are disjoint subsets of D, but their union need not be all of D – this is a three-valued system that accommodates partial interpretations of T(x). Let £(S_1,S_2) be the interpretation of £ with T(x) thus interpreted. Let $S_1^/$ be the set of (codes of) true sentences of £(S_1,S_2), and $S_2^/$ be the set of (codes of) false sentences of £(S_1,S_2) together with all elements of D that are not codes of sentences of £. The unary function φ is given by $\varphi(<S_1,S_2>) = <S_1^{/},S_2^{/}>$. Given the strong Kleene valuation scheme, it is straightforward to show that φ is monotonic.[5]

The key construction of Kripke's theory is the *minimal fixed point*. The minimal fixed point provides an interpretation of the truth predicate T(x) that satisfies the following intuition: whenever you can assert a sentence S, you can assert the sentence T(S) (and vice versa). If you can assert '7+5=12', then you can assert "'7+5=12' is true". And we can keep going, for sentences that themselves contain 'true': if you can assert "'7+5=12' is true", then you can assert '"'7+5=12' is true" is true'. And so on. We can put the intuition this way: S and T(S) are intersubstitutable in any transparent context.

We reach the minimal fixed point by a transfinite series of stages. At stage 0, we assign to the 1-place predicate T(x) the function *f* that assigns to each element of D the value ½. So, at stage 0, *f* does not assign the values 1 or 0 to any element of D: both the extension and the anti-extension of T(x) are $\wedge$, the empty set. Let $£_0$ be $£(\wedge,\wedge)$, the interpretation of £ which for which T(x) is completely undefined. Let $S_{1,1}$ be the set of codes of true sentences of $£_0$, and let $S_{2,1}$ the set of all elements of D that are either codes of false sentences of $£_0$ or not codes of sentences of $£_0$. Then the interpretation of £ one level up from $£_0$ is $£_1 = £(S_{1,1},S_{2,1})$. In general, if α is a successor ordinal ($\alpha = \beta+1$), $£_\alpha = £(S_{1,\alpha}, S_{2,\alpha})$, where $S_{1,\alpha}$ is the set of codes of true sentences of $£_\beta$, and $S_{2,\alpha}$ is the set of all elements of D that are either codes of false sentences of $£_\beta$ or not codes of sentences of $£_\beta$. And if λ is a limit ordinal, $£_\lambda = £(U_{\beta<\lambda}S_{1,\beta}, U_{\beta<\lambda}S_{2,\beta})$.

The intuitive idea is that the extension and anti-extension of T(x) increase as we go up the levels – as a consequence of the monotonicity of φ, once a sentence gets into the extension or anti-extension of T(x), it stays in. But this process does not go on forever – we reach a fixed point of φ, call it $£_\sigma$, in which the truth predicate of $£_\sigma$ is true of exactly the true sentences of $£_\sigma$.

And similarly for the falsity predicate. So $£_\sigma$ is a language that contains its own truth and falsity predicates. $£_\sigma$ is the *minimal* fixed point, since its construction starts out from the empty interpretation of T(x).[6] Every sentence of $£_\sigma$ that can be declared true is declared true in $£_\sigma$. And every sentence of $£_\sigma$ that can be declared false is declared false in $£_\sigma$. Only ungrounded sentences (such as the Liar and the Truth-Teller) are without a truth value in the minimal fixed point.[7]

Kripke's minimal fixed point is semantically closed to a remarkable degree – it contains its own truth and falsity predicates. But consider Kripke's notion of groundedness: a sentence is *grounded* if it receives the value 1 or 0 in the minimal fixed point, and ungrounded otherwise. If we were to add the grounded predicate to $£_\sigma$, a second-order revenge paradox is generated by, for example, 'This sentence is false or ungrounded' (contradiction follows whether we assume the sentence is true, false or ungrounded). The escape route is an ascent to a metalanguage. Central terms of Kripke's theory, like 'grounded' and 'paradoxical', are not in the object language, but in a metalanguage in which the theory is expressed. Even if paradoxes involving truth and falsity are handled by Kripke's theory, paradoxes involving groundedness are not. The notion of groundedness is beyond the expressive capacity of $£_\sigma$. So Kripke's theory is vulnerable to second-order revenge.

Kripke's minimal fixed point is expressively incomplete in other ways too. As Kripke points out:

> "Liar sentences are *not true* in the object language, in the sense that the inductive process never makes them true; but we are precluded from saying this in the object language by our interpretation of negation, and the truth predicate"[8]

Liar sentences are not true, but that's not because they are false – to be false is to be in the antiextension of 'true', and Liar sentences aren't in this antiextension. The inductive process generating the minimal fixed point places Liar sentences lie outside the extension *and* the antiextension of 'true' – Liar sentences are neither true nor false. And it follows from this that they're not true. But for us to say that Liar sentences aren't true presupposes that we have the resources to express the notion *neither true nor false*.   This notion is not expressible in Kripke's object language. If it were, paradox would be generated by sentences such as 'This sentence is false or neither true nor false', or 'This sentence is not true', where 'not true' is not coextensive with 'false', but rather expresses the full complement of truth.[9]

Similarly with a Truth-Teller sentence, a sentence that says of itself that it is true. This will receive no truth value in the minimal fixed point – "In particular, it will never be called 'true'" (Kripke, p66). But this fact, that the Truth-Teller is not true, cannot be stated in the language of the minimal fixed point, on pain of contradiction.[10]

One way to characterize the expressive incompleteness here is to say that the object language cannot express the notion of a truth-value gap. Another way is to distinguish between two senses of negation. We saw that Kripke adopts the choice negation of Kleene's strong 3-valued logic. Choice negation is contrasted with exclusion negation, where ~P is false iff P is true, and ~P is true iff P is false or undefined. Given this distinction, we can say that exclusion negation is not expressible in Kripke's object language.   As far as exclusion negation is part of natural language, then Kripke is also subject to *direct* revenge, in the form of the Liar sentence "This sentence is not true".[11]

The impact of revenge is felt not just by Kripke's theory, but by many others, and in much the same way. We seem forced to accept expressive incompleteness on pain of contradiction. We start with a target semantic notion – in the present case, truth – and provide a theory of that notion which is not vulnerable to the associated paradoxes. Kripke provides a precise characterization of a language that can express consistently its own notion of truth. But nevertheless $£_{\sigma}$ is expressively incomplete – it cannot express the semantic notions introduced by the theory, such as groundedness and truth-value gaps. I have argued elsewhere that parallel remarks can be made about a variety of theories of truth and the semantic notions they introduce, whether stable truth, or definite truth, or fuzzy truth, and so on.[12]

How might the theorist respond? One response might go like this: these revenge paradoxes turn on technical notions, and the proper setting of the semantic paradoxes is ordinary language.[13] Terms like 'true' and 'denotes' are terms of ordinary language; terms like 'grounded' and 'truth-value gap' are not. So, for example, if Kripke's minimal fixed point language L is a plausible model of English, then it's plausible to say that we have a solution to the liar in its natural setting. The problem with this response is that these introduced notions are supposed to be intuitive. We can readily grasp the thought that the evaluation "'Snow is white' is true" is grounded in a sentence free of the truth predicate, while "This sentence is true" is not; or the thought that the Truth-Teller is neither true nor false; or the idea that the truth value of "This sentence is false" is unstable, flip-flopping between truth and falsity (if it's true, then it's false, so then it's true, so then it's false, …); or the claim that "'Harry is bald' is true" can be regarded as no more definitely true than "Harry is bald"; and so on. Indeed, if these notions were not natural and intuitive, the theories would face the charge that they're artificial and unmotivated. So the objection remains: the theories cannot deal with semantic paradoxes generated by natural enough semantic notions.

A second response might go like this: why expect the theory to deal both with the original target concepts *and* with the theoretical concepts of the theory itself? The basic concepts of denotation, extension and truth are to be treated one way, and the theoretical concepts another. For example, why not treat the revenge paradoxes that turn on groundedness or stable truth by a distinction between levels of language, and treat the language of the theory as a metalanguage for the target object language? The problem with this response is twofold. First, the family of revenge paradoxes, both direct and second-order, seems too close-knit to require distinct kinds of resolution. The sentences that generate second-order revenge (e.g. 'This sentence is false or ungrounded', 'This sentence is not stably true', etc.) seem very like those that generate direct revenge, and the contradiction-producing reasoning looks very similar. The concepts may be different, but the structure of paradox remains the same. Second, whatever additional way out is offered for the introduced concepts, that too will face its own second-order revenge. If, for example, we appeal to a distinction between language levels, then we face the challenge posed by 'This sentence is not true at any level'. The problem of second-order revenge is just postponed.

8.3 Field's theory of truth

8.3.1 Kripke and non-classical logic

Kripke accepts that there are notions that are beyond the scope of the object language. He accepts the need to ascend to a metalanguage in order to say that Liar sentences are not true, or to express the notions of *grounded* or *paradoxical*.[14] But this concession presupposes that the predicates 'grounded', 'paradoxical' and 'gappy' themselves have *exhaustive* extensions and anti-extensions, and that there really is an exclusion-negation operator in natural language. According to Field, we should reject these presuppositions.

Field suggests that it is better to understand Kripke's theory not as committed to truth-value gaps, but rather as committed to a non-classical logic.[15] In this *paracomplete* version of Kripke's theory, call it KFS,[16] the law of excluded middle is not valid – for example, where Q is a Liar sentence, Qv~Q is to be rejected. According to Field, Kripke's construction "implicitly gives a non-classical *model theory* for a language with a truth-predicate" (Field 2008, p.65). The two key features of the theory KFS are that (1) it is based on a logic, which Field calls K3, which is appropriate to the strong Kleene semantics, and (2) it models naïve truth, since it contains the Intersubstitutivity Principle, according to which <A> and True<A> are fully intersubstitutable in non-opaque contexts. And "a large part of the value of this model theory" (p65) is that it provides a consistency proof of a theory of truth that is built into the Kripkean fixed points.[17] The intersubstitutability principle and the consistency proof are incompatible with the inclusion of the law of excluded middle. So KFS is a non-classical theory that, it might be claimed, provides an idealized model of truth in natural language – and, moreover, a model in which truth is provably consistent.

It is clear that in KFS, having semantic value 1 in the minimal fixed point[18] should not be identified with truth. Field gives two examples.[19] First, the notion of *having semantic value 1* is a classical notion, defined in classical set theory, and so every sentence of the object language $£_\sigma$ either has value 1 or it doesn't. If truth is identified with having semantic value 1, then KFS will claim that A is true or A is not true, for any sentence A, including Liar sentences. But the sentence "Q is true or Q is not true", where Q is a Liar sentence, is not in the minimal fixed point, so it is no part of the theory.[20] So there is a divergence between the truths of the theory and the truths of the minimal fixed point. And since the theory KFS declares untrue any sentence not in the extension of ‘true’ in the minimal fixed point, it would declare untrue a claim of the theory. Second, observe that the Liar sentence Q does not have value 1 (it has value ½). If truth is identified with 1, then, since Q does not have value 1, then Q is not true. But the claim that Q is not true does not appear in the minimal fixed point, and so again we have a divergence between the minimal fixed point and the theory, and the anomalous consequence that the theory declares untrue one of its own claims.

So we should not identify *having semantic value 1* with *being true*. It follows that we should not identify *having the value ½* with *being a truth-value gap*. To say that Q receives the value ‘undefined’ (or ‘u’ or ‘1/2’) is not be understood as saying that Q is neither true nor false, or gappy. There are sentences that, in the semantics, receive neither value 1 nor value 0, so there are ‘semantic value gaps’; but “failure to have semantic value 1 or 0 is not failure to be true or false” (p71). So what is the relation between having semantic value 1 and truth (and between having semantic value 0 and falsity)? We can say that having semantic value 1 is sufficient for being true (and having semantic value 0 is sufficient for being false). Field continues:

> “For sentences with semantic value ½, we can’t say they’re true, or that they aren’t, or that they’re false, or that they aren’t. We cannot say whether or not they are “gappy” … . And our inability to say these things can’t be attributed to ignorance, for we don’t accept that there is a truth about the matter. This isn’t to say that we think there is no truth about the matter: we don’t think that there is, and we don’t think that there isn’t. And we don’t think there either is or isn’t. Paracompleteness runs deep”.[21]

Though Field prefers this way of taking Kripke’s theory, he observes that KFS has three serious weaknesses: it does not contain a “decent” conditional (and so cannot carry out ordinary reasoning),[22] it does not validate the truth schema,[23] and there are things we would like to say, but cannot say, about the Liar sentence – for example, that it isn’t true.[24] These deficiencies of

KFS lead Field to a theory that significantly improves KFS, by providing a more reasonable conditional and the resources to assess semantically defective sentences.

8.3.2 Field's theory of truth

I turn first to Field's conditional. The formal idea behind the conditional draws on Kripke's minimal fixed point construction and the revision theory of Gupta and Belnap (though Field uses a 'revision rule' for the conditional, not for the truth predicate). We add to a base language the predicate 'true' and the new conditional →. At the initial starting point, 'true' receives the empty extension, and every conditional (that is, every sentence whose main connective is →) receives the value ½.[25] We then proceed from this starting point to the minimal fixed point, following Kripke's construction and the strong Kleene valuation scheme. All conditional-free sentences will be interpreted in the usual Kripkean way. And we use this minimal fixed point to determine a value for a conditional A→B as follows: if A's value in the minimal fixed point is less than or equal to the value of B, the value of A→B is 1; otherwise, its value is 0. This provides a new starting point from which to construct the next minimal fixed point. We move in the same way through all successive minimal fixed points and starting points. At a limit stage, we look back to all previous minimal fixed points to see if there is a fixed point after which the value of A is always less than or equal to the value of B, or always greater. If the former, then the value of A→B is 1, and if the latter, the value of A→B is 0. If there is no such fixed point the value of A→B is ½. This provides a new starting point from which a minimal fixed point at the limit stage is constructed. Finally we can define the *ultimate value* of a sentence. The ultimate value of a sentence A is 1 if there is a fixed point after which the value of

A is always 1; 0 if there is a fixed point after which the value of A is always 0; and ½ if there is a fixed point after which the value of A is always ½, or if A fails to stabilize.[26]

By this construction, all sentences, including all conditionals, receive an ultimate value. The construction has several key features. The ultimate values obey the strong Kleene valuation scheme.[27] The construction validates the Intersubstitutivity Principle and the truth schema, and provides a logic for the conditional "strong enough so that the Intersubstitutivity Principle for truth follows from the Tarski schema as well as entailing it".[28] And Field provides a consistency proof, showing that the introduction of the conditional does not result in any new paradoxes.[29] So we now have, according to Field, a reasonable conditional and the validation of the truth schema. This overcomes two defects of KFS. It remains to show how we can assess semantically defective sentences.

The problem for KFS is that, though Liar sentences are not true in the object language (the inductive process never makes them true), we cannot say that in the object language, given the interpretation of negation and the truth predicate. Exclusion negation would allow us to say that Liar sentences are not true, but that's not available to Kripke, or to Field. What is needed is a stronger notion of truth, so that when we say that Liar sentences are not true, we're saying that Liar sentences are not *strongly true*. For Field, *strong truth* is to be understood in terms of a *determinately* operator applied to the Kripkean notion of truth - so when we say that a Liar sentence is not true, we're saying that it's not *determinately true*. So Field introduces a determinately operator *D*. Applied to the Kripkean truth predicate, this yields the notion of determinate truth. Iterating, we obtain the notions of determinate determinate truth, determinate determinate determinate truth, and so on – a transfinite hierarchy of increasingly strong notions of truth. This determinately operator is definable within the object language, in terms of Field's conditional '→' (and conjunction and negation), and Field provides a consistency proof for this language. And so this hierarchy of increasingly strong truth predicates cannot generate new paradoxes – the theory is "revenge-immune", in Field's phrase. In contrast to Kripke's theory, there is no need to ascend to a metalanguage – all these notions of truth are definable within the object language itself. Field writes:

> "If we think of a determinately operator as attaching to a truth predicate to yield a predicate of "strong truth", we can think of the theory as providing an account of hierarchy of "stronger and stronger truth predicates". But unlike most approaches that allow a hierarchy of "truth predicates", no infinite hierarchy of metalanguages is required. Indeed there need be no distinction between metalanguages and object languages at all: if the object language is rich enough to include standard set theory (ZFC) and a single notion of truth that obeys the truth schema (and of course the Kleene connectives and the new→), then all these other "truth predicates" are definable within the object language."[30]

Field defines DA as A ∧ ~(A->~A).[31] This yields some expected inferential laws for the determinacy operator, for example:

(i) $\models$ DA -> A

(ii) A $\models$ DA

(iii) DA $\models$ A

(iv) If $\models$ A->~A then $\models$ ~DA

(v) (If $\models$ A->B then $\models$ DA->DB.

Where Q is a liar sentence which asserts its own untruth, it follows from (iv) that $\models$ ~DQ. So given a Liar sentence, we can now say that it is not true - that is, not determinately true. It's straightforward to check that the semantics confirms this. Q has value ½ at every fixed point. So at every fixed point after the first, DQ (that is, Q ∧ ~(Q->~Q)) has value 0, and so ~DQ has value 1.[32] That is, the claim that the Liar sentence Q is not determinately true has ultimate value

1. It's also easy to check that ~D~Q has ultimate value 1. So the conjunction ~DQ $\wedge$ ~D~Q has ultimate value 1. That is, neither Q nor its negation is determinately true.[33]

But now what about a *determinate liar sentence* – a sentence that says of itself that it isn't determinately true? Let Q_1 be the sentence ~DQ_1.[34] According to Field's theory, Q_1 receives the value ½. But we cannot assert that Q_1 is not determinately true – that would contradict Field's consistency proof. So not only can we not assert that Q_1 is determinately true, we also cannot assert that Q_1 is not determinately true. That is, we cannot assert DQ_1 v ~DQ_1 (and in fact we can reject it). That is, the Law of Excluded Middle cannot be assumed even for claims of determinateness (in the case of Q_1 and Determinate Liar sentences generally, it can be rejected). Yet Q_1 is defective – so how can we capture the intuition that Q_1 is not true (the intuition about Liar sentences that KFS fails to capture)? Notice that we can assert that Q_1 is not determinately untrue, but we cannot assert that Q_1 is not determinately true. However, we *can* assert that Q_1 is not *determinately* determinately true. And we can assert that both Q_1 and its negation are not determinately determinately true, which is a way of expressing the defectiveness of Q_1. The iteration of the determinateness operator provides the means for expressing a way in which Q_1 is defective.[35]

And we can keep going. Let Q_2 be ~DD(Q_2), that is, a sentence that says of itself that it is not determinately determinately true. We can assert that ~DDDQ_2 (that Q_2 is not determinately determinately determinately true), and capture its defectiveness, in a way, by the assertion that ~DDDQ_2 $\wedge$ ~DDD~Q_2. In general, let Q_σ be the sentence which says of itself that it is not D^σ-true, and D^σ is the σ-fold iteration of D. Then we cannot assert ~$D^\sigma Q_\sigma$, or $D^\sigma Q_\sigma$ v ~$D^\sigma Q_\sigma$. But we can assert ~$D^{\sigma+1}Q_\sigma$ $\wedge$ ~$D^{\sigma+1}$~Q_σ (and also ~D~Q_σ). The series of Q_σ-sentences corresponds to a series of stronger and stronger truth predicates. But provably, the Q_σ-sentences cannot generate any new paradoxes – they're contained within Field's provably consistent language. For this reason, Field takes his account to be "revenge-immune". And each Q_σ-sentence can be classified as defective (in the sense that neither it nor its negation is determinately$^{\sigma+1}$ true) *within* the language – there is no need to ascend to a metalanguage.

A substantial restriction should be noted here: this hierarchy can only be defined for ordinals for which an ordinal notation exists. This is because at limit stages we need to form infinite conjunctions: for a limit ordinal λ, $D^\lambda A$ is the infinite conjunction of all the $D^\alpha A$ for $\alpha<\lambda$. That is, for a limit ordinal λ, '$D^\lambda A$' abbreviates '(for all $\alpha<\lambda$)(True(<$D^\alpha A$>)', where the truth predicate is employed in its logical role of expressing infinite conjunctions. So as Field puts it, the superscript in the sentence '$D^\lambda A$' must be "a piece of notation for the ordinal λ, not the ordinal λ itself".[36]

There is no doubt that Field's theory is an impressive technical achievement. It's reasonable to suppose that Field takes the Kripkean approach just about as far as it can go. The question remains, however, whether or not it provides an adequate account of the Liar, and, more broadly, of semantical paradox generally. The proper setting for the semantical paradoxes is natural language – an adequate solution to the Liar and the other paradoxes should deal with the paradoxes as they arise in natural language. I shall argue that Field's theory is too far removed from natural language to come properly to grips with semantical paradox. This concern has a number of connected strands.

8.3.3 Field's conditional

Let me start with Field's conditional →. Field's conditional does behave in a number of expected ways (for example, we have: $A, A\rightarrow B \models B$; $\models A\rightarrow{\sim}{\sim}A$ $\models A\wedge B\rightarrow A$).[37] And if excluded middle holds for A and B, → behaves just like the material conditional ⊃. But, as Yablo points out,[38] it is hard to find an intuitive, "antecedently comprehensible", meaning for →. Field's conditional is not truth-functional -- as Field points out, when the ultimate values of both A and B are ½, the ultimate value of A→B is *either* 1 *or* ½; when the ultimate value of A is 1 and of B is ½, the ultimate value of A→B is either ½ or 1; and when the ultimate value of A is ½ and of B is 0, the ultimate value of A→B is either ½ or 0.[39] So an understanding of → is not to be had from an extensional truth-table. Yablo shows that → is somewhere between the Lukasiewicz conditional[40] and the necessitation of the Lukasiewicz conditional (stronger than the former and weaker than the latter) – but still we have no definite intuitive understanding of →.[41]

Yablo also points out that Field's conditional yields some counterintuitive results. We do not expect the Truth-Teller

(E) E is true

to be true. But consider a version of the Truth-Teller couched in the terms of Field's conditional

(S) (A→A) →T(S).

S says that given A if A, S itself is true. Since A→A is a tautology, S in effect says of itself that it is true. According to Field's semantics, S is true at every minimal fixed point except the very first.[42] So S has ultimate value 1. And this seems unacceptably arbitrary.[43]

The Truth-Teller leads to puzzling results for Field in other ways too. Yablo asks us to consider Jones and Smith, who hold opposite views about the Truth-Teller sentence E:

(J) E is true (i.e. T(E)).

(S) E is false. (i.e. T(~E) – falsity is truth of the negation).

These are incompatible utterances, and it's natural to try to express the incompatibility using Field's conditional:

(1) $T(E) \to {\sim}T({\sim}E)$.

This has ultimate value 1 on Field's semantics for ->, as we might expect.[44] However, the sentence

(2) $T(E) \to T({\sim}E)$

also has ultimate value 1.[45] So (1) fails to express the incompatibility between T(E) and T(~E). Similarly, suppose that Jones and Smith each say that the other is lying:

(J) ${\sim}T(S)$

(S) ${\sim}T(J)$.

We might try to express their disagreement via $J \to {\sim}S$. But again it's easily checked that for Field's → both $J \to {\sim}S$ and $J \to S$ have ultimate value 1.

The Truth-Teller also compromises the equivalence of A and T(A) in Field's semantics. According to Field's semantics, the ultimate value of $T(A) \leftrightarrow A$ is 1. But Yablo points out that for the full equivalence of T(A) and A, we would also need the ultimate value of $T(A) \leftrightarrow {\sim}A$ *not* to be 1. But when A is the Truth-Teller, the ultimate value of $T(A) \leftrightarrow {\sim}A$ *is* 1.[46] This is the case not only when A is the Truth-Teller, but when A is any sentence whose value stabilizes at ½.[47] Although Field's conditional behaves as it should when A has ultimate value 1 or 0, in the case where A's ultimate value is ½, we can assert both $T(A) \leftrightarrow A$ and $T(A) \leftrightarrow {\sim}A$. And, as Yablo says, Field's conditional was supposed to help with sentences which have ultimate value ½, such as the Truth-Teller.[48]

Field acknowledges these counterintuitive consequences of the conditional →.[49] Field suggests that we might avoid them by modifying the definition of the conditional so as to take into account Kripkean fixed points other than the minimal fixed point. But then we lose the intuitiveness of the minimal fixed point, and the revised definition of the conditional is more complicated and more removed from ordinary language.[50]

In sum, then, it is not at all clear that there is any straightforward understanding of Field's non-truth-functional conditional; the semantics for the conditional, though designed with defective sentences in mind, produces counterintuitive results in cases like the conditional truth-teller (and other sentences that stabilize at ½); and Field's suggested way of avoiding these results leads us further away from natural language.

8.3.4 Determinate truth

I turn now to Field's *determinately true* operator and the claim that the theory is revenge-immune. The basis for Field's claim is this: each defective sentence in the language can be assessed as defective *within* the language – there is no need to ascend to a metalanguage. Given a Liar sentence Q, we can assert within the language that neither Q nor ~Q is determinately true. This corrects one shortcoming of the Kripkean theory KFS, where one cannot assert, within the language, that a Liar sentence is not true. And given the (1st-level) Determinate Liar sentence Q_1 – which says of itself that it is not determinately true (that is, $\sim DQ_1$) - we can assert that neither it nor its negation is determinately determinately true (that is, we can assert $\sim DDQ_1 \wedge \sim DD{\sim}Q_1$). In general, as we saw, if Q_σ is the (σ-level) Determinate Liar sentence $\sim D^\sigma Q_\sigma$, we can assert $\sim D^{\sigma+1}Q_\sigma \wedge \sim D^{\sigma+1}{\sim}Q_\sigma$. In this way, every defective sentence of the language can be evaluated as defective within the language. As Field puts it: "…you can state in this logic the way in which certain sentences of the logic are "defective"; because you can do so, and because there is a consistency proof of naïve truth theory in the logic, the notion (or notions) of defectiveness cannot generate any new paradoxes".[51]

But we can ask: in what sense are no new paradoxes generated by the "the notion (or notions) of defectiveness"? What we have in Field's theory is a transfinite hierarchy of distinct notions of defectiveness, as follows: a sentence A is defective$_0$ iff ~DA, and ~D~A, defective$_1$ iff ~DDA and ~DD~A, and so on, through the ordinals for which a notation exists. But what we don't have within the theory is a general notion of defectiveness: "Unfortunately we can't get a single unified notion of defectiveness, but must rest content with an increasing hierarchy…".[52] For suppose there was a general determinate truth operator D* in the language. Then we could express the general notion of defectiveness: a sentence A is defective iff ~D*(A) ∧ ~D*(~A). But if we assume D* is an operator in the language, we can form the Liar sentence

(Q*) ~D*(Q*).

Q* is not determinately true, but to assert that – that is, to assert ~D*(Q*) – is just to assert Q*. So we are led to paradox. On pain of paradox, then, there is no general determinateness operator expressible in the language, and consequently no general defectiveness notion expressible in the language.[53]

However, it seems these general notions of determinate truth and defectiveness are intelligible notions. We seem to understand what we mean when we say that all Liar-like sentences (liar sentences, truth-tellers, Curry sentences, and so on) are defective, or when we say that all these sentences are neither determinately true or determinately false. The notions of defectiveness found in the hierarchy – defectiveness$_1$, defectiveness$_2$, and so on – are highly specific ways of being defective. It's hard to see how we could grasp the significance or point of these distinct hierarchical notions without an intelligible general notion of defectiveness as a guide. This seems like a notion expressible in natural language – in particular, in the present paragraph.

One possible response to this worry is simply to deny the appearances, deny the existence of intelligible general notions of determinate truth and defectiveness. If these notions are unintelligible, then they cannot establish the expressive incompleteness of *any* language. Field does in a few places take this way out. According to Field, his model theory "ought not to make sense of a notion of super-determinateness meeting intuitive preconceptions: my claim is that the notion is ultimately unintelligible";[54] a super-determinateness operator "doesn't really make sense (though I grant that this is initially quite surprising)".[55] But it is a distinct drawback of any solution to paradox if it is forced to deny the intelligibility of notions that appear quite intelligible to us, especially if the main motivation for the denial is to protect one's theory from the threat of paradox.

Field allows that this kind of argument against his theory "is perhaps the one with most intuitive force: it is that we *just need* a unified account of determinacy or defectiveness".[56] Field continues:

> "Note however that this argument cannot very well be advocated by the classical theorist, since the classical theorist has no such unified notion either. Nor can it very well be advocated by the proponent of any other solution to the paradoxes in which such a notion is unavailable. Indeed, I'm not sure that there are any demonstratively consistent theories (or even non-trivial dialetheic ones) that have such a notion available and hence are in a position to advocate this argument. I'm willing to concede (for the moment anyway) that it would be a point in favor of a solution to the paradoxes that it had a unified notion of defectiveness."[57]

Here it is no longer clear whether Field is maintaining the view that there is no intelligible general notion of determinacy or defectiveness, or whether he is allowing that there is, but doubting that any theory can express it. The trouble with the former position is that the notion seems intelligible to us (and is needed to grasp Field's account). The trouble with the latter is that it places Field's theory among all those who buy consistency at the price of expressive incompleteness. This is the familiar trade-off. And no theory that makes this trade, Field's included, can claim immunity from revenge.

There is a further *internal* version of this kind of objection to Field's theory – that Field's theory itself contains the resources to generate paradox-producing general notions of determinateness and defectiveness. Can't we conjoin all the levels of the determinacy hierarchy and thereby obtain a general notion of determinateness – call it *hyper-determinateness* -- within the theory? And isn't that general notion of determinate truth already captured within the theory by the notion of *having ultimate value 1*? I turn to these two questions now. As Field points out, negative answers to both are guaranteed by the consistency proof for his language. However, there is a price to be paid: paradox is avoided by disengaging the theory from natural language and intuitive semantic notions.

To repeat the first question: Why can't we simply conjoin all the determinate truth operators in Field's hierarchy and obtain a general *hyper-determinateness* operator? And similarly for defectiveness: why can't conjoin all the defectiveness predicates 'defective_0', 'defective_1', 'defective_2', … , to obtain a general notion of defectiveness? Field argues in considerable technical detail that this cannot be done *within* the theory.[58] Suppose we try to define a hyper-determinateness operator H by conjoining all the determinately true operators D^{α}, for *every* ordinal α. For this, we need to define the hierarchy of iterations of D. Successor stages are straightforward: $D^{\alpha+1}$ is just $D(D^{\alpha})$. For limit stage λ, we take '$D^{\lambda}A$' as abbreviating '(for all $\alpha<\lambda$)True($D^{\alpha}A$)'. But there is a complication, as we noted above: '$D^{\lambda}A$' is to be a sentence, so the superscript must be a notation for the ordinal λ, not the ordinal λ itself. Field shows that as a consequence of this, the process of iterating D collapses. That is, for sufficiently big ordinals, all the iterations of D corresponding to these ordinals are equivalent. For example, given that there are only countably many sentences, for any uncountable ordinals α and β, D^{α} and D^{β} are equivalent. And this collapse is "bad": when α is sufficiently big in this way (as with the uncountable ordinals), D^{α} takes straightforward truths into falsehoods. So if we tried to generate a paradox by defining H as an infinite conjunction of absolutely *all* the D^{α}, we will fail: H would be a trivial operator, turning every sentence into a falsehood.[59] We might try to reinstate paradox by defining H as the infinite conjunction of all the D^{α} for which the iterations are well-behaved and genuine (those prior to the collapse). Field argues that we cannot do this either: to put it intuitively and loosely, it is a "fuzzy" question where the hierarchy of well-behaved iterations ends, and the collapse begins.[60] Field concludes: "no amount of iteration, or conjunction or quantification over what has been iterated, can achieve a useful operator that is immune to further strengthening by D".[61] We cannot define an intuitive hyper-determinateness predicate in terms of all the D^{α} for which the iterations are genuine, and so no paradox is forthcoming.

So Field argues that attempts to produce a hyper-determinateness paradox by conjoining the iterations of D will fail: "… none of the defined notions of hyper-determinacy meet the joint expectations of well-behavedness and maximality… ".[62] Iterations of D, when we reach high enough ordinal levels, are not well-behaved, and any 'hyper-determinateness' operator that we define in terms of these iterations will not be the intuitive, unified notion we're after. Still, it is one thing to say that the theory itself will not generate an intuitive paradox-producing notion of hyper-determinateness (but instead only ill-behaved notions that do not threaten any contradictions). It is another to say that there simply does not exist an intelligible notion of hyper-determinateness. As Field says:

> "it just seems as if we have a unified notion of hyper-determinate truth ('determinate truth in every reasonable sense of that term') corresponding to '*True* and *DTrue* and D^2*True* and … '. Or if you like, a unified notion of

> ‘defective in some reasonable sense of that term’, viz, ‘(~*True* and ~*False*) or (~*DTrue* and ~*DFalse*) or (~D^2*True* and ~D^2*False*) or …’
> I don’t want to deny that we have these notions; but not every notion we have is ultimately intelligible when examined closely.”[63]

Again, it is a drawback of a theory if it dismisses as unintelligible notions that we do possess, especially when the breakdown of attempts to define such notions within the theory may be due to artifacts of the theory itself – perhaps the restriction of the law of excluded middle for certain notions, or the use of the truth predicate to effect quantification.

Turning to the second question: Isn’t the general notion of determinate truth *already* captured within Field’s theory, by the predicate ‘has ultimate semantic value 1’. This predicate seems to apply to exactly the sentences that are determinately true. And, likewise, the general notion of defectiveness seems to be captured by the predicate ‘has ultimate semantic value ½’. But that can’t really be so – these general notions lead to paradox, and Field’s theory is consistent. Now it is the case that the predicates ‘has ultimate semantic value 1’ and ‘has ultimate semantic value ½’ are part of Field’s language. The base language that initiates Field’s construction includes the language of set theory, and the model for the base language is definable in the base language. So Field’s set-theoretical construction can be turned into an explicit definition of these predicates.[64] And since excluded middle is assumed for the base language, excluded middle must also hold for attributions of the semantic values 1, ½ (and 0). This raises the question:

> “Can’t we then reinstitute a paradox, based on sentences that attribute to themselves a semantic value of less than 1?”[65]

The answer is no. Paradox is avoided – but at the cost of separating Field’s theory from natural language and the intuitive notion of determinate truth. Take the sentence:

(A) A has ultimate value other than 1,

Field observes that his construction yields the consistency even of this instance of the truth-schema:

True(A) ↔ A has ultimate value other than 1.

Paradox is avoided because the predicate ‘has ultimate value 1’ “can’t correspond to any normal notion of truth”.[66] In fact, given Tarski’s undefinability theorem, and given that Field defines the notion of ultimate value 1 in classical set theory, that notion will not correspond to the intuitive notion of truth even for sentences that are not indeterminate or paradoxical.[67] So, since the notions of truth and determinate truth coincide for these unproblematic sentences, the notion of having ultimate value 1 cannot correspond to any intuitive notion of determinate truth. Field’s formal theory seems out of step with any ordinary notion of determinate truth, or any ordinary notion of being other than determinately true. And, again, the fact that his theory does not contain any notion corresponding to the intuitive notion of determinate truth does not show that this notion does not exist or is unintelligible.

What then is significance of the explicit definition of ‘has ultimate value 1’? According to Field, a crucially important role of his semantic theory is to provide a consistency proof:

> “The semantics I’ve provided for truth theory, despite its distortions, gives a proof within classical set theory of the consistency of naïve truth theory in a nonclassical logic, so that we know that that logic is indeed consistent.”[68]

It is clearly significant to demonstrate the consistency of one’s theory of truth. At the same time, it is important for a theory of truth to match up with our intuitive semantic notions. Field argues that he has captured the intuitive notion of truth, since his theory adheres closely to the naïve theory of truth, as captured by the Intersubstitutability Principle. But the question here is whether determinate truth is captured by ‘has ultimate value 1’. It isn’t, and in fact Field suggests that we shouldn’t expect it to be. The notion of semantic value, he says, “is a technical notion of formal semantics (sometimes a technical notion for giving consistency proofs, sometimes a technical notion for heuristic "explanations" of logical principles, sometimes both)".[69] Determinate truth, on the other hand, is not a mere technical notion (ibid.). In the same vein, Field writes:

> "… the model theory is primarily just a model theory, used for explaining validity; no sense has been given to an assignment of values to sentences in an absolute sense, independent of a model. I now add that *a value space itself* has no significance for the real world, for it is of use only for models of cardinality no greater than a given cardinal C; for higher cardinality models you need a bigger value space, and the real world has higher "cardinality" than any C. Given that the value space has no significance for the real world, we shouldn't be disturbed by any "monsters" that it contains."[70]

Field concedes that "there may well be a place for a theory that postulates a model-independent space of values"[71]. The pull towards such a theory is the desire to provide a more satisfactory match-up between Field's model theory and natural language (in particular, the notion of determinate truth). Such a theory would require a non-classical set theory for which excluded middle failed (so that we could not assume that a sentence either has value 1 or doesn't, thereby blocking paradox). But now the challenge will be to wrestle not just with the Liar, but with the Russell paradox too.[72] Field does express some skepticism about such a project, on the grounds that the notion *having real semantic value 1* would be too much like the problematic super-determinateness operator D* discussed above.[73] Again, then, the worry is that there is a notion of determinate truth that cannot be captured by Field's theory - or by any clear variant of it.[74]

8.3.5 Field's theory and natural language

It is not just the general notions of determinate truth and defectiveness that cannot be expressed in Field's theory. Clearly, both exclusion negation and the full notion of a truth-value gap are beyond its scope, as they are for Kripke's theory, and for any paracomplete theory that relies on a monotonic fixed point construction. Field cannot allow exclusion negation or the full notion of a gap while maintaining naïve truth and the Intersubstitutivity Principle, and he denies the intelligibility of these notions.[75] This is counter to our ordinary semantic usage. The notion of a truth-value gap drives Kripke's own presentation of his theory, and it naturally arises not just in connection with the Liar, but also in connection with, for example, category mistakes, presupposition failures and category mistakes. Field himself remarks that his paracomplete theory doesn't *postulate* gaps, but does *allow for* them.[76] This seems to be compatible with the idea that while Field's theory cannot express the general notion of a truth-value gap, that notion is nevertheless intelligible. And once we have the notion of a truth-value gap, we have exclusion negation. From the claim "A is gappy", it is natural to infer "A is not true". This is a use of exclusion negation, where being not true is not simply a matter of being false. To deny that we use negation this way just seems to deny that we do what we in fact do.

Further, it is still possible to use exclusion negation in a non-classical setting. One can recast Field's theory in terms of an iteration of gaps - gaphood gaps, gaphood gaphood gaps, etc - corresponding to the levels of the determinacy hierarchy.[77] The ordinary Liar sentence Q is gappy$_0$, where the notion of gappiness$_0$ itself is gappy. And a stronger Liar sentence Q_1 that is itself couched in terms of gappiness$_0$, will not be gappy$_0$ but suffers a new kind of gap – it's gappy$_1$, say. And a still stronger Liar sentence Q_2, couched in terms of gappiness$_1$, will be gappy$_2$. And we can continue in this way through the hierarchy. Excluded middle fails for gappiness all the way along, and there is no paradox-producing, exhaustive notion of gappiness expressible within the theory. In this way, we can always say of any defective sentence that it's gappy (in one of its iterations), and thereby capture Field's sense of revenge-immunity. But parallel worries arise here for this version of Field's theory concerning a general notion of gappiness, as they did for Field's theory and a general notion of determinate truth. Isn't the gap hierarchy itself driven by a general notion of a gap? And can't we conjoin all these gaphood notions into one general notion?

But beyond these questions, it seems that there is *still* a place for natural uses of exclusion negation, even where excluded middle fails for the notion of a gap. Given that Q is gappy$_0$, it is natural to infer that Q is not true. Given that Q$_1$ is gappy$_1$, it is natural to infer that Q$_1$ is not true. And so on.

It is one thing to claim that determinate truth, defectiveness, truth-value gaps and exclusion negation are inexpressible in Field's theory, so that paradox is not forthcoming. It is another to claim that these notions are unintelligible, and that *no* language can express them.[78] As long as these notions are expressible in some language, Field's theory has restricted scope, and cannot provide an adequate response to the semantic paradoxes. And they do appear to be expressible in English – we find the general notions of determinate truth and defectiveness intelligible, we find the difference between choice and exclusion negation intelligible, and we understand the claim that the Moon is neither true nor false, or that '5 is triangular' suffers a truth-value gap. To say that these notions are unintelligible because they lead to paradox is not a way to deal with the Liar.[79]

It might be responded that Field's theory should be regarded as revisionary or prescriptive, not descriptive. Field sometimes seems to support this line; for example, he writes:

> “The paradoxes show that there’s something wrong with firmly held patterns of reasoning… . What’s of interest is to figure out how best to modify this reasoning: to find a new way of reasoning that we can convince ourselves is intuitively acceptable, and which avoids plainly unacceptable conclusions.”[80]

And in a discussion of postulates that seem to articulate a general notion of determinate truth, or what he calls ‘super-determinate truth’, Field writes:

> “I don’t doubt that that these postulates might be so central to someone’s intuitive understanding of super-determinate truth as to be ‘part of his concept’; but if so his concept is defective and needs to be replaced.”[81]

If we adopt Field’s non-classical theory – that is, if we restrict the law of excluded middle, define the new conditional →, take truth to be naïve truth, and stratify a determinately-true operator - then we have a prescription for a consistent theory that is ‘revenge-immune’ in Field’s sense. The trouble is that the notions of determinate truth, defectiveness, truth-value gaps, exclusion negation, and others, don’t really go away, and we’ve only side-stepped the Liar.[82]

I’ve been arguing that Field’s theory is too distant from natural language. This worry has a number of strands: the technical nature of Field’s conditional, the lack of a matchup between Field’s semantic model and natural language, and Field’s claim that seemingly intelligible notions are ultimately unintelligible. But I think there is another strand, which demonstrates a still wider gulf between Field’s theory and ordinary language. This turns on Field’s determinacy hierarchy.

Consider Jane, an ordinary speaker who is familiar with the Liar. Given a liar sentence, Jane will recognize that it is defective, and conclude that it’s not true. And given her familiarity with the Liar, Jane will readily be able to produce anaphoric liar sentences. For example, given the sentence, say, “‘2+2=5’ is not true”, it’s a simple matter to produce the liar sentence “’2+2=5’ is not true, and neither is this very sentence”.

So since Jane is familiar with the Liar, she can readily produce the following stretch of reasoning, which combines her readiness to move from a sentence's being defective to its being untrue, and her ability to form anaphoric liar sentences. Suppose there is a liar sentence L on the board. Here's how Jane reasons:

L is defective, and so:

(1) L is not true.

Now form a new anaphoric liar sentence, building on (1):

(1*) L is not true and neither is (1*).

Since (1*) is a liar sentence, (1*) is defective. So

(2) (1*) is not true.

Now form a new liar sentence, building on (2):

(2*) (1*) is not true and neither is (2*).

Since (2*) is a liar sentence, (2*) is defective. So

(3) (2*) is not true.

And so on, through a series of liar sentences (3*), (4*), … .

Here is a simple analysis of the reasoning. It's the analysis that the singularity theory provides, but it's simple enough that it's compatible with other contextual accounts as well (for example, Burge's – see below). Let 'true$_{cL}$' represent the occurrence of 'true' in the liar sentence L. Then in the usual way, we can establish that L is defective – L fails to have truth$_{cL}$-conditions. So Jane's conclusion that (L) is not true, that is, (1) above, is represented by:

(1) (L) is not true$_{cL}$.[83]

And the rest of the discourse is represented by:

(1*) (L) is not true$_{cL}$ and neither is (1*).

Since (1*) is a liar sentence, (1*) is defective. So

(2) (1*) is not true$_{cL}$.

Now form a new liar sentence, building on (2):

(2*) (1*) is not true$_{cL}$ and neither is (2*).

Since (2*) is a liar sentence, (2*) is defective. So

(3) (2*) is not true$_{cL}$

And so on.

This analysis identifies a single truth predicate that does not undergo any change of extension. That is as it should be – it respects the fact that the anaphoric back-reference throughout the discourse is tied to the occurrence of ‘true’ in L. We attribute to Jane the use of a single, constant truth predicate throughout the course of her reasoning – the truth predicate that first appears in L. And there’s a single notion of defectiveness running through the discourse – the failure to have truth$_{cL}$-conditions.

Now Field’s determinacy hierarchy is driven by the need to accommodate the intuition that Liar sentences are *not true*. And this is of course a very natural intuition – given a semantically defective sentence like the Liar, we want to say that it isn’t true. According to Field’s theory, we accommodate the intuition that Liar sentences are not true by introducing the notion of determinate truth – when we say that a Liar sentence such as L is not true, we are to be interpreted as saying that L is not determinately true. And when we say that L is defective, we are saying that neither it nor its negation is determinately true. So on Field’s account, Jane’s (1) is to be interpreted as:

(1) (L) is not determinately true.

(1) involves the use of a notion of truth – determinate truth – that is different from and stronger than the notion used in L.

Now (1*) is:

(1*) L is not determinately true and neither is (1*).

So (1*) is a determinately true liar sentence. (1*) is defective: neither L nor not-L are determinately true. On pain of contradiction, we cannot say that (1*) is not determinately true. The sense in which Jane can say that (1*) is defective is in terms of a notion of truth stronger than not only truth but also determinate truth – neither (1*) nor its negation are determinately determinately true. And the sense in which we can say that (1*) is not true is this: (1*) is not determinately determinately true. So Jane's (2) is interpreted as:

(2) (1*) is not determinately determinately true.

In turn, (3) is interpreted as:

(3) (2*) is not determinately determinately determinately true.

And so on. Jane is interpreted as employing a sequence of stronger and stronger notions of truth and defectiveness that are further and further removed from ordinary language.

Field does take these notions to be outside ordinary talk. "The fact is", he writes, "that people rarely iterate determinately operators very far".[84] Yet here Jane is interpreted as iterating the determinately operator repeatedly – and this is implausible, given how few resources she needs to carry out this reasoning. The reasoning constantly repeats the same cycle of steps, and we have no reason to identify any increasing conceptual complexity. Jane's reasoning requires only the use of the same truth predicate throughout, expressing the same notion of truth.

The worry about Field's determinacy hierarchy is that it fails to respect the intuitions that motivated the hierarchy in the first place. If the claim that any Liar sentence is defective really is a natural claim, then any account of it should be in terms readily available to the ordinary speaker. Similarly with the inference from 'A is defective' to 'A is not true'. Even if we suppose that determinate truth, as defined by Field in terms of →, is readily available, it is clear that iterations of it are not. If the evaluation of a liar sentence as defective or not true is taken to involve an iterated determinately operator, then we have left behind any ordinary sense of 'defective' or 'not true'. It is natural for Jane to say of any of the liar sentences (1*), (2*), (3*), … that it is defective and not true. She has the few resources she needs to do that. To attribute to Jane the claim that, say, (3*) is not determinately determinately determinately determinately true is unrealistic, to say the least.

The same kind of criticism of Field's account can be given when we broaden the perspective beyond examples like Jane's discourse, and compare Field's theory with contextual theories generally. The singularity theory is one; Burge's hierarchical account is another. The singularity theory will identify Jane's (1) as a repetition of (L), and (1) is true when evaluated reflectively (since (L) is a singularity of the occurrence of 'true' in (L)). A contextual-hierarchical view such as Burge's will say that (L) is not true$_{cL}$, as (1) says – and since (L) says the same thing, (L) itself is true – not true$_{cL}$, but true$_{rL}$, say, where 'true$_{rL}$' has a broader extension than 'true$_{cL}$'.[85] These two accounts will agree on the simple analysis of Jane's discourse. They'll also agree on the idea that 'true' is a single context-sensitive expression that shifts its extension according to context. But the singularity theory does not stratify the truth predicate, while Burge's account does. Despite this difference, both can claim the following advantage over Field's theory – both stay close to ordinary language, while Field's departs from it.

Let's consider this claim in more detail. Since Field offers a hierarchical account of determinate truth, a comparison with stratified theories of truth is apt. Field draws the following contrast:

> "in classical truth theories that involve stratification, the stratification consists of there being a whole hierarchy of primitive truth predicates … But in the case of the paracomplete theories… there is no need for a hierarchy of primitives. Rather, there is a single primitive notion of truth, and a single notion of determinateness…"[86]

It may be true of the simplest kind of hierarchical approaches to the liar that there is a hierarchy of distinct primitive truth predicates. But a contextual-hierarchical theory such as Burge's has a single, indexical truth predicate – so on this score, Field cannot claim an advantage, especially since Field admits a hierarchy of "stronger and stronger truth predicates". Field also claims that iteration isn't really stratification because predicates of the form 'D^{α}True' can be significantly applied to sentences containing 'D^{β}True', where β is greater than α.[87] (Consider, for example, '2+2=4 v D^{α}True(S)', for any ordinal α and sentence S – by the strong Kleene valuation, this is D^{0}True.) But Burge can make the parallel claim for his Constructions 2 and 3.[88] And for Field iteration really *is* stratification in the cases that matter – for example, the series Q_{σ} of determinate-liar sentences.

Further, a theory like Burge's has the resources to assess any defective sentence of the language. Consider a Liar sentence L that says of itself that it is not true_{α}, for ordinal α. According to Burge's account, L does not have truth_{α} conditions, and so it isn't true_{α}, and so, since that is what it says, it is $\text{true}_{\alpha+1}$. We're led through strengthened reasoning to a more comprehensive, reflective use of the context-sensitive predicate 'true' by which we assess the defective sentence L. So Burge's theory is just as revenge-immune as Field's, in the sense that any defective sentence of the language can be assessed as defective within the language.[89] Indeed, it might seem that the contextual-hierarchical approach has an advantage here: Burge's theory has a single truth predicate, while Field's theory splits off naïve truth from determinate truth, and then develops a hierarchy of distinct, increasingly strong, notions of truth.

Now Field claims the following major advantage for his determinacy hierarchy: stratification "is applied only to the relatively peripheral notion of determinateness, not to the crucial notion of truth".[90] Determinate truth and its iterations of determinateness are not features of everyday talk: it would take a "fairly fanciful story" for two speakers to be interested in the *determinate* truth of each other's remarks.[91] If we must rest content with an increasing hierarchy, as Field thinks we must, then he prefers that it be the peripheral notion of determinate truth that is stratified, rather than the central notion of truth. But it seems to me that, far from being an advantage, it is a distinct drawback of Field's theory that the determinacy hierarchy is composed of notions that are increasingly remote from natural language.

As we've seen, Field's determinacy hierarchy is driven by the natural intuition that Liar sentences are *not true*. But Liar sentences appear at each level of the determinacy hierarchy, and this intuition extends to all of them – we want to say of *any* Liar sentence that it isn't true. However, the determinacy hierarchy doesn't accommodate our intuition here. On this point it is instructive to compare Burge's hierarchy with Field's.

Burge's account, like the singularity account, seeks a natural way of representing strengthened reasoning, where we reason through the defectiveness of a Liar sentence. According to Burge's account, there is a contextual shift from a less comprehensive use of the truth predicate to a more comprehensive use – consider the case of L again, and the shift from ‘true$_{cL}$’ to the more comprehensive ‘true$_{rL}$’. We abandon the true$_{cL}$-schema (since we find that L does not have truth$_{cL}$-conditions), and assess it as true$_{rL}$ via the more comprehensive true$_{rL}$-schema, which accommodates the defectiveness of L.

It’s the same story at higher levels. We can add to our final evaluation of L as true$_{rL}$ the following perverse addition “but this very sentence isn’t” to obtain a new Liar sentence that may be represented as:

(L$_1$) (L) is true$_{rL}$ but this very sentence isn’t.

L$_1$ is a defective Liar sentence: it fails to have truth$_{rL}$-conditions. And so it isn’t true$_{rL}$. Here the evaluation of L$_1$ as defective and untrue is analyzed in terms of the truth predicate fixed by the context of L$_1$. So the claim that L$_1$ is not true is treated in just the same way as the claim that L is not true: in both cases, we use a single context-sensitive truth predicate, whose extension is fixed by the context of the sentence we are assessing. (This is also true of the non-hierarchical singularity theory.) And so on, as we go up the levels of this truth hierarchy, through further Liar sentences L$_2$, L$_3$, … , and increasingly comprehensive uses of ‘true’. On Burge’s account (and the singularity account), when an ordinary speaker declares a Liar sentence not true, all we need to require of their use of ‘true’ is that it be tied to the context of the sentence they are assessing. Our ability to say of *any* Liar sentence that it is not true is captured in terms of a single, context-sensitive truth predicate. Again, no new notion of truth is introduced and the account stays close to natural language usage.

Notice that the sequence L, L$_1$, L$_2$, … is to be sharply distinguished from Jane’s sequence L, (1*), (2*) … . And, appropriately, Burge’s account handles them quite differently. Burge’s account – like the singularity account -- handles Jane’s discourse via a single truth predicate with a constant extension. But Burge’s account handles the sequence L, L$_1$, L$_2$, … by identifying a shift in the extension of 'true', from less comprehensive to more comprehensive. In the case of Jane's discourse, Burge's account (and the singularity theory) had this advantage: Field's account appealed to a hierarchy where none was needed. In handling the sequence L, L_1, L_2, …, Burge does appeal to a hierarchy. But still Field's theory is at a disadvantage, and for the same kinds of reasons as before.

As we go up Field's determinacy hierarchy, the departure from natural language becomes more and more pronounced: we attribute to a speaker increasingly iterated notions of determinate truth. In contrast, on Burge's account, if a speaker evaluates a Liar sentence as 'not true' – wherever it is in the hierarchy – then their evaluation is analyzed in terms of 'true' as it appears in that Liar sentence. All that is attributed to the speaker here is a use of a single context-sensitive truth predicate, tied to the context of the sentence being assessed. Again, no sequence of stronger and stronger notions of truth is required.

Parallel remarks hold for the notion of defectiveness. On Burge's account, when we say that a Liar sentence is defective, this is analyzed as the failure to have truth-conditions assigned to it by its associated truth-schema. On Field's account, to evaluate a Liar sentence as defective requires the introduction of a new, stronger notion of truth, more remote from ordinary language than anything contained within the Liar sentence itself.

Both hierarchies – Burge's truth hierarchy and Field's determinacy hierarchy – are motivated by the observation that we can evaluate Liar sentences as defective and untrue. Whatever means we employ for that (whether an appeal to context-sensitivity, or a distinction between truth and determinate truth, or some other), there is the prospect that new paradoxical sentences will emerge. But if we were originally motivated to accommodate the claims that Liar sentences are defective, and not true, then we should want to accommodate those claims for the new paradoxical sentences too. Whether or not we are persuaded by Burge's account, it provides a clear example of a theory where these subsequent paradoxical sentences can be evaluated as defective, and as not true, in just the same way as the original ones, without introducing new notions of truth, and without going beyond the semantic repertoires of speakers. If Max has produced a Liar sentence at some level or other of Burge's hierarchy, and you are told on unimpeachable authority that Max has produced some Liar sentence or other, it is natural for you to infer that what he said is defective and untrue. Now according to Burge's account, if 'true_M' represents Max's use of 'true', then your evaluation that what he said is defective is understood as saying that it's neither true_M nor false_M; and your evaluation that what he said is not true is understood as saying that it's not true_M. Whatever position Max's utterance occupies in the hierarchy, however high it is, the story is the same: your evaluation is analyzed simply in terms of a truth predicate tied to Max's context of utterance. But on Field's accounts, if you're told that Max has produced some Liar sentence, and you draw the natural conclusion that it is defective and not true, then your evaluation will be interpreted as involving as many iterations of the determinately operator as are required by the level of Max's utterance. This is surely implausible, even for very low levels of the hierarchy, let alone high levels.

I think that this shows that Field's determinacy hierarchy moves off in the wrong direction. The hierarchy does not illuminate our ability to assess semantically defective sentences. Neither Burge's theory nor the singularity theory requires a further notion of truth, or a further series of notions of truth – they require only a single context-sensitive predicate. But on Field's theory, determinate truth is separated off from truth, and we have to be prepared to attribute to ordinary speakers increasingly technical and artificial notions of truth. Perhaps the determinacy hierarchy is suited to a resolution of the Sorites paradoxes, where our motivation to avoid sharp cutoffs may encourage iterations of the determinateness operator. But it does not seem well-suited to the Liar.[92]

Field says less about the paradoxes of denotation, but it seems that the story there will be similar. According to the singularity account, our ability to say of a pathological denoting expression, like C, that it is defective does not require a new notion of denotation. It requires only a use of 'denotes' tied to the context of C – a repetition of that use of 'denotes', though in a different context. According to Field, definability paradoxes show that the notion of definability in a given language does not have sharp boundaries (in common with a vague term like 'old'), and in response we should restrict the law of excluded middle for the notion of definability (in a given language).[93] Determinateness enters the picture, as it does with vagueness, and we cannot say of a Berry- or König-like phrase that it either determinately defines or determinately fails to define a number within the given language. Higher order paradoxes threaten (in parallel to the worry about higher-order vagueness), but presumably we should restrict the law of excluded middle for *determinately defines* and subsequent iterated notions.[94] As before, our natural inclination to say of any defective denoting expression that it does not denote will be treated in terms of notions that involve the iteration of the determinately operator, notions that are artificial and far removed from ordinary language.

8.4 Dialetheism and revenge

If expressive incompleteness signals a failure to deal with paradox, and if second-order revenge forces expressive incompleteness on any consistent theory, then perhaps inconsistency is the price we should pay. According to the dialetheist, there are true contradictions, and liar sentences, for example, are both true and false. In classical logic, of course, everything follows from a contradiction – and the dialetheist cannot allow that everything is true. So the contradictions associated with the paradoxes are quarantined by some suitable paraconsistent logic. Accept these quarantined contradictions and the paradoxes are tamed. The very notion of revenge seems misplaced now, for what worse could a purported revenge paradox produce than a contradiction? For the price of inconsistency we can buy expressive completeness. However, despite appearances, there are revenge paradoxes for the dialetheist. Since dialetheists focus mainly on truth, I shall focus here on revenge *liars*.

According to the dialetheist, some sentences are true ('2+2=4'), some are false ('2+2=5'), and some, like the Liar sentences, are both true and false. (Dialetheists differ over truth value gaps – Priest, for example, rejects gaps, while other dialetheists admit them. For simplicity, I will consider only dialetheism without gaps.) Some Liar sentences, such as

(1) This sentence is not true

are true and false, *and* not true. Given a sentence that is true and false, it may further be the case that the sentence is not true (or not false, for that matter). According to Priest, the information that (1) is not true is more information, *in addition* to the information that (1) is true.[95]

Now it is natural to think that a revenge liar for the dialetheist is generated by the sentence:

(2) (2) is false only.[96]

(2) is a Liar sentence, so according to the dialetheist, it is both true and false. Since it is true, (2) is false only. So (2) is both true and false, and false only. If we now claim that (2) cannot be both true and false, *and* false *only*, the dialetheist will say that it can, and in fact is – by dialetheist's lights, being true and false does not preclude being false only. That (2) is false only is additional information, additional to the information that it is true and false. Being false does

not preclude being true, and neither does being false *only* preclude being true. After all, the dialetheist will say, we can capture the status of (2) by

T(2) & F(2) & ~T(2),

where the third conjunct adds more information. We may feel, with some justification, that the dilaetheist is not taking the exclusionary character of 'only' in (2) seriously, and that (2) does pose a revenge problem for the dialetheist. But I think there is a still more clear-cut revenge liar.

Let's assign the value 1 to sentences that are true, and the value 0 to sentences that are false. According to Priest, these are the *only* values. Some sentences relate to these values consistently – '2+2=4' relates to 1 consistently, and '2+2=5' relates to 0 consistently. Liar sentences relate, inconsistently, to both 0 and 1.[97] Priest writes:

> "… there are only two truth values: true and false. Different sentences just relate to them in different (consistent and inconsistent) ways."[98]

These are the basic semantic categories for the dialetheist: true, false, true and false. But specifying the values that a sentence relates to need not tell the whole story about its semantic status. For example, the values of (1) are 1 and 0 – and, in addition, (1) is not true. The values of (2) are 1 and 0 – and, in addition, it is false only.

Now define the *value set* of a sentence A as the set of A's values. The value set of '2+2=4' is 1; the value set of '2+2=5' is {0}. The value set of sentence (1) is {1,0}; the value set of (2) is {1,0}. A sentence can have only one value set. Non-paradoxical sentences that are true have the value set {1}, non-paradoxical sentences that are false have the value set {0}.[99] And paradoxical sentences, according to the dialetheist, have the value set {1,0}. If a sentence A is paradoxical, so that its value set is {1,0}, we can *at least* say this about A:

T(S) and F(S).

This may not, however, complete its semantic profile. There may be more to say about A -- A may also be, for example, untrue or false only. But A's value set will still be {1,0}. For example, the value set of (2) is {1,0}, even though its semantic profile is, as we saw,

T(2) & F(2) & ~T(2),

where there is a conjunct beyond T(2) and F(2). The third conjunct adds information additional to the first two conjuncts, and, for the dialetheist, it does not in any way cancel or remove the first conjunct.[100] For any paradoxical sentence A, the conjuncts T(A) and F(A) will be part of its semantic profile. And since truth and falsity are the only values, no conjuncts of A's semantic profile other than T(A) or F(A) can contribute to the value set of A. If A is paradoxical, its value set is {1,0}; A's value set cannot be smaller or larger.

Now consider the sentence

(3) The value set of (3) is {0}.

Suppose first that the value set of (3) is {1}. So (3) is true. By the truth-schema - which Priest endorses - it follows that the value set of (3) is {0}. Since any sentence has just one value set, it follows that {1}={0}, so 1=0, and everything is true. This is unacceptable to the dialetheist.[101] Suppose second that the value set of (3) is {0}. Then, by the truth-schema, given what (3) says, (3) is true. So the value set of (3) is either {1} or {1,0}. Either way, 1=0 again. Suppose third that (3) has value set {1,0} – this is presumably the option most in line with dialetheism, since (3) is a liar sentence. Then (3) is true (as well as false). By the truth-schema, the value set of (3) is {0}. So, since (3) has only one value set, {1,0}={0}, and so again 1=0. All three cases lead to triviality. So (3) generates a revenge liar for the dialetheist.

We should be careful to distinguish this revenge liar from other attempts to produce problems for the dialetheist. Consider the 'false only' liar, generated by the sentence (2). The problem, supposedly, is that the Liar reasoning yields an unacceptable result, that (2) is true, false *and* false only. Priest's response to this purported revenge liar is that this result is perfectly acceptable by dialetheist lights: *false only* and *true and false* are not mutually exclusive. Of course, (2) is related to truth and falsity in inconsistent ways – but that's part of the dialetheist diagnosis of the Liar. But, in contrast, the reasoning about (3) yields not an inconsistency, but the triviality result that everything is true.

It's worth noting that this revenge liar does not preclude a sentence from being *false only* and true, since the semantic profile of (3) can be given as:

$$T(3) \,\&\, F(3) \,\&\, {\sim}T(3).$$

But from this we can read off the value set for (3) – it's given by the first two conjuncts as {1,0}. Again, this is the value set for (3) in line with the dialetheist account of liar sentences – but assuming that (3) has value set {1,0} leads to triviality, as we just saw.

In a discussion of revenge paradoxes,[102] Priest considers a liar related to (2), but apparently more damaging – because, like (3), it yields the triviality result.[103] Instead of talking in terms of a sentence being true or false, we introduce the notion of the semantic value of a sentence. Let 'Val(A)' abbreviate 'the value of A'. Since Priest does not admit gaps, we have T(A)vF(A), for any sentence A. And given the Law of Excluded Middle (which Priest endorses), we can show that:

Trichotomy (T(A)&~F(A)) v (~T(A)&F(A)) v (T(A)&F(A))

So it is natural to define Val as follows:

(i) Val(A) = {1} iff T(A)&~F(A)

(ii) Val(A) = {0} iff F(A)&~T(A)

(iii) Val(A) = {0,1} iff T(A)&F(A)

Now consider the sentence:

(4) Val(4) = {0}

By Trichotomy, Val(A) = {1} v Val(A) = {0} v Val(A) = {1,0}. In the first and the third cases we have T(4), so by the truth-schema, Val(4) = {0}. So in the first case, we have Val(4) = {1} and Val(4) = {0}, and in the third case, we have Val(4) = {0,1} and Val(4) = {0}. In either case, we obtain 0=1. So for any A, Val(A) = 1, and we have the triviality result – everything is true. In the second case, where Val(A) = {0}, it follows from the truth-schema that Val(A) = {1}. Again, we obtain 0=1, and the triviality result.

In response, Priest argues that Val is not a well-defined function. As with (1) and (2), a dialetheist account of (4) will yield: T(4) & F(4) & ~T(4).[104] But then cases (ii) and (iii) of the definition of Val overlap, since we have both ~T(4)&F(4) and T(4)&F(4). So the 'function' Val is not well-defined – it produces two distinct outputs for input (4).

But this way out cannot be taken with the paradox generated by (3). The dialetheist account of (3) will yield, in parallel with (4), the result that T(3) & F(3) & ~T(3).[105] But from this it follows that (3)'s value set is {1,0} and nothing else. The value set of a sentence A is the set of *all* A's values, not just some of them.

Priest points out that if we switch from functions to relations, we can readily obtain a well-defined relational analogue of Val. We can define the relation Rel as follows:

Rel(A,1) iff T(A)

Rel(A,0) iff F(A)

The would-be Liar sentence is now

(5) Rel((5),0) & ~Rel((5),1).

From (5) we can infer

$$\mathrm{Rel}((5),1)\ \&\ \mathrm{Rel}((5),0)\ \&\ {\sim}\mathrm{Rel}((5),1),$$

and this is perfectly acceptable to the dialetheist.

But now Priest considers a new attempt to reinstate revenge by defining a suitable function in terms of Rel. Using set-abstraction, we can define Val* as follows:

$$\mathrm{Val}^*(A) =\{x \mid \mathrm{Rel}(a,x)\}.$$

We have:

$$[\mathrm{Rel}(A,1)\ \&\ {\sim}\mathrm{Rel}(A,0)]\ \mathrm{v}\ [\mathrm{Rel}(A,0)\ \&\ {\sim}\mathrm{Rel}(A,1)]\ \mathrm{v}\ [\mathrm{Rel}(A,1)\ \&\ \mathrm{Rel}(A,0)].$$

And we make it explicit that there are just two values:

(+) $\forall x(\mathrm{Rel}(A,x) \text{ -> } (x=1 \text{ v } x=0)).$

Now the proposed Liar sentence is

(6) $\mathrm{Val}^*(6) = \{0\}.$

Priest goes on to identify a place where the paradoxical reasoning breaks down. Consider the second case, where we assume $\mathrm{Rel}(6,0)\&{\sim}\mathrm{Rel}(6,1)$. We need to get from this to $\mathrm{Val}^*(6) = \{0\}$, so that we can go on to apply the truth-schema. For this inference to go through, we need: $\forall x(\mathrm{Rel}((6),x) \leftrightarrow x=0)$. But we cannot obtain the left-to-right direction of the biconditional for all values of x, since, in analogy with (4) and (5), we have not only ${\sim}\mathrm{Rel}((6),1)$ but also $\mathrm{Rel}(6,1)$. So we cannot use disjunctive syllogism to move from $\mathrm{Rel}((6),0)\&{\sim}\mathrm{Rel}((6),1)$ to $\mathrm{Rel}((6),x)\rightarrow x=0$.

There is no such breakdown in the case of (3). In the analogous second case, we assume that (3) has value set {0}, which is just to assume (3) itself. So the application of the truth-schema to (3) is immediate. In the case of (6), there is a gap between the assumption

Rel((6),0)&~Rel((6),1) and the target sentence Val*(6) = {0} (that is, (6) itself), since Rel((6),0)&~Rel((6),1) does not require (6) to take only the value 0.

Priest considers a stronger version of (5) in a final attempt to generate revenge:

(7) $\mathrm{Rel}((7), 0)\ \&\ \forall x(\mathrm{Rel}((7),x) \rightarrow x=0)$.

The sentence

(5) Rel((5),0) & ~Rel((5),1)

is too weak to require (5) to take only the value 0, and (7) removes that weakness. But still, Priest argues, this attempt at revenge fails, for the same reason that the attempt via (6) fails. In the second case of the reasoning, where we assume Rel((7),0) & ~Rel((7),1), we will need to prove $\forall x(\mathrm{Rel}((7),x) \rightarrow x=0)$, and this is impossible, as before.

In the cases of (6) and (7), the assumptions we make in the second case of the paradoxical reasoning are too weak to get where we want. So a natural thought, taken up by Bromand,[106] is to strengthen those assumptions. Instead of working from Trichotomy, we start with a stronger axiom:

*Trichotomy** $\mathrm{Rel}(A,1)\ \&\ \forall x(\mathrm{Rel}(A,x) \rightarrow x=1)\ \mathrm{v}$

$\mathrm{Rel}(A,0)\ \&\ \forall x(\mathrm{Rel}(A,x) \rightarrow x=0)\ \mathrm{v}$

$\mathrm{Rel}(A,1)\ \&\ \mathrm{Rel}(A,0)\ \&\ \forall x(\mathrm{Rel}(A,x) \rightarrow (x=1\ \mathrm{v}\ x=0))$

Given Trichotomy*, an apparent revenge paradox arises, as follows. The dialetheist is committed to

(++) Every sentence is true only or false only or both true and false.

(Trichotomy*) expresses (++). So the dialetheist is committed to Trichotomy*. And now the argument to the unacceptable conclusion 0=1 goes through, given the stronger axiom Trichotomy*.

But Priest has a response: while the dialetheist is committed to (++), (Trichotomy*) does *not* express (++). Rather, it is (Trichotomy), together with (+), that expresses (++). For example, the claim that A is false only is expressed by

$$\mathrm{Rel}(A,0) \;\&\; {\sim}\mathrm{Rel}(A,1) \;\&\; \forall x(\mathrm{Rel}(A,x) \rightarrow x=1 \vee x=0),$$

and not by

$$\mathrm{Rel}(A,0) \;\&\; \forall x(\mathrm{Rel}(A,x) \rightarrow x=0).$$

So the dialetheist need not be committed to (Trichotomy*), and the revenge paradox doesn't get off the ground.

It is clear that the dialetheist cannot accept (Trichotomy*) as an expression of (++). Consider, for illustration, the case of 'false only'. Suppose 'A is false only' was expressed by the second disjunct of (Trichotomy*):

$$\mathrm{Rel}(A,0) \;\&\; \forall x(\mathrm{Rel}(A,x) \rightarrow x=0).$$

Then the Liar sentence

(7) $\mathrm{Rel}((7), 0) \;\&\; \forall x(\mathrm{Rel}((7),x) \rightarrow x=0)$.

will say that (7) is false only – we have a version of the 'false only' paradox. And as we saw the dialetheist response to this is to say that (7) is false only, and true: (7)'s semantic profile is given by $\mathrm{Rel}((7),0) \;\&\; {\sim}\mathrm{Rel}((7),1) \;\&\; \mathrm{Rel}((7),1)$. So an instantiation of the second disjunct of Trichotomy* yields $(\mathrm{Rel}((7),1) \rightarrow 1=0$, which is false, since the antecedent is true and the consequent false.

Unlike this last attempt at revenge, the paradox generated from (3) does not assume that Trichotomy*, or an equivalent, expresses (++). Rather, the revenge paradox from (3) rests on the notion of a value set, and this notion is compatible with the dialetheist way of understanding ‘false only’ (and ‘true only’). In the case of (7), for example, we can say that (7)’s value set is straightforwardly {1,0}. A semantic profile for (7) is Rel((7),0) & ~Rel((7),1) & Rel((7),1), or equivalently, F(7) & ~T(7)) & T(7), and (7)’s value set is easily read off from this, as {1,0}. We can still maintain the dialetheist reading of ‘(7) is false only’ as

$$\mathrm{Rel}((7),0)\ \&\ {\sim}\mathrm{Rel}((7),1)\ \&\ \forall x(\mathrm{Rel}((7),x) \text{ -> } x=1 \lor x=0),$$

since that reading is quite compatible with (7)’s having a value set of {1,0}. And that’s because, for the dialetheist, a false only sentence can also be true.

The reason that (3) generates a genuine revenge paradox lies in the strength of the notion of a value set. The value set of a sentence A is the *totality* of the values 1 and 0 to which A is related, consistently or inconsistently. Suppose we’re given that F(A)&~T(A) (together with (+), which tells us that 1 and 0 are the only values). Then, if we follow the dialetheist, we can say that A is false only. But that’s not enough to guarantee that the value set of A is {0}. By dialetheist lights, there may be more to the semantic profile of A – it may be that, for example, F(A)&~T(A)&T(A), so that A’s value set is {1,0}. If A’s value set is {0}, then T(A) is not a conjunct of A’s semantic profile. And it’s not just that A isn’t consistently related to T -- it’s neither consistently nor inconsistently related to T. It’s not related to T at all. If we accept the intelligibility of the notion of a value set, then *as a consequence* we have

A has value set {1} v A has value set {0} v A has value set {1,0}.

This is the value set version of trichotomy, and its disjuncts provide the three cases of the paradoxical reasoning from (3), leading to 1=0. These disjuncts are mutually exclusive, as with

Trichotomy*. But the role of the value set version of trichotomy in the paradoxical reasoning is different. It is not assumed to express (++). And it is not an axiom, but a consequence of the definition of a value set. If dialetheists are to take issue with the paradox generated from (3), they must take issue not with its version of trichotomy, but with the notion of a value set itself.

A dialetheist might try to model a response to the paradox from (3) on the dialetheist response to the 'false only' paradox. Just as a sentence can be false only *and* true, so, it might be said, a sentence can have value set {0} *and* be true. But if a sentence (A) has value set {0}, then F(A) is a conjunct of its semantic profile. And if (A) is also true, then T(A) is another conjunct of its semantic profile. That establishes that the value set of A is {0,1}. So again {0}={0,1}, and 1=0. The dialetheist might also try to respond by saying that (3) has value set {0} and *doesn't* have value set {0}. But if (3) doesn't have value set {0} then it has value set {1} or {1,0}. And since (3) also has value set {0}, the result again is 1=0.

Alternatively, the dialetheist might challenge the intelligibility of the notion of a value set. But it is hard to see how this could succeed. According to the dialetheist, there are just two values, true and false, and a sentence can be related to these in consistent and inconsistent ways. Some sentences are related to the value 1, and not to the value 0 – their value sets are {1}. Other sentences, like all of the paradoxical sentences we've considered here, are related to 1 and, inconsistently, to 0 as well – their value sets are {1,0}. There are sentences that are not true that have value set {1,0} – those with a semantic profile given by F(A)&~T(A)&T(A), where being untrue does not 'take back' or exclude being true. Once a conjunct of a semantic profile, always a conjunct. Further inconsistent valuations do not remove or cancel it, but add further information. Membership in the value set is not undone, or somehow made indeterminate or unstable, by inconsistency. If we accept that there is a complete accounting of the relations –

consistent and inconsistent - that a sentence bears to the values true and false, then the notion of a value set is not only intelligible, but perfectly intuitive. And how could we understand the dialetheist account, yet not understand the notion of a value set? The notion depends only on there being a fact of the matter, for any given sentence, about the relations it bears to the values true and false.[107] Either T(A) appears as a conjunct of A's semantic profile or it doesn't; either F(A) appears or it doesn't. To deny the intelligibility of the notion of a value set is to deny that there is a fact of the matter whether or not T(A) and F(A) are conjuncts in the semantic profile of certain sentences. But then the semantic profile of a liar sentence would be something essentially incompleteable, or unstable, or indeterminate. And the dialetheist will reject any treatment of the Liar in these terms. Indeed, the dialetheist is committed to the value set of a liar sentence being completeably, determinately and stably {0,1}.

So it seems implausible that the notion of a value set is unintelligible. An alternative is to allow that it is intelligible, but not expressible in the dialetheist language. But then the dialetheist language is expressively incomplete. And it's the worst of both worlds to buy expressive incompleteness at the price of inconsistency.

Or the dialetheist might say that the intelligibility of the notion of value set depends on a prior understanding of the notion of set – and that set theory should be treated in a paraconsistent way. But then, it seems to me, the revenge liar generated from (3) would have demonstrated its significance. It would be a paradox that cannot be treated along the lines that the dialetheist has treated other semantic paradoxes. It would show that in order to resolve semantic paradox, the dialetheist must develop an alternative set theory that will bear on this revenge liar.

Notes to Chapter 8

1. Richard 1905.

2. Kripke 1975.

3. Martin (circulated xerox), Maddy 1983.

4. See Herzberger 1981 for a vivid demonstration of the problem posed by revenge Liar paradoxes. Herzberger 1970 discusses one kind of second-order revenge paradox - paradoxes of grounding.

5. That is, if $<S_1,S_2>\leq<S_1^*,S_2^*>$ then $\varphi(<S_1,S_2>)\leq\varphi(<S_1^*,S_2^*>)$. For a proof, see, for example, Simmons 1993, p50.

6. For a proof of the formal results here – that (1) the extension and anti-extension of T(x) increase with increasing α, (2) there is a fixed point of φ, and (3) the fixed point $\pounds_\sigma$ is the minimal fixed point (i.e. extended by all other fixed points) – see, for example, Simmons 1993, pp.50-52.

7. We can reach a non-minimal fixed point by, for example, throwing the Truth-Teller into the extension of T(x) at level 0. The Truth-Teller will remain true at all subsequent levels. In contrast, the Liar sentence is paradoxical (not just ungrounded), and never receives a truth value at any fixed point.

8. Kripke, in Martin 1984, p.80.

9. So Kripke's theory is vulnerable to direct revenge insofar as *neither true nor false*, or *not true* (in a sense not coextensive with *false*) may be regarded as ordinary, non-technical notions, and suitable as initial targets. See Simmons 1993, Chapter 3 for more on this.

10. See Kripke, in Martin 1984, p.66.

11. Maudlin argues that Kripke succumbs too quickly to direct revenge. Maudlin argues that we can retain the truth predicate of the minimal fixed point as *the* truth predicate, and *still* say, in the object language, that the Liar sentence is not true: "The object language, in this case, contains a truth predicate, and contains negation, and contains individual terms and descriptions that denote the Liar sentence. These afford all the resources one needs to say that the Liar is not true, by means of the Liar itself." (Maudlin, in Beall 2007, p.193). But when we assert the Liar sentence, we are asserting an ungrounded sentence – we cannot *truly* assert the Liar. So, according to Maudlin, we must distance assertion from truth: there are sentences, like the Liar, which are not true but which we can permissibly assert. And now, Maudlin claims, direct revenge is no longer a problem. And second-order revenge is not a problem either: if we declare

a Liar sentence ungrounded, for example, then our claim is a permissible assertion – but it isn't true, and so no new paradox is forthcoming. For Maudlin's theory, see Maudlin 2004 and 2007.

But Maudlin's theory faces a problem of self-refutation. It is a consequence of the theory that the Liar sentence is not true, yet this very consequence is not true. Similarly with the claim that the Liar sentence is ungrounded. A second problem is that the point of assertion seems lost once assertion is divorced from truth – how can we explain what makes a sentence permissibly assertable in the absence of truth? Third, the introduction of the notion of permissible assertion encourages new revenge paradoxes, generated by sentences such as 'This sentence is not permissibly assertable'. Maudlin's response to this new form of revenge appeals to a hierarchy, compromising the unity of his response to the paradoxes, and inviting in all the problems that attend hierarchical approaches. Fourth, since Maudlin's theory takes over Kripke's monotonic fixed point construction, it cannot accommodate exclusion negation. This presents a dilemma: either the theory is restricted to languages that don't contain exclusion negation, or the very notion of exclusion negation is, counterintuitively, to be regarded as incoherent. For more critical discussion of Maudlin's account, see, for example, Scharp 2007 and Priest 2005b.

12. See Simmons 1993, especially Chapters 3 and 4, which contain critical discussions of the approaches of Herzberger, Gupta, McGee and Feferman.

13. Kripke suggests a response along these lines in Kripke 1975, pp.79-80 and fn.34.

14. See Kripke, in Martin 1984, p80, and fn 34.

15. As I indicated in the previous paragraph, Kripke himself seems to encourage the reading that the theory is committed to fully defined truth-value gaps, and a fully defined *grounded* predicate. But I leave this textual matter aside.

16. The label is due to Reinhardt 1986.

17. The consistency proof is based on these observations: (1) There are no contradictions in any fixed point, and (2) the inferences allowed by K3 and the Intersubstitutability Principle never lead "from premises in the fixed point to a conclusion not in the fixed point", or equivalently, "from premises with value 1 to a conclusion with value less than 1" (Field 2008, pp.65-66).

18. Field focuses almost exclusively on Kripke's minimal fixed point construction.

19. Field 2008, pp.68-9.

20. See Field 2008, p70.

21. Field 2008, p.72.

22. Suppose we define the conditional from negation and disjunction in the classical way: A→B is ~AvB. Then A→A is equivalent to ~AvA. But Av~A is ½ when A is ½. So, in the absence

of the law of excluded middle, not even $A\rightarrow A$ is valid. And without a reasonable conditional we cannot begin to accommodate ordinary reasoning.

23. Given the intersubstitutability of A and T(A), the two sides of the truth schema are equivalent to A, and so the schema is equivalent to $A\leftrightarrow A$. And since $A\rightarrow A$ isn't valid in KFS, neither is $A\leftrightarrow A$.

24. Field 2008, pp.72-73. As we saw above, the claim that a given Liar sentence is not true does not appear in the minimal fixed point.

25. Field does recognize a certain arbitrariness in this choice of starting valuation. In Field 2005, for example, he considers other possibilities, and reports that "all seem a bit *ad hoc*" (Field 2005, p.73).

26. Field's conditional provides for a marked improvement over KFS**.** For example, the following inferences are valid: $\vdash A\rightarrow A$; $A, A\rightarrow B \vdash B$; $\vdash {\sim}{\sim}A\rightarrow A$; $\vdash (A\rightarrow{\sim}B)\rightarrow(B\rightarrow{\sim}A)$; $\vdash A\&B\rightarrow A$ (see Field 2003b, p.292).

27. This follows from Field's 'Fundamental Theorem' (see Field 2008, pp.251-2, and pp.257-8).

28. Field 2008, p,253.

29. See Field 2003a.

30. Field 2003, p.140.

31. See, for example, Field 2008, p. 236.

32. Since Q has value ½, so does ~Q, and so $Q\rightarrow{\sim}Q$ has value 1 (since the value of the antecedent is less than or equal to the consequent). So ${\sim}(Q\rightarrow{\sim}Q)$ has value 0, and so the conjunction of this with Q (that is, DQ) has value 0.

33. See, for example, Field 2003b, pp.298-9.

34. Notice that ${\sim}DQ_1$ is equivalent to ${\sim}DT(Q_1)$, given the intersubstitutability of A and T(A).

35. Here is what we can assert: we can assert that Q_1 is not determinately untrue (${\sim}D{\sim}Q_1$ has ultimate value 1), we can assert that Q_1 is not determinately determinately true (${\sim}DDQ_1$ has ultimate value 1), and we can assert that neither Q_1 nor its negation is determinately determinately true (${\sim}DDQ_1\wedge{\sim}DD{\sim}Q_1$ has ultimate value 1). Here is what we cannot assert: we cannot assert that Q_1 is not determinately true (${\sim}DQ_1$ has ultimate value ½), and we cannot assert that Q1 is either determinately true or not determinately true (DQ_1 v ${\sim}DQ_1$ has ultimate value ½). In the terms of Field's construction, Q1 gets value

> "½ at all even ordinals and 1 at all odd ordinals. DQ1 gets value ½ at all even ordinals and 0 at all odd ordinals; ~DQ1 thus has the same value as Q1, as desired. ~D~Q1 gets ultimate value 1, as we might expect: so we can assert that Q1 is *not* determinately *un*true. As for the claim that Q1 is determinately true, its ultimate value is ½, so we can't assert DQ1 v ~DQ1 (and indeed, can reject it). So excluded middle can't be assumed (and indeed, can be rejected) *even for claims of determinateness*: that is, we have a kind of second-order indeterminacy. But we can assert that Q1 isn't *determinately* determinately true. So we can assert that Q1 is BAD_2, where BAD_2(x) means that ~DD(True(x))∧~DD~True(x) " (Field 2003b, p.299).

Notice that for the 'regular' Liar sentence Q, which does not contain the determinateness operator D, badness or defectiveness is a matter of neither Q nor its negation being determinately true; for Q1, which contains a single application of D, defectiveness is a matter of neither Q1 nor its negation being *determinately* determinately true. This continues up a determinateness hierarchy – the defectiveness of a sentence is captured by the next higher level.

36. Field 2008, p.327. See also Field 2003b, p.299 and fn 34, and Field 2007, p.123.

37. See, for example, Field 2003b, p.292.

38. Yablo 2003, pp.316-318.

39. See Field 2003b, p.272.

40. The Lukasiewicz conditional is given as follows, where the possible values of a sentence A are 1, 0, ½, and '|A|' stands for the value of A: |A→B| is 1 if |A| ≤|B|; and if |A|>|B|, then |A→B| is 1-(|A|-|B|). In Field 2008 (p.84ff), Field considers a Lukasiewicz continuum-valued semantics, where the possible values are extended to every real number in the interval [0,1]. Field observes that the semantics works very well for the quantifier-free part of the language – but once we introduce quantification or infinite conjunction, paradox emerges (see Field 2008, pp.92-3). This motivates Field's search for an alternative conditional.

41. Yablo 2003, pp.317-8.

42. Recall that all conditionals receive the value ½ at stage 0. So at the first minimal fixed point, A→A has value ½, and so does S itself. Given that S and T(S) are intersubstitutable, T(S) gets the value ½ at the first minimal fixed point. So the antecedent and consequent of S both receive value ½ at the first minimal fixed point – and since the value of S's antecedent is less than or equal to the value of S's consequent, the conditional S gets the value at the next starting point determined by the first fixed point. But now T(S) gets the same value as S, so the consequent of S is 1 – so the conditional S will be 1 at the next starting point, and at every subsequent starting point. So the ultimate value of S is 1.

43. Here is another of Yablo's cases: let S_n be the sentence $T(S_n) \rightarrow T(S_{n+1})$ (think of an infinite queue of people where the nth person in line says 'If I am speaking truly, then so is the person

behind me'). Field's semantics yields ultimate value 1 to each S_n. This is easily seen: each S_n (S_0, S_1, … S_n, S_{n+1} …) receives ½ in the first minimal fixed point, and so, by the Intersubstitutivity Principle, each T<Sn> also receives ½ in the first minimal fixed point. So each conditional T(S_n) →T(S_{n+1}) receives the value 1 at the next starting point – that is, each S_n gets 1 at the next starting point – since the value of S_n's antecedent is equal to the value of its consequent. And once each S_n gets value 1, it will get the value 1 forever.

Yablo points out that on a classical understanding of this chain of conditionals, every conditional comes out true (if any S_n were false, then S_n would have a false antecedent, in which case S_n would be true). So the classical assignment is not *arbitrary*, but it is *ungrounded* – the truth attributions are not based on any prior facts (Yablo 2003, p.319). So the classical treatment is objectionable because it assigns definite truth values to ungrounded sentences. Similarly, Yablo argues, with Field's account: ultimate values 1 are assigned to ungrounded sentences. One might add that Field's assignment of values is arbitrary too, given that the initial assignment of ½ to any sentence with main connective -> is itself arbitrary. (Field does remark on "some seemingly arbitrary features of the semantics, such as the choice of a starting valuation for conditionals (and the special role of *minimal* Kripke fixed points at each stage of the construction…)" (Field 2008, p277)).

44. At stage 0, the sentence (1) is ½, because it contains ->. Given the Kripke valuation, at the first fixed point E is ½, and so are T[E] and T[~E] and ~T[~E]. So the conditional is 1 at stage 1, since the value of the antecedent is (less than or) equal to the value of the consequent. And it will remain so for ever, since E is always ½. So the absolute value is 1.

45. At the initial starting point, (2) receives the value ½. At the first fixed point, the antecedent and consequent both have value ½, so at the next starting point, the conditional will be 1, and will stay that way – so its ultimate value is 1.

46. The Truth-Teller E has ultimate value ½, so T(E) and ~E also have ultimate value ½. So the two sides of the biconditional T(E)↔~E have the same value, so each conditional has ultimate value 1. And so the biconditional itself has ultimate value 1.

47. And also for certain sentences that do *not* stabilize at ½. Given a sentence A and the conditional 0=0→A, one might hope that the ultimate values of the biconditionals T(A)↔(0=0→A) and T(A)↔~(0=0→A) are respectively 1 and 0. But the *reverse* is true for certain choices of A, for example when A is a Curry sentence. See Yablo 2003, p.321.

48. Yablo 2003, p.321.

49. See Field 2008, p272.

50. Yablo presents other alternatives that give a possible world semantics for → (see Yablo 2003, p.322ff). Field provides a critical discussion of Yablo's approach in Field 2008, p.244ff (see also pp.272-4).

51. Field 2003b, p.273.

52. Field, 2003b, p.307.

53. Priest argues that it will not help to introduce D* into the language while withholding the law of excluded middle for it. Expressive incompleteness will still result. Consider Q* again. If Q* is determinate, then we have Q*v~Q* - but we're assuming that we cannot assert the law of excluded middle for Q*. So Q* cannot be determinate. But then we cannot express this by ~DQ*&~D(~Q*), since that entails ~D(Q*), which is Q* - so by v-introduction, we obtain Q*v~Q*, and a contradiction again. So the theory will be unable to express the indeterminateness of Q*. See Priest 2005a, p.45.

54. Field 2008, p.356.

55. *op. cit.*, p.357.

56. Field 2007, p.140.

57. Field 2007, p.141.

58. See especially Field 2007, pp.120-141, and Field 2008, pp.325-338.

59. See Field 2008, p.333.

60. Field points out that one cannot assume excluded middle for well-behavedness here, on pain of a König-like paradox (see Field 2008, pp.330-331).

61. Field 2008, p.338.

62. Field 2008, p.340.

63. Field 2007, p.119.

64. See, for example, Field 2003b, p.302.

65. Field, 2003b, p.302.

66. Field 2003b, p.303.

67. The reason is that "in order to give a *definition* of semantic value we have to pretend that the quantifiers of the language range only over the members of a given set, namely, the domain of the starting model, rather than over absolutely everything. What we've defined should really be called 'ultimate semantic value relative to the particular starting model M_0'." (Field 2003b, p.303)

68. Field 2003b, p.304.

69. Field 2003b, p307.

70. Field 2008, p.356.

71. *ibid*.

72. Priest writes: "The fact, then – if it is a fact – that the revenge problem for the theory of truth has turned out to be the same as that for ZF is not reassuring." (Priest 2005, p.46)

73. See Field 2003b, p.306.

74. Relatedly, Rayo and Welch argue that the key semantic notion underlying Field's proposal, *having real world value 1* (as opposed to having value 1 relative to a model) *can* be expressed in a higher-order language. This puts pressure on Field's claim that there is no intelligible notion of determinate truth, if we take *having real world value 1* as capturing that notion. And it reinforces the idea that Field's theory escapes paradox at the cost of expressive incompleteness (see Rayo and Welch 2007).

75. For a discussion of exclusion negation, see Field 2008, pp.309-12; for a discussion of truth-value gaps, see Field 2008, pp.70-72, pp.121-141, pp.206-8.

76. Field 2008, p.311.

77. See Scharp, in Beall 2007, pp.277-8.

78. Scharp stresses the difference between what he calls a *weakly internalizable* theory and a *naturally internalizable* theory – see Scharp 2014. Field's theory is weakly internalizable because it contains within it the expressive capacity to evaluate every sentence containing 'true' or 'determinately true' or iterations of 'determinately true'. But according to Scharp, Field's theory is not naturally internalizable because the theory is not applicable to notions, including exclusion negation and a fully defined notion of a truth-value gap, that are expressible in a natural language such as English. Scharp notes that Field's weakly internalizable theory has no need for the object language/metalanguage distinction in order to maintain consistency. But dispensing with that distinction is no guarantee of expressive completeness – for that, natural internalizability is needed.

79. It is true that if one adds, say, a general notion of determinate truth, or exclusion negation, to Field's theory, then inconsistency results. But of course that does not establish the unintelligibility of determinate truth or exclusion negation – it establishes only their inconsistency with features of Field's theory. One might, for example, give up the naïve account of truth and admit exclusion negation. See more on deflationary truth in the next chapter. One might put the point this way: according to Tarski, it is a mark of natural languages that they are universal, in the sense that they have the potential to say anything that can be said in any language. But this is not true of the language of Field's theory – and so it cannot represent a natural language, and so it cannot provide an account of the Liar in natural language.

80. Field 2008, p.17.

81. Field 2008, p.344.

82. See also Scharp 2007, pp.287-290.

83. Recall the discussion in 2.8. Jane is right to conclude that (1) is not true$_{cL}$. If L was true$_{cL}$, then the c$_L$-schema would apply to L, and we'd land in contradiction. So L is not true$_{cL}$. Note again that this conclusion does not lead back to paradox. For that, we'd need the c$_L$-schema – but that schema does not apply to L.

84. Field 2008, p.351.

85. See Burge 1979.

86. Field 2008, p.347.

87. See Field 2008.

88. Burge 1979, pp.102-106 Martin 1984.

89. Revenge 'from outside' threatens Burge's theory, just as it does Field's. In Field's case, there is the threat from an external general notion of determinate truth, or from quantifying over or conjoining all the determinate truth predicates in the hierarchy. In Burge's case, there is the threat posed by quantification over all the levels, and the sentence 'This sentence is not true at any level'.

90. Field 2008, p.349.

91. Field 2008, p.350.

92. There are (subsidiary) problems for Field's hierarchy, as he points out (Field 2008, pp.350-353). One is the Nixon-Dean case transposed to the determinacy hierarchy. What if Nixon and Dean talk about each other's utterances in terms of determinate truth and its iterations? If one iterates further than the other, then his utterance will be determinately true, and the other's false. So they can engage in "superscript contests". (Field criticizes Burge's treatment of the Nixon-Dean case in Field 2008, p.219.) A second problem is generated by this case: suppose I remember that Brown said something semantically defective yesterday, but I can't remember the level of his utterance. I want to report that Brown said something defective, but I run the risk that I will choose too low a level to make a successful report. Field says that he sees no way around the Nixon-Dean case, and acknowledges a risk in the Brown case - but he thinks that both problems are mitigated by the degree to which these iterations are removed from ordinary usage. How likely is it that Nixon will be interested in whether Dean's utterance is determinately determinately true, or that Brown will have iterated the determinately operator more than two or

three times? This distance from ordinary language may help with the Nixon-Dean and the Brown problems, but I've argued that it presents a serious problem for Field's theory as a resolution of the Liar.

As we saw in the Chapter 7, the singularity theory has a natural way of dealing with the Nixon-Dean example – each is a singularity of the other's utterance. And when I say that what Brown said was defective, I am simply denying that it has truth-conditions, as fixed by Brown's context of utterance.

93. See Field 2008, pp.106-108 and pp.291-3.

94. Field discusses the paradoxes of definability of König and Berry in Field 2008, pp.106-108. On pp.291-3, he extends his discussion to paradoxes of denotation more generally, to include paradoxes couched in terms of languages with a description operator, and concludes that his earlier treatment of the definability paradoxes carries over to this more general setting. Field does not directly discuss *determinately defines* paradoxes, but the parallel he draws with vagueness (in Chapter 5, pp.106-8) suggests that they are to be handled by a determinateness hierarchy (as are the determinate Liar sentences).   As I noted in Chapter 4, Field also considers the simple paradox of denotation we discussed in Chapters 3 and 4 (see Field 2008, pp.293-4, note 7). But he seems to regard it as a different kind of paradox of denotation, suggesting that it's to be resolved in terms of set theory.

95. Priest 2006, p.90, fn. 11.

96. Versions of this paradox are discussed by Smiley 1993, Everett 1993, Bromand 2002, and Littmann and Simmons 2004.

97. I set aside here discussion of nonparadoxical but ungrounded sentences, such as the Truth-Teller, that says of itself that it is true.

98. Priest p.90, fn.12.   Elsewhere Priest does seem to endorse the idea that there are many (in fact, infinitely many) dialetheist values. See Littmann and Simmons 2004 for more on this, and for a development of a related, but different, revenge liar for the dialetheist from the one presented below.

99. Again, we're setting aside truth gaps. They are easily accommodated by the addition of empty value sets.

100. To the charge that the third conjunct 'takes back' what the first one says, Priest responds: "It does not: negation is not cancellation" (Priest 2006, fn. 11, p.90). The third conjunct "adds *more* information" (ibid.)

101. Of the conclusion that 1=0, Priest writes that it "is not just an inconsistency, but triviality: eveything is true. This is unacceptable to any rational dialetheist" (*op. cit*., p.89).

102. Priest, *op. cit*., 20.3 (pp.88-92).

103. *op. cit.*, p.89.

104. The dialetheist reasoning runs as follows. (4) is either true or false. First, If (4) is true, then by the truth-schema, Val(4)={0}, i.e. F(4)&~T(4). So in this case, we have T(4)&F(4)&~T(4). Second, if (4) is false, its negation is true, so we have ~(~T(4)&F(4)), and so T(4)v~F(4). By the exhaustion principle (which Priest endorses), ~F(A)->T(A). So from T(4)v~F(4) we have T(4), and it follows from the truth-schema that F(4)&~T(4). So in this second case, we have T(4)&F(4)&~T(4). So in both cases we have T(4)&F(4)&~T(4).

105. The reasoning is as follows: (3) is either true or false. First, if (3) is true, then (3)'s value set is {0}, and so (3) is false and not true. So in this first case, we have T(3)&F(3)&~T(3). Second, if (3) is false, then the value set of (3) is {1} or {1,0}. If the value set is{1}, then (3) is true (and not false); and if the value set is {1,0}, then (3) is true. Either way (3) is true. So by the truth-schema, (3) is false and not true. So in this second case we have T(3)&F(3)&~T(3). So in both cases we have T(3)&F(3)&~T(3).

106. Bromand 2002.

107. There is a separate issue about whether a dialetheist could communicate the idea that a sentence is not true – that is, *really* not true (see Priest 2006, 20.4). That is a separate issue, because here we're concerned only with the semantic facts of the matter, not with whether a dialetheist could communicate them.