Revenge and Context

Keith Simmons
UNC Chapel Hill

I. Direct and second-order revenge

It's 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. This is familiar in the case of the liar. Ignorant of my whereabouts, I write on the board:

(L) The sentence written on the board in room 101 is not true.

But once it's realized that (L) is written on the board in room 101, we can reason in the usual way to the conclusion that (1) is in some way pathological, perhaps gappy or ungrounded. But if a sentence is pathological, it isn't true. So the sentence written on the board in room 101 is not true. And this true conclusion says just what (L) says – so the treatment of (L) as pathological has led to the re-emergence of (L) as a truth.

Though it's rarely noted, paradoxes of denotation and versions of Russell's paradox are also subject to direct revenge. Suppose I write the following expressions on the board:

(A) pi

(B) six

(C) the sum of the numbers denoted by expressions on the board in room 102.

But I've unwittingly written these expressions in room 102. So we can reason as follows. Suppose that (C) denotes a number, say k. Then $k = \pi + 6 + k$, which is absurd. So (C) is pathological, and fails to denote. (A) and (B) denote π and 6 respectively, and (C) fails to denote, it follows that the sum of the numbers denoted by expressions on the board in room 102 is $\pi+6$. But the definite description in the previous sentence that refers to $\pi+6$ is composed of exactly the same words as (C) with the same meanings. So (C) re-emerges as an expression that successfully denotes a number.[1]

Or consider a version of Russell's paradox. Suppose I write on the board (in room 103, of course):

(D) moon of the Earth

(E) unit extension of a predicate on the board in room 103

(where a unit extension is an extension with exactly one member). We now reason: the extension of (D) is a unit extension. So the extension of (D) is a member of the extension of (E). But does (E) have an extension? If it does, the extension is either self-membered or it isn't. Suppose first that it is. Then the extension of (E) has two members, so it is not a unit extension - and so it is not a self-member. Suppose second that it is not a self-member. Then the extension of (E) has just one member, so it is a unit extension - and so it is a self-member. Either way, we have a contradiction, and we conclude that (E) fails to have an extension. Now if (E) does not have an extension, then in particular it does not have a unit extension. So the only unit extension of a predicate on the board in room 103 is the extension of (D). But in the previous sentence there is a predicate composed of the same words as (E) with the same meanings. And since this predicate has a well-determined extension, so does (E).

In all these cases, our reasoning to the conclusion that (L) or (C) or (E) is pathological is not the end of the matter. We reason past pathology; indeed it is the pathological nature of these expressions that provides for their rehabilitation. (L) is true because it is pathological and so not true, which is what it says it is. (C) denotes $\pi+6$ since, given that (C) is pathological, the sum of the numbers denoted by expressions on the board is $\pi+6$. And (E) has a well-determined extension, since, given that (E) is pathological, the only unit extension on the board in the extension of (D). But once rehabilitated, (L), (C) and (E) can exact their revenge: if (L) is true, then it isn't; if (C) denotes $\pi+6$, then the sum of the numbers denoted by expressions on the board is $\pi+6+(\pi+6)$, so that $\pi+6=\pi+6+(\pi+6)$; and if (E) has an extension with just one member, then it has an extension with two members, the unit extension of (D) and the unit extension of (E).

Direct revenge, then, makes 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 if (L), (C), (E), and their ilk are not pathological, what are they?

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, perhaps technical notions – for example, gaps (in truth, reference, or predicate-application), levels of a hierarchy, groundedness, determinate truth, stability, context. Second-order revenge takes these new notions, and constructs new paradoxes for old. 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 – truth, denotation, extension - 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 ordinary 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.

II. Dealing with direct revenge

Direct revenge seems to put the semantic theorist in a bind. Any account of paradox will surely characterize (L) or (C) or (E) as pathological in some way or other. For example, (L) is characterized by Kripke as ungrounded;[2] in one treatment of his paradox of definability, Richard suggests that the analogue of (C) fails to denote;[3] and Martin and Maddy suggest that the extension of a Russell predicate, analogous to (E), falls into a membership gap, failing to belong to its own extension or anti-extension.[4] But direct revenge seems to show that the withholding of a truth value to (L), of successful reference to (C), or of a determinate extension to (E), leads only to the reinstatement of a truth value, a reference, or an extension, and the apparent return of paradox.

But I believe there is a natural way of treating direct revenge. Rather than regarding the reasoning as a threat, we can regard it as intuitive and valid reasoning that instructs us about our use of semantic concepts. There is a common pattern in all three direct revenge discourses. We start with the paradox-producer, the expression (L) or (C) or (E). We reach the conclusion that it is pathological. We then go on to *repeat* the very expression that caused the problem in the first place. But when we repeat the expression, when we use the very same words that compose (L) or (C) or (E), we find that the repeated expression *does* have a determinate truth value, reference, or extension. In each case, we have two tokens of the same type – one pathological, one not. That is, we have two tokens of the same type but with different semantic status. The natural thought is that these tokens are produced in different context with different results. This is a thought that I've explored elsewhere.[5] Let me sketch the main ideas here.

It's a familiar idea that *context acts on content* – consider indexicals like 'I' and 'now'. But it is increasingly being recognized that this is not a one-way street. The reverse direction holds as well: *content acts on context*. Stalnaker writes:

> "context constrains content in systematic ways. But also, the fact that a certain sentence is uttered, and a certain proposition expressed, may in turn constrain or alter the context. ... There is thus a two-way interaction between contexts of utterance and contents of utterances."[6]

At a given point in a discourse, the context will in part depend on what has been said before. For example, the context may change as new information is added to the discourse. Over the last twenty years or so, the kinematics of context-change has been studied by philosophers, semanticists and linguists alike.[7]

According to Stalnaker the connection between context and available information is very tight indeed. Stalnaker writes:

> "I propose to identify a context (at a particular point in a discourse) with the body of information that is presumed, at that point, to be common to the participants in the discourse."[8]

To put it another way, a context is to be represented by the shared presuppositions of the participants[9] - or the 'common ground', to use a phrase from Grice.[10] As new utterances are produced, and new information is made available, the context changes. For a specific example, consider the speech act of assertion: "Any assertion changes the context by becoming an additional presupposition of subsequent conversation."[11]

The shared presuppositions of conversants also figure in David Lewis's account of context-change. Lewis introduces the notion of a *conversational score*.[12] Following Stalnaker, Lewis identifies the set of shared presuppositions of the participants (at a given stage of a conversation) as one component of the conversational score. "Presuppositions can be created or destroyed in the course of a conversation"[13] - and as the set of presuppositions changes, the conversational score changes. Of course, the notion of conversational score is a vivid way of capturing the notion of context. A change in the set of presuppositions is a change of context.

Another component of the conversational score, according to Lewis, is the *standard of precision* that is in force at a given stage of the discourse. Suppose I say 'France is hexagonal'. If you have just said 'Italy is boot-shaped', and got away with it, then my utterance is true enough. The standards of precision are sufficiently relaxed. But if you have just denied that Italy is boot-shaped, and carefully pointed out the differences, then my utterance is far from true enough - the standards of precision are too exacting. The acceptability of what I say here depends on the conversational score, on the context, which in turn depends on what has been said before. The extension of 'hexagonal' shifts with changes of context. Or, for another example, suppose I say 'The pavement is flat' under standards of flatness where the bumps in the pavement are too small to be relevant. Then what I say is true. But if the conversational score changes, and I say 'The pavement is flat' under raised standards of flatness, what I say will no longer be true. But "[t]hat does not alter the fact that it *was* true enough *in its original context*."[14] Like the extension of 'hexagonal', the extension of 'flat' changes with the context.[15]

Let's return to direct revenge. Let $\mathcal{P}$ stand for the pathological token (L) or (C) or (E). It is natural to divide our direct revenge discourses into four segments: first, where I produce tokens on the board; second, where we reason to the conclusion that $\mathcal{P}$ is pathological and fails to denote; third, where we repeat the pathological expression; and fourth where we conclude that the repetition, call it $\mathcal{P}^*$, and hence $\mathcal{P}$, have a semantic value (a truth value, a referent, or an extension).[16]

Now consider in particular the transition from the second to the third segment of the discourse. The culmination of the reasoning of the second segment is the proposition that $\mathcal{P}$ is pathological and fails to have a value. This is new information, and the proposition becomes one of our shared presuppositions, part of the common ground. So in the transition from the second segment to the third, there is a context change - a shift in the body of information that is presumed to be available.[17]

Let us say that the new contexts associated with the third and fourth segments are *reflective with respect to C*. In general, a context associated with a given point of a discourse is *reflective with respect to a given expression* if at that point it is part of the common ground that the expression is semantically pathological, and fails to have a value. So as we move from the second segment to the third, there is a context-change - a shift to a context that is reflective with respect to $\mathcal{P}$. This context-change is an essential ingredient of direct revenge.

How do these changes in context act on content? Let *t* be the semantic term ('true', 'denotes' or 'extension') that appears in $\mathcal{P}$ and $\mathcal{P}^*$. Let 'i' denote the initial context in which $\mathcal{P}$ is produced, and let 't_i' represent a use of a token of *t* that is coextensive with the occurrence of t in $\mathcal{P}$. This by itself is not to commit us to the claim that t is context-sensitive, that the extension of t may shift with context – for all that we've said so far, it may be that all uses of t are coextensive, whatever the context. When we reach the conclusion that $\mathcal{P}$ is pathological, we will have employed a t-schema – either a truth schema ('s' is true iff s) or a denotation schema ('a' denotes b iff a=b) or an extension-schema (a is in the extension of 'F' iff Fa). In each case, we will have employed the t_i-schema, that is, the schema that contains an occurrence of t coextensive with the occurrence of t in $\mathcal{P}$. This is how a contradiction is reached; for example, we reach a contradiction by assessing the sentence (L) by the true_i-schema, obtaining (L) is true_i iff (L) is not true_i.

In the subsequent reflective context, we produce $\mathcal{P}$*, a repetition of $\mathcal{P}$. The occurrence of t in $\mathcal{P}$ is again to be represented by 't_i'. Consider for example the repetition C* of C. Recall our reasoning: since C fails to denote, "it follows that the sum of the numbers denoted by expressions on the board in room 102 is π+6". But C's failure to denote is a failure to denote_i – C's pathology is the result of its assessment by the denotes_i-schema. So we arrive at the sum π+6 as the referent of C* by way of C's failure to denote_i a number. It is because C fails to denote_i, while A and B succeed, that the sum is π+6. So our reasoning is represented as follows: since C fails to denote_i, it follows that the sum of the numbers denoted_i by expressions on the board in room 102 is π+6. The occurrence of 'denotes' in (C*) is represented by 'denotes_i': the occurrence of 'denotes' in C* inherits its extension from that of the occurrence of 'denotes' in C. Similarly, it is because (E) fails to have an extension_i while (D) succeeds that we are led to the conclusion that the only unit extension of an expression on the board in room 103 is that of (D) – and so the occurrence of 'extension' in (E*) is represented by 'extension$_i$'. And the occurrence of 'true' in (L*) ("The sentence written on the board in room 101 is not true") is represented by 'true$_i$', since we infer (L*) from (L)'s failure to have a truth value when assessed by the true$_i$-schema.

So $\mathcal{P}$* is a repetition of $\mathcal{P}$ in a strong sense: it is composed of the same words with the same meanings *and* the same extensions. And yet in our discourse we provide a definite value for $\mathcal{P}$*. But if $\mathcal{P}$* is assessed by the t_i-schema, then, just as with $\mathcal{P}$, no value is forthcoming, So $\mathcal{P}$* is evaluated by a different schema, a schema associated with the reflective context in which $\mathcal{P}$* is produced, a schema that provides for the evaluation of $\mathcal{P}$* in the light of $\mathcal{P}$'s pathology. If we are to respect the reasoning, we must discern a shift in the extension of *t*, and a shift in the schema by which $\mathcal{P}$* is evaluated. For example, (C*) does not denote$_i$, but it does denote – let us say it denotes$_r$, to mark the shift in the extension of 'denotes' when we assess (C*) reflectively, in the light of C's pathology.

What produces this shift in the extension of *t*? The change in context - specifically, the shift to a context which is reflective with respect to $\mathcal{P}$. At the third stage, the reflective character of the context had the effect of disengaging $\mathcal{P}$ from the i-schema. Now, at the fourth stage, it has the effect of bringing into play a new schema - the reflective r-schema. When we assess $\mathcal{P}$*, and declare that it has a value, we assess it in a context where it is part of the common ground that $\mathcal{P}$ is pathological. The schema by which we assess $\mathcal{P}$* provides an assessment of $\mathcal{P}$* *in the light of* $\mathcal{P}$ *'s pathologicality*. For example, here's what the instance of the r-schema looks like in the case of (L*):

(L*) is true$_r$ iff (L) is not true$_i$.

Given the information that now forms part of the common ground – *that (L) is pathological and is not true$_i$* – the right hand side of the biconditional holds, and it follows that (L*) is true$_r$. Similarly for (C*) and (E*). The difference between the assessments of $\mathcal{P}$ and $\mathcal{P}$* is explained this way: $\mathcal{P}$ is assessed by the unreflective t_i-schema, and $\mathcal{P}$* by the reflective t_r-schema. With the change in context, there is a change in the implicated schema. There is no intrinsic difference between $\mathcal{P}$ and $\mathcal{P}$* - the difference lies in the schemas by which they are assessed.

Notice that $\mathcal{P}$ also has a value when assessed by the t_r-schema, the same value that $\mathcal{P}$ * does. So, for example, (E) fails to have an extension and also has an extension. But there is no contradiction here: (E) fails to have an extension$_i$, but does have an extension$_r$. Compare Lewis's treatment of 'hexagonal' or 'flat'. Sometimes an utterance of 'France is hexagonal' (or 'The pavement is flat') is true, and sometimes it isn't. The extension of the predicates 'hexagonal' or 'flat' depend on the conversational score, in particular on the standards of precision that are in force. In a loosely analogous way, whether or not it is true to say that (E) has an extension will depend on the standard of assessment: do we apply the unreflective t_i-schema or the reflective t_r-schema?

The upshot is that t is a context-sensitive term that may shift its extension, depending on the schema of assessment that is in force in the given context. This in turn depends on the common ground, the information that is presumed to be available – in particular, information concerning the pathologicality of denoting expressions. When $\mathcal{P}$ is first assessed, the information that it is pathological is not part of the common ground. The initial schema of assessment is unreflective with respect to $\mathcal{P}$. Once the information that $\mathcal{P}$ is pathological is incorporated into the common ground, we have a new standard: the subsequent schema of assessment is reflective with respect to $\mathcal{P}$. We have identified a contextual parameter – the *reflective status* of a context – to which the term t is sensitive.

If we do not attend to our ability to reason past pathology, the claim that t is a context-sensitive term will come as a surprise, and *reflective status* will not be an obvious contextual coordinate (unlike the familiar coordinates of speaker, time and place, for example). But once we pay careful attention to discourses where we reason past pathology, it is natural and intuitive to conclude that t is indeed sensitive to the reflective status of a context. Cresswell once wrote:

> "It seems to me impossible to lay down in advance what sort of thing is going to count [as a relevant feature of context] … The moral here seems to be that there is no way of specifying a finite list of contextual coordinates."[18]

Along with Cresswell, Lewis, Stalnaker and others, we should be open to contextual coordinates beyond the familiar ones. If we recognize that content acts on context, that new information or new presuppositions can change the context, then we can identify contextual coordinates that we might otherwise miss. Reflective status is such a coordinate. When we conclude that $\mathcal{P}$ is pathological, a new presupposition is created, and the context changes. And the context-change is a change in reflective status. The key difference between context i and context r is the difference in reflective status:

context *i* is not reflective with respect to $\mathcal{P}$, but *r* is. If we are after the extension of some token of the indexical 'I', we must determine who the speaker is in the given context. If we want to know whether $\mathcal{P}$, or any semantically pathological expression, is in the extension of a token of *t*, we must determine whether or not the given context is reflective with respect to that expression.

It is not initially obvious that 'true' or 'denotes' or 'extension' is context-sensitive, or that reflective status is a contextual coordinate along with speaker, time and place. But if these claims provide the most plausible treatment of direct revenge, then surprise can give way to an improved understanding of how our semantic terms work. This is why we study paradoxes: we hope to learn from them.

III. Second-order revenge

Consider Kripke's theory of truth, and its central notion of groundedness. Though the object language $\mathcal{L}$ of Kripke's minimal fixed point is semantically closed with respect to its truth and falsity predicates, it is not with respect to 'grounded'. If we add the 'grounded' predicate to $\mathcal{L}$, 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 $\mathcal{L}$.

This is typical of second-order revenge: the semantic theorist is forced to accept expressive incompleteness on pain of contradiction. We start with a target semantic notion – in Kripke's case, truth – and provide a theory of that notion which is not vulnerable to the associated paradoxes. In Kripke's case, we have a precise characterization of a language that can express consistently its own notion of truth. But nevertheless $\mathcal{L}$ is expressively incomplete – it cannot express the semantic notions introduced by the theory, such as groundedness. 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, definite truth, determinate truth, or fuzzy truth, and so on.[19]

This failure of expressive completeness seems to compromise a theory's claim to resolve semantic paradox: a second-order revenge paradox is a semantic paradox beyond the scope of the given theory. 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.[20] Terms like 'true' and 'denotes' are terms of ordinary language; terms like 'grounded' and 'stably true' are not. So, for example, if Kripke's minimal fixed point language $\mathcal{L}$ 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 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 truth, denotation and extension 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 (‘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.[21]

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 relate just to the value T ('2+2=4'), some relate just to the value F ('2+2=5'), and some, like the liar sentences, relate both to T and to F.[22] (Some dialetheists, though not Priest, will also allow that there are sentences that relate to neither value. For simplicity, I'll set this form of dialetheism aside.) Let the *evaluation set* of a sentence be the set of values to which the dialetheist relates the sentence. So, for example, the evaluation set of '2+2=4' is the unit set {T}, the evaluation set of '2+2=5' is the unit set {F}, and the evaluation set of 'This sentence is false' is the set {T, F}. Now consider the sentence:

(X) The evaluation set of (X) is {F}.

Since X is a liar sentence, the dialetheist will say that it relates to both T and F. So we can infer:

(i) The evaluation set of (X) is {T, F}.

But since X is true, it follows that:

(ii) The evaluation set of (X) is {F}.

From (i) and (ii) we obtain that {T, F} = {F}, from which it follows that T = F.[23] But then, since everything is true or false for our dialetheist (there are no gaps), everything is true. This is unacceptable to the dialetheist – and we have a (second-order) revenge liar.[24]

How might the dialetheist respond? Perhaps along the following lines. The sentence (X) is related *just* to F (as it truly says of itself), and it's *also* related to both T and F. In terms of evaluation sets that is to say that the evaluation set of (X) is {F}, *and* that the evaluation set of (X) is {T, F}. This is inconsistent - but the dialetheist is not constrained by consistency when dealing with the liar.

The trouble with this response is that, on dialetheist grounds, the evaluation set of a sentence will be unique. There will be just one set of values to which a sentence is related by the dialetheist account. If a dialetheist says that a sentence is related to T, then T is in the evaluation set of the sentence, whether or not the dialetheist also says that the sentence is related only to F. So, in particular, if the dialetheist says that (X) is related just to F, that is not to say that {F} is the evaluation set of (X) – it will depend on what *else* the dialetheist relates (X) to. Consider the likely dialetheist response to the sentence:

($L^{\dagger}$) This sentence is false only.

This liar sentence is true and false according to the dialetheist, and since it's true, it's false only. But the dialetheist will say that ($L^{\dagger}$) is false only, *and* both true and false. In relational terms, one can put it this way: ($L^{\dagger}$) is related just to F, and is also related to both T and F. Being false only does not rule out the additional information that ($L^{\dagger}$) is true too. In general, for the dialetheist, being false does not preclude being true, and neither does being false *only* preclude being true. When we list all the values to which ($L^{\dagger}$) is related by the dialetheist, there will be a single, determinate list, namely, the list T, F. Similarly for (X). The evaluation set for any liar sentence will be the unique set {T, F} because the dialetheist will relate any liar sentence to both values, whether or not she will also allow that it is false only (or true only) as well. So the problem posed by (X) remains: (X) has a unique evaluation set, with the consequence that T=F.

Notice also that dialetheists do not suggest that the semantic status of liar sentences is in any way unstable or ambiguous. It is no part of the dialetheist account that the truth value of a liar sentence can somehow shift, say, from *true and false* to *false only*. According to some theories of truth -- such as the revision theory of truth or contextual theories - we should pay close attention to shifts in our evaluations of liar sentences. But dialetheists reject these theories. According to the dialetheist, paradoxical sentences are supposed to receive a single, stable evaluation – they're true and false. An evaluation that was somehow ineffable or inexpressible would be quite against the spirit of dialetheism: part of the motivation for the view is to avoid counterintuitive restrictions on expressibility. For the dialetheist, the evaluation set of a Liar sentence does not change, and neither is it inexpressible.

Why does (X) pose a problem for the dialetheist? The case of ($L^{\dagger}$) provides a clue, I think. There, the dialetheist's response turns on a certain kind of open-endedness to the evaluation of ($L^{\dagger}$). If the value *false only* were to close off even the inconsistent addition of the value *true* (or vice versa), then ($L^{\dagger}$) would generate a revenge paradox. But the addition of *true* to *false only* (or vice versa) is acceptable by dialetheist lights:

neither *false only* nor *true and false* counts by itself as a complete evaluation of ($L^{\dagger}$) – neither tells the whole story. But the notion of an *evaluation set* forces the dialetheist to tell the whole story about a sentence, in the sense that the members of the evaluation set will exhaust the values to which the dialetheist relates the sentence. Since the dialetheist is committed to the truth of (X), (X) truly identifies its own evaluation set – that is, whatever values are in the identified set are the values to which the dialetheist relates (X). Given the uniqueness of the evaluation set, we obtain the result that {F} = {T, F}, and the consequence that T=F.

IV. Dealing with second-order revenge

Let's return to the contextual view defended in (II). There I explained the difference between the assessments of $\mathcal{P}$ and $\mathcal{P}$* this way: $\mathcal{P}$ is assessed by the unreflective t_i-schema, and $\mathcal{P}$* by the reflective t_r-schema. There is a difference between the extension of the terms t_i and t_r – for one thing, $\mathcal{P}$ and $\mathcal{P}$* are in the extension of t_r, but not in the extension of t_i. What else is excluded from the extension of t_i? And what is the relation between the extensions of t_i and t_r?

A possible response here is a Tarskian one: t_i and t_r are associated with distinct levels of language.[25] The predicate t_r is associated with $\mathcal{P}$'s unreflective context of utterance; the predicate $\mathcal{P}$* is the more comprehensive denotation predicate of a semantically richer language associated with a context reflective with respect to $\mathcal{P}$. On such a hierarchical account, the extension of t_i is properly contained in the extension of t_r.

There are a number of *prima facie* worries about such a hierarchical approach. For one, the stratification of English into a hierarchy of languages seems artificial. For another, we can ask how levels can be assigned to occurrences of *t* in a systematic way. But let me focus here on the problem of revenge for the hierarchical approach. Second-order revenge helps itself to the talk of levels and constructs new paradoxes generated by sentences and phrases such as (a) 'This sentence is not true at any level', or (b) 'the least ordinal not denoted by an expression at any level' or (c) 'non-self-membered extension of a predicate at some level'. Revenge here seems to force a choice: disallow unrestricted talk of all levels (expressive incompleteness) or ascend to a theoretical metalanguage which absorbs talk of all levels. The newly introduced semantic notions – *true at some level*, *denotes at some level*, *has an extension at some level* – apply more widely that any occurrence of *t*, which on the hierarchical account must always occur at some particular level. For example, the predicate 'denotes at some level' applies to every denoting phrase at every level of the hierarchy – and no occurrence of the context-sensitive term 'denotes', appearing at some level of the hierarchy, applies so comprehensively. The pattern then is the familiar one: revenge is avoided either by admitting expressive incompleteness or by ascent to a metalanguage.

I want to suggest an alternative contextual account. Here I will provide a brief sketch, and then go on to consider revenge.[26] This alternative is in a strong sense anti-hierarchical: there is no stratification of the semantic term *t*. The leading idea is that *t* applies almost everywhere, except for certain singular points, or *singularities*.[27] More precisely, it is *occurrences* of *t* that have singularities. For example, $\mathcal{P}$ is a singularity of the occurrence of *t* in $\mathcal{P}$ ((L) is a singularity of the occurrence of 'true' in (L), (C) a singularity of ‘denotes’ in (C), and (E) a singularity of ‘extension’ in (E)). In general, suppose we are given a context α and a phrase or sentence σ containing the term *t*. If σ cannot be given t_α conditions, if it cannot be evaluated by the t_α-schema, then σ is a *singularity* of t_α. And if σ is a singularity of t_α, and α is σ's context of utterance, then σ is *pathological*. So, for example, (C) is a singularity of ‘denotes_i’, and it is pathological too, since the subscript stands for (C)'s context of utterance. (C*) is also a singularity of 'denotes_i', but (C*) is *not* pathological, since its context of utterance is the reflective context r, and (C*) does denote_r. Similarly with (L) and (L*), and (E) and (E*).

It’s the job of the singularity theory to provide a systematic way of identifying the singularities of a given occurrence of *t*. Notice that something is missing if we represent, say, the token (C) as an ordered pair <type(C),i>, where the first element is the type of (C), and the second indicates the appropriate representation of 'denotes' in (C) (viz., 'denotes_i’). This representation does not distinguish (C) from (C*), yet the former denoting expression is pathological and the latter isn't. There is something more to consider: the schema by which (C) is given denotation conditions.

In the second segment of the revenge reasoning, (C) is evaluated by the i-schema; in the fourth segment, (C*) is evaluated by the r-schema, a schema that is reflective with respect to (C). So we capture the discourse more perspicuously if we represent (C) by the ordered *triple* <type(C),i,i>, where the third element indicates that the schema by which (C) is assessed is the i-schema, and (C*) by the triple <type(C),i,r>, indicating that (C*) is assessed by the r-schema. In the course of the revenge reasoning, (C) is evaluated by the i-schema, and it is this evaluation that leads to the conclusion that (C) is pathological; and (C*) is evaluated by the r-schema, and it is this evaluation that leads to the conclusion that (C*) has a determinate denotation. So if we are after an analysis of the revenge discourse, the representation <type(C),i,i> of (C) is privileged, and <type(C,i,r> is likewise a privileged representation of (C*). Call these the *primary representations* of C and C*.[28] In general, a primary representation of a sentence or phrase σ will represent σ as evaluated by the α-schema, where α is σ's context of utterance. We are interested in identifying semantic pathology, and if a token is pathological, it will lack a value if assessed by its associated α-schema. If pathology is there to be found, primary representations will help to reveal it.

We will characterize the notions of *pathology* and *singularity* via a certain kind of tree. Consider, for example (C). (C) makes reference to denoting phrases, and to determine C's denotation we must first determine what these phrases denote - denote_i, that is, because that occurrence of 'denote' in (C) is represented by 'denotes_i'. So the appropriate schema by which to assess the phrases to which (C) refers is the i-schema. All this is captured by the *primary tree* for (C). The top node of the tree is the primary representation of (C), the triple <type(C),i,i>. This is the node at the top of the tree. At the second tier are the phrases to which (C) makes reference, namely (A), (B) and (C), where (C) is represented as evaluated by the i-schema (and for simplicity (A) and (B) are represented simply by their types, since they contain no context-sensitive terms). So the primary representation of (C) appears again at the second tier, and this in turn generates a third tier of nodes. And so on, indefinitely. The primary tree for (C) is unfounded:

```
            <type(C),i,i>
          /      |         \
type(A)   type(B)    <type(C),i,i>
                           /       |       \
                  type(A)   type(B)    <type(C),i,i>
                                           /         |        \
                                  type(A)   type(B)      .
                                                          .
                                                           .
```

We now say that (C) is pathological because *its primary representation repeats on its primary tree*. The unfounded tree shows that (C) cannot be assessed by the i-schema – and so we can also say that (C) is a *singularity* of ‘true_i’.

Contrast the primary tree for (C*). The primary representation of (C*) is <type(C),i,r>, and the primary tree for (C*) is:

```
            <type(C),i,r>
          /      |         \
type(A)   type(B)    <type(C),i,i>
                           /       |        \
                  type(A)   type(B)      <type(C),i,i>
                                              /       |       \
                                  type(A)   type(B)      .
                                                          .
                                                           .
```

The phrases at the second tier are represented as evaluated by the i-schema, since the occurrence of ‘denotes’ in (C*) is represented by ‘denotes_i’. (More formally, the second member of the triple at the top of the tree is the third member of any ordered triple at the second tier.) At the second tier, then, we have the primary representation of (C), which then repeats at all subsequent tiers. But the representation of (C*) does not repeat, indicating that, unlike (C), (C*) is *not* pathological, and *not* a singularity of ‘true_r’. Intuitively, (C*) stands above the circle in which (C) is caught. (C)'s pathology is not the end of the matter - we can reason past pathology. When we determine a value for (C*), we will need to determine the denotation$_i$ of the phrases to which is refers. Since (C) does not denote$_i$ - as the tree indicates - we determine a value for (C*) in terms of (A) and (B) only.[29]

Notice that if we evaluate (C) by the r-schema – that is, evaluate it from a context that is reflective with respect to (C) – we will find that it does denote, just like (C*). This is shown by a *secondary tree* for (C). The triple <type(C),i,r> is a secondary representation of (C), since the third element is not (C)'s context of utterance. And the secondary tree for (C) is identical to the primary tree for (C*), indicating that (C), like (C*), does denote$_r$.

Now suppose that we evaluate (C) from some context other than i or r, where it is *not* part of the common ground that (C) is pathological. Then the corresponding secondary representation of (C) is <type(C), i, n>, where n is the neutral context. The secondary tree will be just like the primary tree of (C*), with 'n' replacing 'r'. And since this secondary representation does not repeat, the tree indicates that C does denote$_n$. This neutral context is treated as if it were reflective with respect to (C). We treat (C) as a denoting phrase – a phrase denoting π+6 – if we possibly can. We cannot allow that (C) denotes$_i$, but we can allow that it denotes$_r$ and denotes$_n$. Restrictions on occurrences of 'denotes' are kept to a minimum.

The results delivered by the primary and secondary trees reflect the basic intuition behind the singularity account. Suppose you say "'The square of 1' denotes 1". Here, your use of 'denotes' is quite unproblematic. Should the pathological token (C) be excluded from its extension? The singularity account says no - because there is no need to exclude it. We have seen that (C) denotes_r π+6 because the sum of the numbers denoted_i by expressions on the board *is indeed* $\pi+6$. And for the same reason, (C) can be counted as a denoting expression in your neutral context of utterance. We have no reason to suppose that (C) must be evaluated from your context of utterance by the contradiction-producing i-schema; instead, your use of 'denotes' is treated as reflective with respect to C. This seems plausible: in general, speakers do not usually aim to produce pathological utterances, or utterances implicated in paradox.

Further, the singularity account respects a basic intuition about predicates. Intuitively, we take a predicate to pick out everything with the property that the predicate denotes. The more restrictions we place on occurrences of the semantic term *t*, the more we are at odds with this intuition. We do expect any solution to a genuine paradox to require some revision of our intuitions. But the more a solution conflicts with our intuitions, the less plausible that solution will be.[30]

An advantage of the singularity account is, I think, that it provides a unified account of the semantic paradoxes, according to which the scope of each occurrence of *t* (whether 'true', 'denotes' or 'extension') is as close to global as it can be. And the corresponding *t*-schema is as close to unrestricted as it can be. Once we have a formal theory that identifies singularities of given occurrences of *t*, we can offer minimally restricted *t*-schemas. For example, here is the the true_i-schema:

If 's' is not a singularity of 'true_i', then 's' is true_i iff s.

Similarly, we can provide close-to-unrestricted schemas for 'denotes_i' and 'extension_i'.

This is a brief sketch of the singularity account, but perhaps it is enough for present purposes.[31] Revenge paradoxes for this contextual account may seem to be generated by the sentence ‘This sentence is not true in any context’, or the phrases ‘the least ordinal not denoted in any context’ and ‘non-self-membered extension of a predicate in some context’. To fix ideas, let $\mathcal{L}$ be the language that the singularity theory is a theory of. For simplicity, $\mathcal{L}$ is taken to be English without any context-sensitive terms, plus ‘true’, or ‘denotes’ or ‘extension’. Let $\mathcal{T}$ be the language in which the singularity theory is expressed. To anticipate a little, we’ll pay close attention to the relation between the language $\mathcal{L}$ and the language $\mathcal{T}$. We’ll see that the crux of the matter is this: $\mathcal{T}$ is *not* a Tarskian metalanguage for $\mathcal{L}$.

For ease of exposition, we’ll work with just one case, the case of truth – but bear in mind that what we say about truth carries over directly to denotation and extension. If we restrict ourselves to the language $\mathcal{L}$, any occurrence of 'true' is an occurrence of a context-sensitive term. But in the language $\mathcal{T}$, we freely quantify over contexts, and we explicitly attach contextual subscripts to uses of 'true' in $\mathcal{L}$. For example, I can say that the sentence (L) does not denote$_i$ and that it does denote$_r$. And I can go on to say that (L) is true in some but not all contexts. In the language $\mathcal{T}$, then, we will find the predicate 'true-in-$\mathcal{L}$', where this is understood as the short form of 'sentence of $\mathcal{L}$ that is true$_\alpha$ for some context α’. This predicate constant may be regarded as the truth predicate for $\mathcal{L}$.

Now we can observe that no occurrence of the context-sensitive predicate 'true' of $\mathcal{L}$ is coextensive with ' true-in-$\mathcal{L}$'. We can establish this in two steps. First, every occurrence of ‘true’ will have singularities – for example, we can add to the innocent statement “’2+2=4’ is true” the paradox-producing “but this very conjunct isn’t”. Second, notice that the pathological conjunct just produced will be true in a suitably reflective context, and so true in some context. So our conjunct is a singularity of ‘true’ in our “innocent” statement, but is in the scope of 'true-in-$\mathscr{L}$'. In this respect, the extension of 'true-in-$\mathscr{L}$' will be broader than that of any occurrence of 'true'. So the question arises: isn't $\mathscr{T}$ a Tarskian metalanguage for $\mathscr{L}$?

The answer is no. The predicate 'true-in-$\mathscr{L}$' is the truth predicate for $\mathscr{L}$ in the sense that it applies to exactly the truths of $\mathscr{L}$. The scope of 'true-in-$\mathscr{L}$' is restricted to the expressions of true-in-$\mathscr{L}$. In contrast, a given occurrence of 'true' applies to all truths except its singularities. It applies to any true sentence of any language, *as long as the sentence is not identified as a singularity*. In particular, the scope of an occurrence of 'true' extends to truths expressed in the language $\mathscr{T}$, including those that cannot be expressed in the language $\mathscr{L}$ - for example, those sentences of $\mathscr{T}$ containing the predicate ‘true-in-$\mathscr{L}$’. (For an example related to revenge, consider ‘This sentence is not true-in-$\mathscr{L}$’. This is a true sentence of $\mathscr{T}$, since it is not a true sentence of $\mathscr{L}$.) So in this respect, any occurrence of 'true' is more comprehensive than the predicate 'true-in-$\mathscr{L}$', since the scope of 'true-in-$\mathscr{L}$' is limited to the sentences expressed in $\mathscr{L}$.

Further, $\mathscr{T}$ is in certain respects expressively weaker than $\mathscr{L}$. $\mathscr{T}$ is a 'scientific' language in which we describe the semantics and pragmatics of a context-sensitive term. In $\mathscr{T}$ we take context-sensitive language to be the object of our study, and stand back from the contexts and the context-sensitive term that we are describing. We formally define notions like *primary tree* and *singularity*. In scientifically describing the behaviour of the context-sensitive term 'true', we do not use context-sensitive terms. There are no context-sensitive terms in $\mathscr{T}$. When we present the singularity theory, we take up an abstract, theoretical point of view. $\mathscr{T}$ is not a language that contains a term tied to context - it is *about* a language that contains a term tied to context.

But if $\mathcal{T}$ is a classical scientific language free of context-sensitive terms, vagueness, ambiguity, and so on, it is provably subject to expressive limitations. In particular, $\mathcal{T}$ cannot contain *its* own truth predicate, 'true-in-$\mathcal{T}$'. For if it did, we could form the sentence 'This sentence is not true-in-$\mathcal{T}$' within $\mathcal{T}$, and derive a contradiction. The truth predicate for $\mathcal{T}$ must be contained in a metalanguage for $\mathcal{T}$, a language essentially richer than $\mathcal{T}$. And this metalanguage will contain truths that cannot be expressed in $\mathcal{T}$. Now none of these truths will be identified as singularities of an occurrence of the ordinary predicate 'true' of $\mathcal{L}$. So the scope of 'true' extends beyond the reach of $\mathcal{T}$. In this regard, $\mathcal{L}$ is essentially richer than $\mathcal{T}$. Clearly $\mathcal{T}$ is no Tarskian metalanguage for $\mathcal{L}$. Notice that the metalanguage for $\mathcal{T}$ is subject to the same kind of expressive limitation, and we are led to a hierarchy of languages with $\mathcal{T}$ at its base. Each metalanguage in the hierarchy contains truths that cannot be expressed at any lower level. But none of these expressions are singularities of a given occurrence of 'true', and so they are all within its scope. In this respect, $\mathcal{L}$ is essentially richer than any language in the hierarchy.

So $\mathcal{T}$ is not a metalanguage for $\mathcal{L}$. But perhaps this should not come as any surprise. Consider any semantic theory of a context-sensitive term. The language of the theory will be 'austere', free of indexical terms, 'scientific'. The language in which we give an adequate semantics for 'I' or 'now' will not itself contain the indexical 'I' or 'now'. There is no requirement that a theory of 'I' or 'now' provide context-sensitive means for referring to myself or the present time. So we should not expect an utterance of 'I am hungry' or 'The meeting starts now' to be translatable into the language of the theory. In general, we should not expect that there will be a way of translating the context-sensitive term into a term or phrase of the language of the theory. Now a Tarskian metalanguage "must contain the object language as a part", or at least it must be the case that "the object-language can be translated into the metalanguage".[32] So in general a language in which we state the theory of a context-sensitive term will not be a Tarskian metalanguage.[33] In particular, the singularity theory of 'true' - and, similarly, of ‘denotes’ and ‘extension’ - is not couched in a metalanguage for the object language $\mathfrak{L}$.

Perhaps it’s worth stressing that the singularity approach is not hierarchical: the terms ‘true’, 'denotes' and 'extension' are not stratified into a series of increasingly comprehensive predicates. Instead we have a single, context sensitive term, and each occurrence of the term has singularities that other occurrences do not have. No occurrence is more (or less) comprehensive than another; each occurrence is minimally restricted. This feature of the singularity account shouldn’t be obscured by the hierarchy generated from $\mathfrak{I}$. This hierarchy is generated from a classical scientific language, and a parallel hierarchy could be generated from, say, the language of arithmetic or chemistry, or any suitably regimented language. In each case we will obtain an infinite series of truth predicates - one series starting with the predicate 'true in $\mathfrak{I}$', another with the predicate 'denoting expression in the language of chemistry', and so on. But these series are composed of predicate constants of the form 'true in L', limited to some suitable scientific language or metalanguage L. But our interest lies in the English predicate 'true' - and according to the singularity solution, this is a context-sensitive predicate that applies to true sentences at every level of all of these hierarchies.

There are of course legitimate questions about these hierarchies. What is their extent? Can we quantify over all their levels? Must we resort to schematic generalizations? If we were offering a Tarskian account of revenge, these questions would be critical. But we are offering a different kind of contextual account, and so these questions are less urgent (though no less interesting). They are questions that do not bear directly on the singularity theory, because the theory does not stratify ‘true’, ‘denotes’ or ‘extension’. These semantic terms of the object language apply to the language of the theory, and beyond.[34]

Bibliography

Beall, J.C. (ed.) (forthcoming). *Deflationism and Paradox*, Oxford University Press.

Blackburn, Simon and Simmons, Keith (1999). *Truth*, in the series *Oxford Readings in Philosophy*, Oxford University Press, Oxford.

Burge, Tyler (1979). 'Semantical paradox', *Journal of Philosophy* 76, pp.169-198; reprinted with a postscript in R. L. Martin (ed.) 1984, pp. 83-117.

Glanzberg, Michael (2001). 'The Liar in Context', *Philosophical Studies* 103, pp.217-251.

Gödel, Kurt (1944). 'Russell's mathematical logic', in Schilpp 1944.

Grice, H.P. (1989). *Studies in the Way of Words*, Harvard University Press.

Grosz, B. and Sidner C. (1986). 'Attention, intention, and the structure of discourse', *Computational Linguistics* 12, pp.175-204.

Heim, Irene (1988). *The Semantics of Definite and Indefinite Noun Phrases*, Garland Publishing, New York and London 1988.

Hilbert, D. and Bernays, P. (1934, 1939). Grundlagen der Arithmetik, vol.1 Berlin: Springer, (vol. 2, 1939).

Kaplan, David (1997). 'What is meaning?: Explorations in the theory of meaning as use' (unpublished draft).

König, Julius (1905). 'On the foundations of set theory and the continuum problem', in van Heijenoort 1967, pp.145-149.

Kripke, Saul (1975). 'Outline of a theory of truth', *Journal of Philosophy* 72, pp.690-716; reprinted in R. L. Martin (ed.) 1984, pp.53-81.

Lewis, David (1970). 'General semantics', *Synthese* 22, pp.18-67.

Lewis, David (1979). 'Scorekeeping in a language game', *Journal of Philosophical Logic*, 8, pp.339-359; reprinted in Lewis 1983, pp.233-249.

Lewis, David (1980). "Index, context, and content", in *Philosophy and Grammar*, S. Kanger and S. Ohman (eds.), Dordrecht, Reidel; reprinted in Lewis 1998, pp.21-44.

Lewis, David (1983). *Philosophical Papers*, vol.1, Oxford University Press.

Lewis, David (1998). *Papers in Philosophical Logic*, Cambridge University Press.

Maddy, Penelope (1983). ‘Proper classes’, *Journal of Symbolic Logic* 48, pp.113-139.

Martin, D.A. ‘Sets vs. classes’, circulated xerox.

Martin, Robert L (ed.) (1984). *Recent Essays on Truth and the Liar Paradox*, Oxford University Press.

Muskens, R., van Benthem, J., and Visser, A. (1997). 'Dynamics' in *Handbook of Logic and Language*, J. van Benthem and A. ter Meulen, The MIT Press/North-Holland, pp.587-648.

Priest, Graham (2002). ‘Paraconsistent logic’, in *Handbook of Philosophical Logic* (*second edition*), vol. 6, D. Gabbay and F. Guenthner (eds.), Kluwer Academic Publishers, Dordrecht, pp.287-393.

Priest, Graham (2005). ‘Spiking the Field-artillery’ in J.C. Beall (forthcoming).

Priest, Graham (2006). *In Contradiction* (2^{nd} edition), Oxford University Press.

Richard, Jules (1905). 'Les principes des mathématiques et le problème des ensembles', in *Revue générale des sciences pures et appliquées* 16, p.541. Also in *Acta Mathematica* 30 (1906), pp.295-6. (English translation in van Heijenoort 1967, pp.143-44.)

Russell, Bertrand (1906). ‘Les paradoxes de la logique’. *Revue du metaphysique et de morale* 14, pp. 627-650.

Russell, Bertrand (1908). 'Mathematical logic as based on the theory of types', *American Journal of Mathematics*, vol. 30, pp. 222-262; reprinted in van Heijenoort, 1967 (ed), pp. 150-182.

Schilpp, P.A. (ed.) (1944). *The Philosophy of Bertrand Russell*, Open Court, pp. 123-153.

Simmons, Keith (1993). *Universality and the Liar: an essay on truth and the diagonal argument*, Cambridge University Press, Cambridge.

Simmons, Keith (1994). ‘Paradoxes of denotation’, *Philosophical Studies* 76, pp.71-104.

Simmons, Keith (2000). ‘Sets, classes and extensions: A singularity approach to Russell’s paradox’, *Philosophical Studies* 100, pp.109-149.

Stalnaker, Robert (1974). 'Pragmatic presuppositions', in *Semantics and Philosophy*, Milton K. Munitz and Peter Unger (eds.), New York: New York University Press, pp.197-213; reprinted in Stalnaker 1999, pp.47-62.

Stalnaker, Robert (1975). 'Indicative conditionals', *Philosophia* 5, 1975, pp.269-286; reprinted in Stalnaker [1999], pp.63-77

Stalnaker, Robert (1978). 'Assertion', in Cole (ed.), *Syntax and Semantics, vol.9: Pragmatics*, New York: New York University Press, pp.197-213; reprinted in Stalnaker 1999, pp.78-95.

Stalnaker, Robert (1988). 'On the representation of context', *Journal of Logic, Language, and Information*, 7, 1998, pp.3-19.

Stalnaker, Robert (1999). *Context and Content: Essays on Intentionality in Speech and in Thought*, Oxford University Press.

Tarski, Alfred (1933/1986). "The concept of truth in formalized languages", in Tarski 1986, pp.152-278.

Tarski, Alfred (1944). ‘The semantic conception of truth’, *Philosophy and Phenomenological Research* vol.4, pp.341-375; reprinted in Blackburn and Simmons 1999, pp. 115-143.

Tarski, Alfred (1986). *Logic, Semantics, Metamathematics* (second edition), edited and introduced by John Corcoran, Hackett 1986.

van Heijenoort, Jean (ed.) (1967). *From Frege to Gödel: A source book in mathematical logic*, Harvard University Press.

Endnotes

---

[1] Even more tightly self-referential is the phrase discussed in Hilbert and Bernays (1939): "the successor of the integer denoted by this phrase". These paradoxes of denotation are related to Richard's paradox (Richard 1905), König's paradox (König 1905), and Berry's paradox (reported in Russell 1906 and 1908).

[2] Kripke 1975.

[3] Richard 1905.

[4] Martin (circulated xerox), Maddy (1983).

[5] Simmons (1993) deals with truth, Simmons (1994) with denotation, and Simmons (2000) with extensions.

[6] Stalnaker 1975, in Stalnaker 1999, p.66.

[7] For a survey article, see Muskens et. al.(1997).

[8] Stalnaker 1988, in Stalnaker 1999, p.98. This is a repeated theme in Stalnaker's writings; for example:

> "... a context should be represented by a body of information that is presumed to be available to the participants in the speech situation" (Stalnaker 1999, p.6).

[9] Stalnaker takes presuppositions here to be *pragmatic presuppositions*:

> "Presuppositions, on this account, are something like the background beliefs of the speaker propositions whose truth he takes for granted, or seems to take for granted, in making his statement." (Stalnaker 1974, in Stalnaker 1999, p.48)

[10] See Grice's 'Logic and Conversation' (The William James lectures), Part 1 of Grice 1989.

[11] Appendix to Stalnaker 1975, in Stalnaker 1999, p.77. In a similar vein, Stalnaker writes:

> "... the essential effect of an assertion is to change the presuppositions of the participants in the conversation by adding the content of what is asserted to what is presupposed." (Stalnaker 1978, in Stalnaker 1999, p.86.)

[12] The analogy is with a baseball score. A baseball score for Lewis is composed of a set of 7 numbers that indicate, for a given stage of the game, how many runs each team has, which half of which innings we're in, and the number of strikes, balls and outs. Notice that correct play depends on the score - what is correct play after two strikes differs from what is correct play after three strikes. Similarly for conversations: the correctness of

utterances - their truth, or their acceptability in some other respect - depends on the *conversational* score. Lewis continues:

> "Not only aspects of acceptability of an uttered sentence may depend on score. So may other semantic properties that play a role in determining aspects of acceptability. For instance, the constituents of an uttered sentence - subsentences, names, predicates, etc - may depend on the score for their intension or extension." (Lewis 1979, in Lewis 1983, p.238)

This last remark of Lewis's is particularly relevant, since we are concerned with the subsentential terms 'true', 'denotes' and 'extension'.

[13] *op. cit.*, p.233.

[14] *op. cit.*, p.246.

[15] The notions of common ground and shared presuppositions also figure in Irene Heim's *file change semantics*, where a 'file' contains all the information that has been conveyed up to that point - and the file is continually updated as the discourse moves on. For Heim, "the common ground of a context be identified with what I have been calling the 'file' of that context" (Heim 1988, p.286). Heim's notion of common ground is more fine-grained than Stalnaker's, since Heim's account is more sensitive to the subsentential structure of sentences. As with Lewis, the extra fine-grainedness of Heim's account is of relevance to us, since the present concern is with the subsentential terms 'true', 'denotes' and 'extension'. But clearly, the accounts of Stalnaker, Lewis and Heim are broadly similar – they track context-change in terms of shifts in the shared presuppositions or common ground of the participants.

[16] The division of the revenge discourse into segments fits naturally into Grosz and Sidner's dynamic theory of discourse structure (see Grosz and Sidner 1986), but I cannot pursue the details here.

[17] According to Heim's account, I will register this shift by updating the file card that stores information about C: I will now add the entries 'is pathological' and 'does not denote a number'. Grosz and Sidner's focusing structure distinguishes the salient objects, properties and relations at each point of the discourse - and as we move from the second segment to the third, and on to the fourth, it will distinguish the pathologicality of C and its failure to denote.

[18] Cresswell, quoted in Lewis 1980, p.30. One target of Cresswell's remark is Lewis 1970, and Lewis takes Cresswell's criticism to heart in Lewis 1980. In a somewhat similar vein, Kaplan writes: "context *provides* whatever parameters are needed" (Kaplan 1989, p.591), though Kaplan's remark is restricted to expressions that are "directly referential".

[19] See Simmons 1993, especially Chapters 3 and 4.

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

[21] In a series of recent papers (see for example Field 2003), Field has offered a formally sophisticated treatment of the liar, which focuses on the revenge problem and achieves a remarkably high degree of semantic closure. But ultimately, it seems to me, it is subject to the familiar complaint: the price of consistency is expressive incompleteness. Prominent features of Field's theory are a non-classical conditional, a three-valued logic, and the notion of determinate truth. Field defines a 'determinately' operator that iterates into the transfinite, so that, it is claimed, for every liarlike sentence there is an operator that expresses its semantic status. But none of these operators captures the general notion of determinate truth, and that seems like a notion that we have the resources to express. (On this, see Priest 2005, pp.44-46, and Beall 2005, pp.23-24.) In a similar vein, Yablo distinguishes two questions:

> "Question 1: is the language able to characterize as defective every sentence that deserves to be so characterized? Question 2: are there intelligible semantic notions such that paradox is avoided only because these notions are not expressible in the language".

Yablo goes on to point out that a 'revenge-monger' may focus on Question 2, paying particular attention to the notion *having an ultimate value other than 1*, where the value 1 is to be understood as determinate truth.

[22] Priest presents dialetheism in relational terms in Priest 2002, ch. 4.6, and in Priest 2006, ch. 20.3.

[23] Perhaps the dialetheist might accept {T, F} = (F}, and also {T}={F}, but deny T = F. This would be a denial of extensionality for at least impure set theory (given that T and F are not sets). But this move seems quite *ad hoc* and unmotivated. If we accept, as the dialetheist does, that T and F are distinct values, it is hard to make sense of the thought that the collection of both values is the same as the collection of just one of them.

[24] If gaps are admitted, we'll obtain that every sentence is true or gappy – also unacceptable to the dialetheist.

[25] Tarski would not endorse this 'Tarskian' response to paradox in the setting of natural language. In Tarski 1933/1986, Tarski turns away from natural languages, and . investigates only formal, regimented languages. Hierarchical contextual accounts can be found in Burge 1979 and Glanzberg 2001.

[26] For full accounts, see Simmons 1993, 1994, 2000.

[27] In a tantalizing passage, Gödel writes: "It might even turn out that it is possible to assume every concept to be significant everywhere except for certain 'singular points' or 'limiting points', so that the paradoxes would appear as something analogous to dividing by zero. Such a system would be most satisfying in the following respect: our logical intuitions would then remain correct up to certain minor corrections, i.e. they could then

be considered to give an essentially correct, only somewhat ‘blurred’, picture of the real state of affairs” (Gödel 1944, in Schilpp 1944, p.228).

[28] A *secondary representation* of a token σ containing *t* indicates the assessment of σ by a *t*-schema other than that associated with σ's context of utterance. As we’ll have occasion to notice later, one secondary representation of (C) (<type(C), i, r>) is the prmary representation of C* - and one secondary representation of C* (<type(C), i, i>) is the primary representation of C.

[29] The primary tree for (L) is an infinite single-branched tree:

```
<type(L),i,i>
      |
<type(L),i,i>
      .
      .
      .
```

The primary representation of (L) repeats on this tree, indicating that (L) is pathological and a singularity of ‘true$_i$’. The primary tree for (L*) has <type(L),i,r> for its top node, and then every subsequent node of this single-branched tree will be the primary representation of (L). So while (L) is pathological, (L*) is not, since its primary representation does not repeat on its primary tree – and neither is (L*) a singularity of ‘true$_r$’.

The primary tree for (E) is:

```
             <type(E),i,i>
             /           \
      type(D)       <type(E),i,i>
                       /        .
                 type(D)         .
                                  .
```

This tree shows that (E) is pathological and a singularity of ‘extension$_i$’. It’s easy to see that the primary representation of (E*) does not repeat on its primary tree, and (E*) is not a singularity of ‘extension$_r$’.

[30]   For example, the Tarskian stratification involves massive restrictions on occurrences of *t*. On a standard Tarskian line, the referring expression 'the only even prime', for example, is of level 0; your unproblematic denoting phrase ("the number denoted by 'the only even prime'") is of level 1, and so on, through the levels. Your use of 'denotes' in an utterance of level 1 has in its extension all referring expressions of level 0, *and no others*. So all sentences of level 1 and beyond are excluded from the extension of such a use of 'denotes'. Gödel remarks of Russell's type theory that "...each concept is significant only ... for an infinitely small portion of all objects" (Gödel 1944, p. 149). A similar

complaint can be made about a standard Tarskian account of *t*: for example, an ordinary use of ‘denotes’ will apply to only a fraction of all the denoting expressions.

[31] Detailed accounts of the singularity theory can be found in Simmons 1993, 1994, and 2000.

[32] Tarski 1944, in Blackburn and Simmons 1999, p.126.

[33] This point in stressed by Kaplan, in Kaplan 1997, p.7.

[34] My thanks to Jamin Asay and Thomas Hofweber for their helpful comments on an earlier draft.