

DERRIDA AND SELF-REFERENCE

Graham Priest

Philosophers are wont to produce theories of very general import concerning, for example, truth, meaning, knowledge and so on. Because of their general import it is not uncommon for these theories to apply to themselves. It is also not uncommon for the result to be a contradiction, usually unintended. Thus, for example, Kant (un)happily says that things in themselves cannot be talked about, thus talking about them; or the Wittgenstein of the *Tractatus* explains certain facts about the structure of language, the world and the relationship between them, the upshot of which is that such facts cannot be expressed.

There are obviously general issues here which are worthy of consideration; however, I do not intend to discuss them now.¹ My aim is to (at least start to) consider how a more recent philosopher fits into the same pattern: Derrida. To see roughly why he fits, it is useful, especially for those more familiar with analytic philosophy, to compare him with (the earlier) Wittgenstein. At *Tractatus* 3.23 Wittgenstein says 'The requirement that simple signs be possible is the requirement that sense be determinate'. What he means is that if there were not some words that have their meaning in a direct way, unmediated by other words, then assertions would have no determinate sense. Wittgenstein thinks that sense must be determinate; hence he postulates the existence of simple objects to be the meanings of certain signs. Derrida, on the other hand, denies the existence of any such things, and so infers that sense is not determinate. Derrida's views apply to language in general, and therefore must apply to his own writings. And, *prima facie*, contradictions seem to follow. For example, in making the claim that language has no determinate sense, Derrida seems to be making a claim with determinate sense. We will come back to this in due course. Let us start with a more careful look at Derrida's views.²

I. Presence

Derrida is a literary philosopher. This is true in two senses. The first sense is that in which his work falls across the traditional divide between philosophy and literary criticism. In what follows I shall not attempt to consider all his work, but only that philosophical part which bears on the theme of the determinacy of sense (though, arguably, this is *the* central part of his work).³ The second sense in which Derrida is a literary philosopher is that, for whatever reason, he often eschews traditional philosophical styles of writing in favour of styles that might be more at home in lit-

¹ For a start, see [6], [10, ch. 4], and [11].

² What follows is a simple exposition of Derrida's views. I make no attempt to evaluate them.

³ For a general introduction to Derrida's thought, a number of works can be consulted, e.g., [9]. The ground we will cover is well traversed in [1].

erary works. This makes him singularly difficult to interpret. For this reason I am unsure that I have understood him. But then — if Derrida is right — there is perhaps nothing determinate to understand.

Still, putting this *caveat* aside, let us start with Derrida's principal philosophical thesis: the denial of presence. In a nutshell, a presence is some kind of non-linguistic entity which serves to provide a determiner of sense; Derrida often calls it the *transcendental signified*. Clearly, Tractarian objects are presences; so are Lockean ideas; so are the entities of Fregean semantics (senses, concepts, etc). Derrida does not explicitly set his sights on these; the examples he does cite are things such as essence, Being and, particularly, consciousness. These are, of course, some of the central concepts from Western metaphysics. In fact, Derrida takes metaphysics to be exactly that subject that endorses the existence and action of some notion of presence. As he puts it [4, p.20]:

... the entire history of metaphysics ... impose(s), and will never cease to impose upon semiological science in its entirety this fundamental quest for a transcendental signified' ...

Whether Derrida is right to see the whole of Western metaphysics in the light of the philosophy of language is an interesting question which we need not follow up here.

II. ... and Absence

So much for what presence is; next to its denial. When Derrida denies presence he denies the existence of any entities which ground meaning, because either the entities themselves do not exist or, if they do, they can not function in the required way. The next question is obvious: if things do not have meaning in virtue of signifying transcendental signifieds, how do they have meaning? Derrida gives an answer to this question by way of a critique of structuralist theories of meaning, and particularly that of Saussure.⁴

Structuralism already rejects the view that a phrase means in virtue of a concept it expresses in and of itself. It observes that phrases come in families, often binary families, and claims that a phrase has its meaning in virtue of its relationships to the other phrases in that family, and, particularly, because of its contrasting place in the family. Thus, for example, consider the word 'red'. This belongs to the family of colour words {red, green, blue ... }, and what makes it mean red, is just that it does not mean green, blue, etc. Similarly for 'blue'. Or, to remove the air of circularity from this, we could put it as follows: each of the words in the family has meaning in virtue of its network of relations of opposition with the other members of the family, that is, its difference from them.

This much of structuralism Derrida accepts, but other aspects of Saussure's thought he rejects. Crucially, he rejects Saussure's notion of the sign. According to this, the sign is composed of two aspects: the signifier (that is, the physical token)

⁴ This can be found in the early pages of [2], and is usefully summarised in [4, pp.19 ff].

and the signified (the concept it expresses). Though the sign may be dependent for its identity on its place a network of relations, the signifier is still conceptualised as corresponding to a unique concept, the signified; and this is a form of presence.

But how are we to locate this concept? Not, according to Saussure, in writing. For when a word or longer linguistic string — which we can call a text — is written, it continues to exist independently of its uttering, and can come to take on all kinds of different meanings in virtue of new contexts. (Derrida calls this the *iterability* of the text.) For example, Shelley's poem *Ozymandias* tells of a traveller who comes across the trunk of a statue in the middle of a desert. This is all that remains of a once great empire. On it are written the words:

My name is Ozymandias, King of Kings:
Look on my works, ye Mighty, and despair!

which obviously meant something quite different in their original context.

To find the concept we therefore have to look to the intentions, i.e., consciousness, of the utterer. (Consciousness therefore becomes a kind of presence.) But what would we find if we examined the intentions of the utterer? They would, presumably, be something like: in uttering such and such I intended to mean so and so. We do not, therefore, find anything to ground meaning; just more words (so and so). And if we were to ask for the meaning of *these* words we would merely find *more* words. Hence, we never break out of the circle of words into a realm of ultimate meanings. Any word is referred to further words indefinitely, and this relationship of referral is just as constitutive of meaning as Saussure's differences. As Derrida puts it [4, p.26]:

Whether in the order of spoken or written discourse, no element can function as a sign without referring to another element which itself is not simply present. This interweaving results in each 'element' — phoneme or grapheme — being constituted on the basis of the trace within it of the other elements of the chain or system.

Thus, meaning is constituted by the total network of both differences and deferrals. For this network Derrida coins the neologism '*différance*'. Hence, meaning is constituted by, as he puts it, the 'play' of *différance*.

III. Deconstruction

Before we are finished with Derrida's basic notions, there is another we need to discuss: deconstruction. Derrida, as I have said, denies presence. He does not, as far as I am aware, provide a general argument against presence. Rather, what he does is take examples of philosophical or metaphysical texts which advocate presence, and show that they are self-undercutting in some sense. This is called 'deconstruction'. Though the details of how to deconstruct depend on the text itself, and so vary from case to case, there is a general frame.

A metaphysical text advocates some form of presence, *p*. By the Saussurian semiotics, this draws its sense from its opposite, *non-p*. And this means that the pair $\langle p, non-p \rangle$, in some sense, provide an under-pinning structure for the text. (There may, of course, be other pairs providing other aspects of the structure.) But the members of the pair are not on equal terms. Since it is *p* that is supposed to be doing all the metaphysical work, *p* is the dominant member of the pair. In the jargon of deconstruction, it is privileged.

Deconstruction comes in two stages (or perhaps 'aspects' is a better word, since these may be carried out simultaneously to a certain extent). The first is to reverse the privileging, to make *non-p* the dominant member of the pair. This is done by showing how, in contrast to the its explicit claims, the text implicitly shows that *non-p* is the more important and fundamental of the pair. This stage is called, for obvious reasons, 'reversal'. The next stage, now that both members of the pair are on an equal footing, is to examine the very ground of the distinction between *p* and *non-p*, and show that it is a false antithesis. There is a space between them in which some new concept lies, which both unites and differentiates the pair, but is not reducible to either. This stage is called 'displacement'. In many respects it is reminiscent of the third stage of Hegel's dialectic.⁵ The new concepts in question might be said to transcend (*aufhebt*) the opposition; they are often called 'undecidables' by (a rather tenuous) analogy with Gödel's Theorem. The concepts are expressed by words that occur in the original text, but come to take on a whole new meaning. These provide much of the distinctive vocabulary of deconstruction: *supplément*, *hymen*, *parergon*, etc.

How deconstruction is supposed to undercut the text is as follows. First, the reversal shows that there are certain contradictions implicit in the pretensions of the text. Showing that a text is self-referentially inconsistent in this way is, of course, a very traditional form of critique. The result of the second phase is much more novel. The undecidable concept that emerges in this undercuts the distinction between, and hence the sense of, the original contrasting pair of concepts, and *a fortiori* any attempt to use the contrast to ground an account of meaning (presence).

It would be pointless for present purposes to examine all Derrida's applications of this frame. However, let us quickly look at one example of it, if only to put some flesh on these rather abstract bones. This is Derrida's discussion of Rousseau in *De la Grammatologie*.⁶ Like Saussure, Rousseau takes speech to be a form of presence in a way that writing is not (since it is always present to consciousness). The couple speech/writing is therefore the underlying structure of his text, speech being the privileged member. Derrida argues that Rousseau is forced to recognise, despite his avowed thesis, that writing is an absolutely necessary supplement of speech; thus writing is conceded to be of more importance. The undecidable that emerges from the discussion is the notion of a supplement, which describes the relationship between speech and writing but, in good Hegelean fashion, means both 'replacement' and 'completion', and so destroys any clear opposition between speech and writing.

⁵ Though see [4, p.43].

⁶ A summary is given in [7, pp.126-127]; another can be found in [8].

IV. Cratylus' Problem

After this whistle-stop tour of Derrida's views, let us now turn to their self-reference. Derrida has given us an account of language, of how it means and how it doesn't. But Derrida is himself using language to express his views, so we can apply his views to his own writings. When we do this certain *prima facie* problems arise. I will discuss two. The first, and shorter, is a very general one.

If Derrida is right, there is no presence, no ultimate ground of meaning. In particular, no text can show immediately and transparently what it means. Thus, all texts must be interpreted — or read, to use a different jargon. Moreover, because of its iterability, a text can be read in many ways; that is, it may be interpreted to mean different things in different contexts (which is not, of course, to say that it can be interpreted to mean anything). There is no question of any one interpretation being correct (though this is not to say that all interpretations are equally good). What is more, the interpretation of a text, by increasing the play of *différance* around it, is actually liable to *change* its meaning. Finally, and in any case, the interpretation of a text is just another text, and so needs to be interpreted, but itself has no privileged interpretation.

Thus, a text expresses no intrinsic meaning, but may be taken to mean indefinitely many things. Now apply this observation to Derrida's own text. We take Derrida to be advocating a certain view, namely, arguing against the existence of presence, the determinacy of sense. Yet, if he is right he is not advocating anything with stable and determinate sense at all. What, then, are we supposed to make of what he says if there is nothing *as such* that he says? Or, to put it the other way, given that he does express certain determinate views (those that I have summarised), he is expressing something that, if he is right, cannot be expressed. This problem has been noted by a number of writers. Wood [13, p.225], for example, pithily puts it thus: 'Derrida has the problem of saying what he means without meaning what he says'.

The problem is similar to that in which Cratylus found himself in Plato's dialogue of the same name. Cratylus was an extreme Heraclitean who believed that everything was in a state of constant flux; and this everything included meanings. Now if Cratylus were right, this flux of meaning would deprive his own assertions of any determinate sense. In particular, they could not mean what he desired them to. Traditionally, Cratylus took the point, and in the end would say nothing but would merely wag his finger.

The situation for Derrida is not exactly the same as that for Cratylus, however. In particular, though Derrida's utterances might mean nothing in any absolute sense, yet the idea of what they are *interpreted* to mean makes perfectly good sense. And this allows the following reply: Derrida's texts have been interpreted (by himself and others) as an attack on the notion of presence (or as subverting the notion, to use another jargon). If this is so, they have served their function. If the texts are interpreted as having some other meaning by some latter-day Shelleyan traveller, who cares?

V. *Différance*

The second problem of self-reference concerns the specific notion of *différance* itself. Many commentators have observed that this is a problematic notion. The problem I have in mind hinges on the following fact. *Différance*, according to Derrida, is the structure that gives rise to meaning; it is the precondition of any meaning at all. It itself is therefore beyond expression: it cannot be described in any way. As Derrida himself puts it [5, p.26]:

... *différance* has no name in our language. But we 'already know' that if it is unnameable, it is not provisionally so, not because our language has not yet found or received this *name*, or because we would have to seek it in another language, outside the finite system of our own. It is rather because there is no *name* for it at all, not even the name of essence or of Being, not even that of '*différance*', which is not a name, which is not a pure nominal unity, and unceasingly dislocates itself in a chain of differing and deferring substitutions.

Let us call this view the 'inexpressibility of *différance*'. It is not difficult to see the problem to which it gives rise. We will come to this in a moment. But in virtue of the fact that it is problematic, it is worth pausing to see that this claim is no mere mistake or mental aberration on Derrida's part; it is, in fact, forced on him by the internal logic of his position. One can see this by putting together two observations.

For the first, note that deconstruction is an operation that is applied to a text, or, more generally, collection of linguistic entities (sentences, pictures, etc.) structured by some binary opposition. In the phase of displacement it produces a notion that is not expressible in terms of that opposition, and so in the text that this opposition structures. As Derrida himself puts it [4, p.42]:

That being said — and on the other hand — to remain in this phase [of deconstruction, viz., reversal] is still to operate on the terrain of and from within the deconstructed system. By means of this double, and precisely stratified, dislodged and dislodging, writing, we must also mark the interval between inversion, which brings low what was high, and the irruptive emergence of a new 'concept', a concept that can no longer be, and never could be, included in the previous regime.

For the second observation, note that the totality of all linguistic entities, that is, textuality itself, is structured by a binary opposition. What *all* such entities have in common is their presupposition of (the) metaphysics (of presence); crucially, the sign itself, of which all linguistic entities are composed, is the metaphysical notion *par excellence*. As Derrida puts it [3, pp.280-281]:

There is no sense in doing without the concept of metaphysics in order to shake metaphysics. We have no language — no syntax and no lexicon —

which is foreign to this history [of metaphysics]; we can pronounce not a single destructive proposition which has not already had to slip into the form, the logic, and the implicit postulations of precisely what it seeks to contest. To take one example from many: the metaphysics of presence is shaken with the notion of *sign*. But . . . as soon as one seeks to demonstrate in this way that there is no transcendental or privileged signified and that the domain or play of signification henceforth has no limit, one must reject even the concept and word 'sign' itself — which is precisely what cannot be done. For the signification 'sign' has always been understood and determined, in its meaning, as a sign-of, a signifier referring to a signified, a signifier different from its signified.

The very notion of presence, the core of metaphysics, is therefore written into any linguistic entity, and hence it is the pair presence/absence that structures textuality itself.

Since this is so we can apply the deconstructive operation to it. What, then, do we get? What notion is obtained when the pair presence/absence is deconstructed? The answer is, of course, *différance*. *Différance* is the notion that arises out of the displacement of the pair presence/absence, and so transcends them. This is because it is precisely *différance* that [2, p.143]:

. . . makes the opposition of presence and absence possible. Without the possibility of *différance*, the desire of presence as such would not find its breathing space. This means by the same token that this desire carries in itself the destiny of non-satisfaction. *Différance* produces what it forbids, makes possible the very thing it makes impossible.

Thus, [4, p.27]:

. . . *différance*, then, is a structure and a movement no longer conceivable on the basis of the opposition presence/absence.

To summarise the two observations: the first is that if a text is deconstructed we arrive at a notion, facts about which, cannot be expressed in the text in question; the second is that the notion of *différance* is obtained by deconstructing the totality of *all* linguistic entities. It follows from these that facts about *différance* can not be expressed by linguistic utterances at all. Hence, we see that the inexpressibility of *différance* is no mere quirk of Derrida, but is inherent in the logic of his position.

VI. Derrida and Wittgenstein

This, though, poses an obvious problem when we apply it to Derrida's own works. These are replete with discussions of *différance* and its doings. How, then, can Derrida say that *différance* is inexpressible? Even to say this you have to refer to it to say what it is that is inexpressible, and so express it.

What does Derrida make of the situation? He never comments on it explicitly as far as I am aware, but there is a remark at the end of the interview with Ronse [4, p.14] which would appear to apply to it. It is as follows:

I try to write (in) the space in which is posed the question of speech and meaning. I try to write the question: (what is) meaning to say? Therefore it is necessary in such a space, and guided by such a question, that writing literally mean nothing [*l'écriture à la lettre ne-veuille-rien-dire*]. Not that it is absurd in the way that absurdity has always been in solidarity with metaphysical meaning. It simply tempts itself, tenders itself, attempts to keep itself at the point of the exhaustion of meaning.

What Derrida is saying, for all his hedging, is that his remarks about meaning, and in particular, presumably, about *différance*, are meaningless. This move may be thought absurd or foolish; it is at least heroic. But before one judges it, one should recall the following. I started off making a comparison between Derrida and the Wittgenstein of the *Tractatus*. As is well known, Wittgenstein, in his discussion of meaning, ended in exactly the same predicament as Derrida: his own ideas turned out to be inexpressible by his own lights. What is more, his reaction to it was identical to that of Derrida: the heroic stance. In the famous penultimate proposition of the *Tractatus* Wittgenstein admits that his own words are meaningless.

VII. The Paradoxes of Self-Reference

What is one to make of the situation? First, the parallel between Wittgenstein and Derrida here is striking. It clearly deserves greater consideration. But this lies outside the scope of this paper. Second, we must consider the outcome for Derrida's views. Some may take the affair to be a simple refutation of his views. Interestingly enough, Wittgenstein's similar predicament has rarely been interpreted in this way. Another possibility is that we try to read profound significance into the situation; just as some have tried to read profound or even mystical significance into the penultimate proposition of the *Tractatus*. I cannot agree with this line of thought. To admit that one's own view is nonsense is as damning a self-indictment as one can find — particularly when it is clear that the views, however wrong they may be, are intelligible, and so not nonsense.

There is, however, a third possibility. There is another area of philosophy where certain things appear to turn out to be both expressible and inexpressible: the paradoxes of self-reference. For example, in Berry's Paradox a certain number turns out to be both definable and undefinable, and similarly with the other definability paradoxes. This suggests that Derrida's contradiction may be of a piece with the paradoxes of self-reference. In fact, the similarity between the two situations is closer than a mere surface analogy. In section V we saw that the contradiction is generated by the fact that there is some operation (deconstruction) that, when applied to a totality of some kind, produces a novel entity of that kind; when applied to the totality of all entities of that kind it must therefore produce something that is both inside

and outside that totality. Though I shall not demonstrate this here, this is precisely the structure that lies behind all the logical paradoxes. To those familiar with Russell's Vicious Circle Principle, the general similarities will be obvious enough.⁷

This third possibility does not, on its own, tell us what to conclude about the situation. Rather, it opens up a number of options: all those that have been suggested for handling the paradoxes of self-reference: metalinguistic ascent, truth value gaps, dialetheism, etc. Which, if any, of these might be appropriate in this context is an issue that transcends the bounds of anything attempted in this essay.⁸

University of Queensland

Received September 1992

Revised June 1993

REFERENCES

1. J. Culler, 'Jacques Derrida' in J. Sturrock (ed.), *Structuralism and Since: from Lévi-Strauss to Derrida* (Oxford: Oxford University Press, 1979) pp.154-179.
2. J. Derrida, *Of Grammatology*, trans. G.C. Spivak (Baltimore: John Hopkins University Press, 1974).
3. J. Derrida, 'Structure, Sign and Play in the Discourse of the Human Sciences' in *Writing and Difference* (London: Routledge, 1978) ch. 10.
4. J. Derrida, *Positions*, trans. A. Bass (Chicago: University of Chicago Press, 1981).
5. J. Derrida, 'Différance' in *Margins of Philosophy* (London: Harvester, 1982) ch. 1.
6. F.B. Fitch, 'Self-Reference in Philosophy', *Mind* 55 (1964) pp.64-73; a revised version appears as Appendix C of *Symbolic Logic* (New York: Ronald Press Company, 1952).
7. K. Hart, *The Trespass of the Sign* (Cambridge: Cambridge University Press, 1989).
8. B. Johnson, translator's preface to J. Derrida's *Dissemination* (Chicago: University of Chicago Press, 1981).
9. C. Norris, *Derrida* (London: Fontana Press, 1987).
10. J. Passmore, *Philosophical Reasoning* (London: Duckworth, 1961).
11. G. Priest, 'The Limits of Thought — and Beyond', *Mind* 100 (1991) pp.361-370.
12. G. Priest, 'The Structure of Paradoxes of Self-Reference', *Mind* (forthcoming); abstracted under the title 'All the paradoxes of self-reference have a common underlying structure generating the contradiction', *The Journal of Symbolic Logic* 57 (1992) p.1501.
13. D. Wood, 'Derrida and the Paradoxes of Reflection', *Journal of the British Society for Phenomenology* 11 (1980) pp.225-238.

⁷ Some details can be found in [11]. For a full discussion see [12].

⁸ This paper was read at the Australasian Association of Philosophy Annual Conference, Melbourne, 1991. I had not intended to publish it, but a number of people said that they found the exegesis of Derrida in the paper helpful, so I decided that there may be some point. I have made some changes as a result of helpful comments by anonymous referees; I am sure, however, that these are unlikely to have produced consensus.