Derrida and Formal Logic: Formalising the Undecidable

Paul Livingston

Abstract

Derrida's key concepts or pseudo-concepts of *différance*, the trace, and the undecidable suggest analogies to some of the most significant results of formal, symbolic logic and metalogic. As early as 1970, Derrida himself pointed out an analogy between his use of 'undecidable' and Gödel's incompleteness theorems, which demonstrate the existence, in any sufficiently complex and consistent system, of propositions which cannot be proven or disproven (i.e., decided) within that system itself. More recently, Graham Priest has interpreted *différance* as an instance of the general metalogical procedure of diagonalisation.

In this essay, I consider the extent to which Derrida's key terms and the essential operations of deconstruction can be formalised. I argue that, if formalisation is indeed the technique of writing *par excellence*, then the formalisation of deconstructive concepts tends to show how the auto-deconstruction of total systems arises from the problematic possibility of writing itself. For instance, since diagonalisation permits the 'arithmetisation of syntax' whereby a formal system is able to formulate claims about its own logico-grammatical properties, we can understand its potential to inscribe the undecidable within the systematicity of language as simply one instance of the potential of writing, in figuring itself, to render inscrutable the trace of its own origin.

*

The aim of this paper is to consider the extent to which some of the key operations of Derrida's deconstruction can be understood as constituting a reflection on *formalism* as such and, therefore, as parallel to key metalogical results arising from reflection on the structure and limits of formal languages. Without excluding other ways of understanding the methods and significance of deconstruction, I argue that several of Derrida's key terms (for instance *trace,* the 'undecidable,' and *différance*) and the textual *praxis* they embody, can indeed usefully be understood as figuring the metalogical consequences of formalism as such. This implication offers to clarify the underlying logical structure of some of Derrida's most important texts, suggests a greater degree of convergence than has usually been noted between these texts and the 'analytic' tradition of meta-linguistic and metalogical reflection, and also facilitates a clearer understanding of the textual *praxis* of deconstruction itself.

I

As early as 1970, Derrida suggested an analogy between what he calls the 'undecidable' and the incompleteness result discovered by Gödel and first announced in the article 'On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems' published in 1931. Derrida draws this connection in the course of a discussion in 'The Double Session' in which he juxtaposes an excerpt from Mallarmé's text *Mimique* with a passage from Plato's *Philebus*. The issue raised by both texts is that of *mimesis*, and of the relationship between a representational text, image, or inscription and the 'original' that it represents. Mallarmé's text, Derrida argues, makes possible a thinking of *mimesis* whereby it is no longer understandable as the hierarchical relationship between a representation and a (present or deferred) original. Rather, Mallarmé's text gives us to think a 'play' of *mimesis* with no original, an order of mirroring defined by *allusion* rather than the hierarchical logic of truth and illusion:

> As its name indicates, allusion *plays*. But that this play should in the last instance be independent of truth does not mean that it is false, an error, appearance, or illusion. Mallarmé writes 'allusion,' not 'illusion.' Allusion, or 'suggestion' as Mallarmé says elsewhere, is indeed that operation we are here *by analogy* calling undecidable. An undecidable proposition, as Gödel demonstrated in 1931, is a proposition which, given a system of axioms governing a multiplicity, is neither an analytical nor deductive consequence of those axioms, nor in contradiction with them, neither true nor false with respect to those axioms. *Tertium datur*, without synthesis. (Derrida 1970, 219)

Since 'undecidable' and 'undecidability' are terms that Derrida retains throughout his career, indeed putting them to a central use in his later analyses of such phenomena as hospitality and the gift, it is worth pausing over this analogy and asking what it shows us about the status of deconstruction *vis a vis* formalism and formalisation, of which Gödel's result is a modern masterpiece.

As is familiar, the essence of Gödel's proof is to construct a sentence in the language of Russell and Whitehead's *Principia Mathematica* that is *undecidable* in the sense that, by way of the construction of a predicate that formalises the rules of proof in that system, it 'asserts' of 'itself' that it cannot be proven or disproven within the system.[1] That is, the Gödel sentence GS for a particular system is defined in terms of a predicate that is intended to hold of all and only sentences in that system that are provable within it; in terms of this 'proof' predicate, the sentence GS 'says' of itself that it cannot be proven. Because of this, neither GS nor its negation can be assumed to be provable, on pain of contradiction. Such a sentence is thus 'undecidable' in the sense that the axioms and constitutive rules of the system do not *decide* it: that is, they themselves do not and cannot establish whether it is true or false. Gödel goes on to take this sentence as demonstrating the *incompleteness* of *Principia Mathematica* in the sense that (assuming *PM* is consistent) there is a truth that it cannot prove. For it is apparently possible to 'see' (by means, however, of an essentially *informal* argument) that the proposition asserted by the Gödel sentence is true (i.e. that, as it says, it itself is indeed not provable within *PM*) although *this* cannot be proven within the system.[2]

In introducing the analogy to Gödel's result, Derrida is quick to emphasise that the undecidability that concerns him is *not* a matter of semantic ambiguity or polysemy:

> 'Undecidability' is not caused here by some enigmatic equivocality, some inexhaustible ambivalence of a word in a 'natural' language, and still less by some '*Gegensinn der Urworte*' (Abel). In dealing here with hymen, it is not a matter of what Hegel undertook to do with German words like *Aufhebung, Urteil, Meinen, Beispiel*, etc., marveling over that lucky accident that installs a natural language within the element of speculative dialectics. What counts here is not the lexical richness, the semantic infiniteness of a word or concept, its depth or breadth, the sedimentation that has produced inside it two contradictory layers of signification (continuity and discontinuity, inside and outside, identity and difference, etc.). What counts here is the formal or syntactical *praxis* that composes and decomposes it. (Derrida 1970, 220)

That is, the undecidable, in the sense in which it concerns him, is not a matter of single terms having a multiplicity of non-equivalent or even mutually contradictory 'meanings.' Quite to the contrary, the undecidability that Derrida finds in Mallarmé's text is a consequence of a total structural relationship that is, in *this* case, figured in the undecidability (for instance between 'inside and outside,' 'continuity and discontinuity,') of the single term 'hymen.' However, since the undecidability here is that of *mimesis* itself, there is nothing essential about the term; others would have done just as well:

> This word, this syllepsis, is not indispensible; philology and etymology interest us only secondarily, and the loss of the 'hymen' would not be irreparable for *Mimique*. It produces its effect first and foremost through the syntax. If we replaced 'hymen' by 'marriage' or 'crime,' 'identity' or 'difference,' etc., the effect would be the same, the only loss being a certain economic condensation or accumulation, which has not gone unnoticed. It is the 'between,' whether it names fusion or separation, that thus carries all the force of the operation. (Derrida 1970, 220)

Thus, the point of emphasising the ambiguous and even contradictory meanings of 'hymen' in Mallarmé's text is not to evince anything intrinsic to this word itself, but rather to show the way in which it (contingently and non-essentially) occupies a particular *position* in this text – the position, as we may say, of the undecidable, what the text itself, and the logic that governs it, does not give us – for *structural* reasons – the resources to decide (Derrida 1970, 220).

Derrida goes on to emphasise, moreover, that what holds for 'hymen' in Mallarmé's text holds, as well, for the (other) key terms of deconstruction, as he had already employed them in readings of Husserl, Plato, Saussure, Rousseau, Heidegger, and others:

> What holds for 'hymen' also holds, *mutatis mutandis*, for all other signs which, like *pharmakon*, *supplément*, *différance*, and others, have a double, contradictory, undecidable value that always derives from their syntax, whether the latter is in a sense 'internal,' articulating and combining under the same yoke, *huph'hen*, two incompatible meanings, or 'external,' dependent on the code in which the word is made to function. But the syntactical composition and decomposition of a sign renders this alternative between internal and external inoperative ... Is it by chance that all these play effects, these 'words' that escape philosophical mastery, should have, in widely differing historical contexts, a very singular relation to writing? (Derrida 1970, 221)

Since the 'representative' sign lives on the distinction between the present and the non-present, or the distinction between the 'original' and representation, any term for the condition of the possibility of such a distinction will exhibit the same kind of undecidability as Mallarmé's 'hymen.'

One such term, employed in relation particularly to Saussure's understanding of language as a 'system of differences without positive terms,' but bearing more general application as well, is the neologism '*différance*.' The term, as Derrida explains in the article of the same title, problematically expresses a kind of general condition for the possibility of presentation itself, given that linguistic presentation is possible only within a system of signs defined both by synchronic difference and diachronic deferring (Derrida 1968, 6). But the consequence of this is that *différance* itself cannot be presented or named, cannot be inscribed or marked by any positively meaningful representative sign. It is thus that '*différance* has no name in our language' or in any other language (Derrida 1968, 26), and that accordingly '*différance* is neither a word or a concept' (Derrida 1968, 7). Rather, in its very incapacity to be named, it 'exceeds the order of truth at a certain precise point,' reserving itself or removing itself 'in regular fashion' from the systematic structural distinctions of truth or falsity or of presence and absence which it itself structures (Derrida 1968, 8).

Given this, it is possible to specify the analogy between Gödel's result and deconstructive undecidability on at least two significant points. First, both trade decisively on the capacity of a *total* system of signs, directed to the establishment of truth or the maintenance of presence, to represent its *own* constitutive conditions of possibility, to figure the basis of the central distinctions that organise the system itself. It is, in both cases, through the figuring of these conditions of possibility at a particular point – the Gödel sentence or Derrida's 'undecidable' terms – that these conditions are shown to undermine themselves, to be at the same time conditions of the *im*possibility of figuring the underlying logic of the system completely. In this way, in each case, the system achieves a kind of total *self-reference*, an 'encoding' at a single point of the total logic that governs the entirety of the system, as well as that point itself. This amounts to a demonstration of the essential *incompleteness* of the system in question, of the necessary existence of points or sentences (in particular, those that express the system's own conditions of possibility) that cannot be decided (as true or false, or as present or absent) in terms of it. In the case of the Gödel sentence itself, these conditions of possibility are represented in the 'provability predicate' for a particular system which encodes the systematic, rule-governed possibilities of proof

within a particular system. In a similar fashion, Mallarmé's 'hymen' inscribes, within his text, the condition for the possibility of a structural distinction between original and representation in *mimesis*, and its inscription marks the point of the self-undermining of this distinction.

Second, in both cases the result can be *generalised*. Just as the existence of a Gödel sentence for *Principia Mathematica* points to the more general phenomenon of the incompleteness of *any* formal system of sufficient complexity, Derrida's *différance* encodes the *general* possibility of undecidability for any system of signs regularly governed by the opposition presence/absence. Thus, the phenomenon of undecidability demonstrated in the particular case of Mallarmé's text by the term 'hymen' is in no way limited to that particular text or term, but in fact can be extended to any mimetic system of signs whatsoever, wherever the general regular conditions of the possibility of representation are themselves represented. The more general terms *différance*, *trace*, and *supplément*, capture this, and are explicitly meant to intervene to destabilise the *totality* of the metaphysics of representation, or what Derrida elsewhere calls the 'metaphysics of presence' itself. They function wherever there is a field or structure of signs which presuppose, for their significative or truth-producing work, a distinction between presence and representation, or between truth and falsity – which is to say wherever there is a system of signs at all.

Thus, much as Gödel's result does not simply evince a contingent limitation of a particular system such as *Principia Mathematica* but rather marks a fundamental problem for the possibility of formalisation *as such* (one which may cause us to reconsider basic and otherwise plausible ideas about the nature of mathematical truth), Derrida's invocation of undecidability suggests a fundamental reconsideration of what is involved in *any* possible system of representation, and hence in any *inscription* or *writing* of any sign as such.[3] The possibility of both operations – the generation of the undecidable Gödel sentence, or the internal/external tracing of the undecidable in Derrida – is based on the way the general structural laws responsible for a system's capacity for presentation – the system's constitutive and determinative logic – can themselves be brought to problematic presentation within that system itself, and so can be seen to undermine themselves at the very point of their positive statement.

II

As we have seen, then, the undecidability long asserted by Derrida as an essential component of deconstruction, and that shown by Gödel's

own meta-mathematical argument, are structurally similar in at least two ways. First, both depend on a kind of 'self-referential' encoding whereby a system's total logic (the conditions for the possibility of its organising distinctions) is formalised at a single point – the Gödel sentence or the 'undecidable term' – which in turn makes it possible to inscribe an 'undecidable.' Second, both suggest a *generalisation* of this result to show that any system of sufficient complexity will allow the inscription of undecidables, and hence be 'incomplete' in a specific sense.

There is, moreover, a third, highly important point of analogy that, although it has been missed by most commentators, verifies this close connection and provides an essential clarification of the basis for *any* deconstructive strategy of reading. It is this: what Derrida calls the 'undecidable' always results from a semantical effect of *syntax* that cannot itself be excluded from any regular system of writing. Although this effect involves, as I shall argue, a kind of essential crossing or confusion between the internal, rule-governed structure of a system and its external 'meaning' or semantics, it essentially *cannot* be captured by any analysis that works on the level of semantic meaning alone.

This essential crossing of syntax and semantics suggests an important analogy with the general metalogical procedure of 'diagonalisation,' which underlies Gödel's result, as well as several other key results of twentieth-century formal and metalogical reflection.[4] The best way to understand diagonalisation, in general, is to think of an infinite number of elements of a system, the totality of which exhausts the system as a whole, or comprises the totality of elements with a certain property within it. For instance, we might take the (infinite) totality of sentences that are provable within a given formal system, or the (infinite) totality of sentences that are assertable as true within the logic of a particular text. Diagonalisation, then, operates on this totality as a whole to produce another element which is both: *formally* a member of the totality in question (that is, it bears the right formal properties to be a member of the totality) and, at the same time, *demonstrably* not the same as *any* of the (infinite number of) elements that already comprise the totality.[5]

Thus, diagonalisation allows the construction, given an infinite set, of an element which is formally part of that set but, as can be shown, differs from each of its (infinite number of) members, and so is not after all part of it. For instance, the Gödel sentence GS for a particular system *diagonalises* the set of all decidable sentences of the system, in the following sense: given the arithmetical specification of the rules that decide provability (or provability of the negation) of any sentence, it can be shown that GS is not a member of this set; the sentence is itself generated by means of reasoning about what must escape these rules.

The existence of the undecidable within a system depends, in each case, on a productive intervention on syntax, whereby the formal/syntactical rules governing the logic of the system as a whole are encoded at one specific point. Derrida is emphatic about this in 'The Double Session'; thus, 'what counts here' is, as we have seen, not the polysemy or ambiguity of the sign, but rather 'the *formal or syntactical* **praxis** that composes and decomposes it' (Derrida 1970, 240); (the syllepsis 'hymen' produces its effect in Mallarmé's text '*first and foremost through the syntax* ... in such a way that the suspense is due only to the placement and not to the content of words' (Derrida 1970, 220); again, it advances this effect by means of what Derrida calls 'the irreducible *excess of the syntactic* over the semantic' (Derrida 1970, 221; emphasis added). The terms that invoke the undecidable are thus, according to the discussion in 'The Double Session', *always* an effect of syntax. More precisely, they are locatable at the point at which syntax situates a kind of semantic gap or void essential to the text as such (Derrida 1970, 221). Thus, for instance *différance* does not name anything that can appear as a positive object or be positively (semantically) signified; if we had to give *différance* a semantic value, we could only say that it names a void of non-being. However, this void is marked *syntactically*, on the level of the formal, systematic structure governing the possibilities of signification, by the formal operation of differing and deferring that *différance* is. This syntactical spacing, though it does not correspond to any semantic correlate, remains structurally necessary, insofar as it *conditions* and *opens* the syntactic possibility of any signification as such (Derrida 1970, 222).

This 'between' of spacing is thus originally and purely a 'syntactic effect'; but, by way of an essential ambiguity involved in the structure of any system of writing, it also can *signify* the possibility of signification itself, and thus 'exceed', in a somewhat paradoxical fashion, a purely syntactic register. Thus,

> One no longer even has the authority to say that 'between' is a purely syntactic function. Through the re-marking of its semantic void, it in fact begins to signify. Its semantic void signifies, but it signifies spacing and articulation; it has as its meaning the possibility of syntax; it orders the play of meaning. Neither purely syntactic nor purely semantic, it marks the articulated opening of that opposition. (Derrida 1970, 222)

That is, the purely syntactical mark of the possibility of signification – for instance 'différance' or the 'trace' – although it does not signify anything itself, and thus *semantically* only signifies the void, nevertheless

articulates the possibility of spacing upon which all signification depends. In this sense it does signify after all, although the general possibility it signifies corresponds to the entirety of the system of signification and not to anything signifiable within it. The outcome of this ambiguous signification is the undecidability of the particular statement that accomplishes it, its retention of a systematic meaning despite its incapacity to be decided as true or false by any of the rules whose total syntactic logic it captures.

We can see how deeply this parallels Gödel's result. As we have seen, the Gödel sentence for any sufficiently complex formal system is a sentence that 'asserts' (in a metalogical, 'semantic' sense of assertion) of itself that it is unprovable in that system. Given that there is such a sentence, it is easy to show that neither it nor its negation can be assumed to be provable in the system, on pain of contradiction. However, how is the Gödel sentence itself obtained? The key step in the proof is the construction of a so-called 'provability predicate', which encodes the syntactic rules governing valid proofs in the system. This predicate is then employed to construct the sentence asserting of itself that it is not provable. But this is itself possible only through what Gödel calls the 'arithmetisation of syntax', the procedure of 'Gödel numbering' according to which syntactically defined formulae in the system are assigned discrete natural numbers. Through this procedure, the syntactical rules of derivation constitutive of proof in the system are rewritten as purely *arithmetical* relationships between natural numbers. It is just such a (purely arithmetical) relationship that the Gödel sentence ultimately asserts to hold. Thus, the arithmetisation of syntax makes possible not only the construction of a 'provability' predicate that encodes the logic of proof in the system, but also a 'fixed point' sentence, like the Gödel sentence, that 'makes reference to itself' by asserting that a certain property holds of its own Gödel number. The undecidable Gödel sentence itself is formed by applying the provability predicate thus derived to this Gödel number, and then negating the result.

In a recent, far-ranging work, Graham Priest has traced a large variety of paradoxes and problems arising in the history of philosophy to a single formal structure that he calls the 'Inclosure Schema', whose underlying basis is typically, again, diagonalisation (Priest 2002, 3–4; 133–6). The problems in which Priest is interested all arise from the consideration of various kinds of *limits*: for instance, the limit of what can be described, the limit of what can be known, or the limit of what can be conceived. There is, Priest argues, a general and formal contradiction that repeatedly arises when philosophers consider these

limits. The contradiction is that it is both *impossible* and *possible* for thought or description to cross these limits, generating an element that is outside the relevant totality and thus thinking the unthinkable (or saying the unsayable, etc.)[6]

More specifically, Priest argues that such a contradiction at the limits of thought will arise whenever it is possible to apply to a certain totality two operations, which he calls *Closure* and *Transcendence*. In general, *Closure* is the operation that defines the limit of the totality under consideration: for instance, we might determine the totality of what is provable in a particular formal system by means of a description of its axioms and inferential rules; or we might determine the totality of the 'metaphysics of representation' by a statement of its constitutive assumptions (including, e.g., that of the difference between the original and its representation). *Transcendence* is a more complex operation, but in general it involves the generation of an arbitrary element outside the given totality delimited by *Closure*. Diagonalisation is, Priest argues, a paradigm of *Transcendence*: for very many systems, it is possible by means of diagonalisation or a closely analogous operation to use the delimitation of the rules governing a system and defining its limits to define an element that *formally* looks like an element of the system, but is demonstrably not an element, and is in this sense 'beyond' the defined totality (Priest 2002, 4; 233). The combination of *Closure* and *Transcendence*, Priest argues, characteristically yields the paradoxical situation whereby the philosopher who attempts to define a limit must be unable to do so but succeeds in doing so anyway. Such is the situation, for instance, of Wittgenstein in the *Tractatus* when he holds both that the '*limits of my language* mean the limits of my world' and yet that the very propositions which articulate this limit must, for this reason, be nonsensical, neither falling within the system they define nor standing for any object that can exist in the world (Wittgenstein 1921, 5–6).[7]

Turning to Derrida, Priest finds a similar inclosure contradiction in the inscription of *différance* and the other 'undecidable' terms that are, as we have seen, formally akin to it (Priest 2002, 214–23). In particular, Priest focuses on Derrida's reading of Saussure and Rousseau in *of Grammatology*. On this reading as Priest understands it, each text is (as a condition for the possibility of its saying anything) structured by at least one 'binary opposition'. (This might be, for instance, the opposition between an original and its representation, or between truth and falsehood). Additionally, the totality of *all* expressions – or at least what we might think of as the entirety of the 'metaphysics of presence' – is structured by one single such binary, the opposition between presence

and absence. By considering the way that this binary, or the space of difference between its two terms, structures the totality of the sayable but without ever appearing as a positive term, we can directly, according to Priest, generate an inclosure paradox:

> We ... have a contradiction typical of a limit of thought. Claims about *différance* are not expressible (Transcendence); yet they are expressed (Closure). In fact, the contradiction fits the Inclosure Schema in a simple way. Let $\varphi(y)$ be: y is a linguistic expression. Let $\psi(x)$ be: x is structured by some binary opposition. Let $\delta(x)$ be some statement that concerns the notion undecidable in terms of such an opposition. (Such statements would typically occur in any text that deconstructs x). As we have seen, if x is a text structured by some binary opposition, $\delta(x)$ cannot be expressed in x. Hence we have Transcendence, but it is clearly a linguistic expression. Hence we have Closure. Finally, the totality of all expressions, Ω, is structured by the pair presence/absence ... The contradiction arises when the deconstructive diagonaliser is applied to the totality of all texts to produce a statement about *différance* (for example, one of Derrida's own). At this point, *le pas au-delà* (the step beyond) is *un pas au-delà* (a non-beyond) to use Derrida's own neat turn of phrase. One might well exploit Derrida's technique of writing under erasure, and call inclosures ~~limits~~. (Priest 2002, 222. The last word in this quotation is crossed out in Priest's text)[8]

According to Priest, then, the undecidable proper to any linguistic system, and determined by the underlying binary which structures it, cannot be inscribed within that system itself; and *if* we generalise this to the totality of language, we can directly produce the inclosure contradiction: a sentence (for instance a sentence 'about *différance*') that, standing for the 'between' of the 'master' binary presence/absence, cannot be inscribed in language at all (although clearly, it is).

Extending Priest's analysis, we can now recognise another general structural feature which (in addition to, or perhaps as a consequence of, the 'binary' presence/absence) structures and preconditions the possibility of all writing as such, according to Derrida.[9] This feature is the *iterability* of the sign, upon which Derrida famously insists in 'Signature, Event, Context':

> My 'written communication' must, if you will, remain legible despite the absolute disappearance of every determined addressee in general for it to function as writing, that is, for it to be legible. It must be repeatable – iterable – in the absolute absence of the addressee or of the empirically determinable set of addresses. This iterability ... structures the mark of writing itself, and

> does so moreover for no matter what type of writing ... A writing that was not structurally legible – iterable – beyond the death of the addressee would not be writing. (Derrida 1971, 315)

This structural iterability, as Derrida goes on to emphasise, implies the radical structural possibility of inscribing the sign within any of an open infinity of contexts or chains, what he later calls the possibility of 'graft' or citationality in general. This open possibility, Derrida says, implies that a written sign 'carries with it' as part of the 'very structure of the written' a 'force of rupture' or possibility of breaking with any particular context of inscription. This possibility of rupture, moreover, arises directly from:

> the spacing which constitutes the written sign: the spacing which separates it from other elements of the internal contextual chain (the always open possibility of its extraction and grafting), but also from all the forms of a present referent (past or to come in the modified form of the present past or to come) that is objective or subjective. This spacing is not the simple negativity of a lack, but the emergence of the mark. (Derrida 1971, 317)

That is, the possibility of iteration, and hence the standing possibility, inscribed in the very structure of the sign, of *rupture* with any determinate context, is a direct consequence of the spacing (both within and without the chain of signifiers) that, according to Derrida, makes possible the inscription (or legibility) of the sign itself. Thus, iterability effectively *diagonalises* any fixed or given context; in Priest's terminology, it ensures the possibility of Transcendence, which effectively generates an arbitrary element outside the totality of any fixed contextual Closure. Once again, moreover, this diagonalisation, as a general feature of inscription as such, is itself a direct outcome of the constitutive *spacing* that, although unpresentable itself and without a semantic correlate, makes possible any and all inscription as such by constituting the very possibility of the sign.

III

So far, I have argued, by way of the close analogy to Gödel's result and to diagonalisation more generally, for three claims about Derrida's undecidables. First, I have argued that they are always the result of a reflective intervention on *syntax*, by means of which the constitutive rules governing a system are problematically 'represented' within that system itself. Second, I have argued that this possibility of inscribing the undecidable is always also the result of the possibility of encoding in

ultimately syntactic rules some feature of the *totality* of the system itself, including in particular the totality of its 'decision procedures'. Third, I have suggested that the inscription of the undecidable, thus understood, gives critical thought access to a complex topology of the limit: neither the *closure* of the system by means of the drawing of a fixed and steady limit, nor its *openness* to the 'infinity' of a transcendent beyond, but what we can term, following Priest (but hyphenating to emphasise the derivation of the term), precisely in-closure. The system is closed only at the price of the inherent paradox of tracing its limits, and open just insofar as this paradoxical closure also operates as the diagonalisation that generates a contradictory point that is both inside and outside. The work of criticism, or deconstruction, in relationship to the problematic totality of metaphysics, or language itself, is henceforth the inscription and tracing (which also takes the form of erasure) of this problematic limit.

Now, as I shall argue in conclusion, it is helpful to understand the 'operation' of deconstruction in terms of this specific kind of topology. In his remarkable 1990 homage to Derrida, '*Pardes:* The Writing of Potentiality,' Giorgio Agamben treats the specificity of deconstruction as consisting in its ability to 'dwell in' the paradoxical topology constituted by the three essential features I have outlined above:

> *The concept 'trace' is not a concept* (just as 'the name "*différance*" is not a name'): this is the paradoxical thesis that is already implicit in the grammatological project and that defines the proper status of Derrida's terminology. Grammatology was forced to become deconstruction in order to avoid this paradox (or, more precisely, to seek to dwell in it correctly); this is why it renounced any attempt to proceed by decisions about meaning. But in its original intention, grammatology is not a theory of polysemy or a doctrine of the transcendence of meaning; it has as its object not an equally inexhaustible, infinite hermeneutics of signification but a radicalisation of the problem of self-reference that calls into question and transforms the very concept of meaning grounding Western logic ... It does not suffice, however, to underline (on the basis of Gödel's theorem) the necessary relation between a determinate axiomatics and undecidable propositions: what is decisive is solely how one conceives this relation. It is possible to consider an undecidable as a purely negative limit (Kant's *Schranke*), such that one then invokes strategies (Bertrand Russell's theory of types or Alfred Tarski's metalanguage) to avoid running up against it. Or one can consider it as a *threshold* (Kant's *Grenze*), which opens onto an exteriority and transforms and dislocates all the elements of the system. This is why the notion of the 'trace' constitutes the specific achievement of Derrida's thought. He does not limit himself to reformulating logical paradoxes; rather, like Heidegger – who in 'On the Way to Language' wrote, 'there is no word for the word,' and

> proposed an experience of language in which language itself came to language
> – Derrida makes these paradoxes into the place of an experiment in which the
> very notion of sense must be transformed and must give way to the concept
> of trace. (Agamben 1999, 213–14)

As Agamben emphasises, echoing Derrida himself, deconstruction is not a hermeneutic of meaning, either of polysemy or transcendence; nor is it originally grounded, in any sense, on any kind of decision on semantical meaning. Its ground is rather the undecidability of the reflection of *syntax* upon itself, and the problematic topology of criticism that this implies. The constitutive paradox of this topology is a result, as Agamben notes, of the more general paradox of reflection that arises from the absence of a 'name for the name,' the radical absence of any coherent reflection of the totality of a system's syntax within itself and without paradox or contradiction. In tracing the possibility of writing, in formalising the possibility of formalisation, deconstruction's task becomes the inscription of and 'dwelling within' this paradox. Its topological structure is not, as I have argued, closure or openness, but rather what Priest calls *inclosure* and what Agamben calls the *threshold*: not the limit of a fixed and determinate line between 'inside' and 'outside,' but rather the threshold of inclosure that, in being closed, opens to the exterior, and in being open, encloses itself.[10]

As Agamben notes, the response of deconstruction is precisely to inscribe and trace the undecidable, to multiply its occurrences and document its syntactical/reflexive necessity, thereby inhabiting – without resolving – the complex critical topology of inclosure. With this, the critical operation of tracing a boundary to thought, language, or expression becomes the inscription of a limitative trace that erases itself in the movement of its own inscription; far from being located at the outside of a bounded and stable totality of language as such, the line of this tracing henceforth inhabits or haunts language at each of its points, wherever (which is everywhere) it attempts or presupposes problematic self-reference through an reflexive invocation, explicit or inexplicit, of its own syntax.

The precise (meta-)structure of the undecidable I have argued for here, moreover, has important consequences for what we should take a deconstructive response to the 'undecidable' actually to be. This is important not only in relation to Derrida's initial examples of undecidability, but equally with respect to the more extended uses of the 'undecidable' of the later period, with relation to such 'ethical' structures as hospitality and the gift. Here, Derrida often employed what he has

called the aporias of 'possibility as impossibility' to demonstrate the inherent undecidables of particular 'ethical' situations. For instance,

> I am trying to elaborate a logic, and I would call this a 'logic', in which the only possible x (and I mean here any rigorous concept of x) is 'the impossible x'. And to do so without being caught in an absurd, nonsensical discourse. For instance, the statement according to which the only possible gift is an impossible gift, is meaningful. Where I can give only what I am able to give, what it is possible for me to give, I don't give. So, for me to give something, I have to give something I don't have, that is, to make an impossible gift. (Glendinning 2001, 55)

These structures of possibility as impossibility, like Mallarmé's *hymen* which inscribes the possibility/impossibility of mimesis, each suggest their own proper undecidable. Thus, for instance, because the pure gift is possible only as the impossibility of the structure of pure giving that makes the gift possible, the question of whether any particular empirical instance of giving is really a proper gift (or is, rather, simply an instance of trade within the general system of exchange, the system which regulates and calculates possibilities of exchange) is undecidable by means of any intra-systematic logic. To decide, then, is what cannot be done on the basis of any purely intra-systematic logic; to decide, we must in some way leap 'outside' the system, and here, we are without its guidance. But this 'outside,' as we have seen, is not a *simple* one; if undecidability, as I have argued, is always generated by the complex topology of inclosure, it is necessarily a *specific* 'outside' grounded in problematic reflection on this system's own limits, and gaining its specificity from the determinacy of the rules or principles that structure these limits. Moreover, if undecidability has the general structure I have suggested, such 'leaping outside the system' will not simply amount to 'breaking with' or leaving the structure of some *particular* system of representation (in favour, as it might be, of another). The moment of decision is (and remains even in Derrida's most 'ethical' moments) rather, the radical moment of the possibility/impossibility of *any* signification as such, of any possibility of representation or signification, and the ethics of 'responsibility' involved in such a decision must be understood in terms of its paradoxical relationship to this structure of possibility/impossibility.

As Derrida often emphasised, far from suggesting hesitation or indifference, what he calls the 'undecidable' is, for him, an essential *precondition* for the possibility of responsibility: for if we have only to decide what is already in some sense decided by the system in which we

operate, there is no responsibility in our decision.[11] Some commentators have emphasised the way in which this situates responsible decision as a relationship to an 'other' which is wholly outside the system of understanding or knowledge itself, what is sometimes called a 'singular.'[12] However, as we can now see, if this is an accurate description of what a 'responsible decision' is for Derrida, it can only be so because the 'singular' other is not *simply* outside the system of decision and knowledge (though of course not simply inside it either), but rather inhabits the paradoxical space of inclosure itself. This is, as we have seen, the paradoxical space marked by the undecidable; indeed, its marking or inscription as such, already amounts to an 'ethical' *praxis* of writing. Thus, if the ethical relationship of responsibility, for Derrida, is indeed always a relationship, 'beyond' knowledge, to a 'wholly other', this 'other' can only be the object of a radical responsibility insofar as it occupies the (non)-space of undecidability, which is not simply 'beyond' the limits of the system of decision but rather involves, in a constitutive way, the undecidability of that system's own closure. This makes it clearer, at least, how the marking or tracing of this problematic inclosure, in radical reflection on the possibility of totality, decision, or tracing as such, can itself amount to a relationship of responsibility, a kind of (arche)-writing of the ethical relationship as such.

Because of Derrida's continuing emphasis on the undecidable, he has sometimes been accused of an ultimately impotent doctrine of infinite textual tracing that removes the possibility of any real choice or action. Inscribing the undecidable, the criticism avers, is simply a way to postpone or preempt decision. But if, quite to the contrary, what Derrida calls the undecidable is the most direct *condition* for the possibility of a genuine decision, the meta-structural topology that inscribes the undecidable as I have suggested makes the nature of this conditioning all the more clear. For it is then possible to say that the genuine decision is conditioned *specifically* by the particular systematic undecidable that it answers to; and that responsible decision *as such* is conditioned by the complex topology of inclosure. That is, without suggesting that the 'correct decision' is determined materially, in an internal way, by the rules of the system, or even that it is determined formally, in an external way, by the demand of a relationship to an other thought as wholly outside them, it is possible to see in the paradoxical crossing of content and form that I have discussed and that has its best model in diagonalisation the opening of a kind of 'ethics', which is precisely what Derrida is insisting upon in these moments in which he affirms the necessity of responsible decision. This ethics has its ultimate basis

not in any form or content, but in the problematic moment of *reflection* whereby the total logic of the system is reflected back into one of its members, producing the diagonal and the undecidable. In a strange and even surprising way, then, we might see Derrida's ethics as the radical practice of reflection on the paradoxical topology of syntactic totality; in this specific sense, it is coextensive with the practical formalisation of formalisation, or the writing of writing, as such.

References

Agamben, Giorgio (1998), *Homo Sacer, Sovereign Power and Bare Life*, trans. Daniel Heller-Roazen, Stanford, CA: Stanford University Press.

Agamben, Giorgio (1999), '*Pardes*: The Writing of Potentiality', in *Potentialities: Collected Essays in Philosophy*, Stanford, CA: Stanford University Press.

Agamben, Giorgio (2005), *State of Exception*, trans. Kevin Attell, Chicago: University of Chicago Press.

Badiou, Alain [1988] (2005), *Being and Event*, trans. Oliver Feltham, London: Continuum.

Caputo, John D. (1997), *The Prayers and Tears of Jacques Derrida: Religion Without Religion*, Bloomington: Indiana University Press.

Critchley, Simon (1999), *Ethics-Politics-Subjectivity: Essays on Derrida, Levinas, and Contemporary French Thought*, London: Verso.

Derrida, Jacques [1968] (1981), 'Différance', in *Margins of Philosophy*, trans. Alan Bass, Chicago: University of Chicago Press.

Derrida, Jacques [1970] (1981), 'The Double Session', in *Dissemination*. trans. Barbara Johnson, Chicago: University of Chicago Press.

Derrida, Jacques [1971] (1985), 'Signature, Event, Context', *in Margins of Philosophy*, trans. Alan Bass, Chicago: University of Chicago Press.

Derrida, Jacques (1995), *The Gift of Death*, trans. David Wills, Chicago: University of Chicago Press.

Derrida, Jacques (2002), 'Force of Law: The 'Mystical Foundation of Authority', trans. M. Quaintance, in *Acts of Religion*, ed. G. Anidjar, London: Routledge.

Floyd, Juliet and Putnam, Hilary (2000), 'A Note on Wittgenstein's "Notorious Paragraph" About the Gödel Theorem' *Journal of Philosophy*, 97: 11, pp. 624–32.

Glendinning, Simon (ed) (2001), *Arguing With Derrida*, Oxford: Blackwell.

Gödel, Kurt [1931] (1965), 'On Formally Undecidable Propositions of Principia Mathematica and Related Systems I', reprinted in *The Undecidable*, ed. Martin Davis, Mineola, NY: Dover.

Hägglund, Martin (2008), *Radical Atheism: Derrida and the Time of Life*, Stanford, CA: Stanford University Press.

Hofstadter, Douglas (1999), *Gödel, Escher, Bach: An Eternal Golden Braid*, 20th Anniversary Edition, Basic Books.

Livingston, Paul M. (2008), 'Alain Badiou: Being and Event' *Inquiry*, 51: 2, pp. 217–38.

Livingston, Paul M. (2009), 'Agamben, Badiou, and Russell' *Continental Philosophy Review* 42, pp. 297–325.

Moore, A. W. (2001), 'Arguing with Derrida' in *Arguing With Derrida*, ed. Simon Glendinning, Oxford: Blackwell.

Nagel, Ernest and James R. Newman (2001), *Gödel's Proof*, Revised Edition, New York and London: New York University Press.

Priest, Graham (2002), *Beyond the Limits of Thought*, Oxford: Clarendon Press.

Schmitt, C. [1934] (2006), *Political Theology: Four chapters on the concept of sovereignty*, ed. George Schwab, Chicago: University of Chicago Press.

Smith, Peter (2007), *An Introduction to Gödel's Theorems*, Cambridge: Cambridge University Press.

Wittgenstein, Ludwig (1978), *Remarks on the Foundations of Mathematics*, Cambridge, MA: MIT Press.

Notes

1. More specifically: the Gödel Sentence, GS, has the form: GS$\leftrightarrow\sim$*Prov*([GS]), where 'Prov' is a so-called 'provability predicate' and the square brackets express the function of taking the Gödel number of what is enclosed in them. There are several good general introductions to Gödel's result available. These include: Nagel and Newman (2001); Smith (2007); Hofstadter (1999).

2. The reasoning goes as follows: suppose the claim asserted by the Gödel sentence GS is false; then it is false that GS cannot be proved; then it is true that it can be proved; but a proof of it is also a proof that it cannot be proved. Thus, we have a contradiction on the assumption that the claim asserted by GS is false; therefore it must be true.

3. Of course, Gödel himself was an avowed Platonist about mathematical truth, and saw his own result simply as demonstrating the capacity of the human mind to have a kind of intuition of truths that no formal system could by itself prove. However, this interpretation of the metamathematical significance of the Gödel sentence is not obligatory. In particular, it seems we may take the sentence to mark an impasse or limit of formalisability, without necessarily taking it to 'capture' a 'truth' that lies beyond the possibility of formalisation. For some considerations along these lines, drawing on Wittgenstein's reaction to Gödel's proof in Wittgenstein (1978), see, for example, Floyd and Putnam (2000).

4. A good general treatment of diagonalisation is Smullyan (1994). See also Hofstadter (1999) and Priest (2002).

5. Another philosopher who has recently made extensive and central use of diagonalisation is Badiou (e.g. 1988). In particular, Cantor's theorem (which shows that the power set of any set is of strictly greater cardinality than that of the set itself) allows us to consider the 'excess' of representation over presentation, which Badiou also connects, in a more political register, to the often necessary excesses of state power, as well as to the possibility of what he terms an Event to intervene in order to 'measure' these excesses and thus limit them. For summary and some related considerations, see Livingston (2008) and (2009).

6. More formally: Priest (2002, p. 134) argues that an inclosure contradiction will result whenever we have i) some total set Ω defined by a property φ and of which another property ψ obtains; and ii) there is some 'diagonalising' function δ such that, given any subset x of Ω that also has the property ψ, $\delta(x)$ is not an element of x and yet $\delta(x)$ is an element of Ω. Then we can produce a contradiction by considering the result of applying the diagonalising function δ to the totality Ω itself; because of the conditions, $\delta(\Omega)$ both is and is not an element of Ω.

7. See Priest (2002), pp. 184–94.

8. There are problems with, or at least limitations to, this as a reading of Derrida. In emphasising the double of presence/absence, and other similar 'binaries,' as structuring the whole of the sayable, Priest focuses on the sense in which *différance* is a result of (spatial or synchronic) difference. However, *différance*

explicitly includes, just as much, the temporal or diachronic dimension of iteration, and this seems to go missing from Priest's account.

9. In this paragraph I am somewhat indebted to the helpful argument of Moore (2001).

10. This topology of the 'threshold' is in fact exactly the same as the structure that defines what Agamben has elsewhere called the 'state of exception'; see for example Agamben (1998) and (2005). Agamben's original source for the idea of the exception and its role in constituting the normal order is Schmitt (1934).

11. See for example, Derrida (2002) and Derrida (1995).

12. See for example, Caputo (1997) and Critchley (1999) for this kind of reading; for a convincing recent argument against it, which I partially follow here, see Hägglund (2008), esp. chapters 3 and 4.