South American Journal of Logic
Vol. 2, n. 1, pp. 83–90, 2016
ISSN: 2446-6719

Gödel's Theorems and the Epsilon Calculus

Hartley Slater

Abstract

The epsilon calculus contains terms of the form '$\epsilon x F x$' for every predicate in the language. This means that it includes what I shall call 'empty' terms (when there are no Fs) and also what I shall call 'indexical' terms (when there is more than one F). These particular terms give the epsilon calculus a very distinctive character in comparison with the standard predicate calculus. For instance, an 'empty' term is central to understanding how we can do things that Whitehead and Russell's *Principia Mathematica* cannot do, in connection with Gödel's theorems. And attention to 'indexical' terms is crucial to solving the major logical paradoxes that have been a puzzle for over a century. Most particularly they are crucial to the solution of what has been called 'Gödel's Paradox', which has been claimed to show that natural language is paraconsistent.

Keywords: Gödel's theorems, epsilon calculus, operators and predicates, minds and machines, liar paradox, paraconsistency.

1 Proofs and Derivations

The overall point that has to be grasped to understand what is true in the standard model of Gödelian sentences is quite simple. For Gödel showed in his first theorem that, if the appropriate system is consistent, then we can say something of the form

$$(\forall z)\neg B(z, `(\forall x)Fx'),$$

although also

$$(\forall x)(\exists z)B(z, `Fx')$$

for a certain 'F', where '$B(z, `Fx')$' says that z is the Gödel number of a derivation of a formula 'Fx' in the system, and 'x' is the numeral for the natural number x in that system [13]. It is important in connection with what follows that only such numerals are involved in the formula, since only such can be used in the Gödel numbering. But on account of the second of these results we get the curious Gödelian conclusion: for it seems that it is thereby proved that every natural number is F. So it seems that we can obtain a universal conclusion that the system itself cannot establish.

But how do we do it? Do we have some 'naïve' method of proof that cannot be formalised? The classic epsilon calculus comes from adding to propositional logic the first epsilon axiom

$$Fy \supset F\epsilon xFx,$$

where 'ϵxFx' is an individual term, defined for all predicates in the language, making it invariably the case that

$$(\exists y)(y = \epsilon xFx).$$

On this basis, one can define the quantifiers, viz:

$$(\exists y)Fx \equiv F\epsilon xFx, \quad (\forall x)Fx \equiv F\epsilon x\neg Fx.$$

Epsilon treatments of arithmetic have been presented by a number of people influenced by Hilbert. Within this tradition one adds an axiom of extensionality

$$(\forall x)(Fx \equiv Gx) \supset \epsilon xFx = \epsilon xGx,$$

and the least number principle in the form

$$(\epsilon x\neg Fx = sn) \supset Fn,$$

where $sn$is the successor of n.

It is the distinctive way that generality is expressed in epsilon arithmetics that matters, and that can be illustrated in an epsilon proof of induction. For if we assume

$$F0\&(\forall m)(Fm \supset Fsm),$$

then, if $\epsilon x\neg Fx = 0$, we know that $F\epsilon x\neg Fx$ and so $(\forall m)Fm$, while if $\epsilon x\neg Fx = sn$, then Fn by the least number principle, and so Fsn from the universal premise, giving $F\epsilon x\neg Fx$ and so $(\forall m)Fm$ again. But in the Gödelian case, while there would be derivable '$(\exists y)(y = \epsilon x\neg Fx$', there would not be derivable '$y = \epsilon x\neg Fx$' for any numeral 'y', since otherwise we would have

$$(\exists y)[(\exists z)(B(z, \text{`}Fy\text{'}\&(\exists z)B(z, \text{`}y = \epsilon x\neg Fx\text{'})]$$

from which to derive

$$(\exists z)B(z, \text{`}(\forall x)Fx\text{'}).$$

And likewise from

$$(\forall y)(\exists z)B(z, \text{`}Fy\text{'})$$

we cannot get

$$(\exists z)B(z, \text{`}F\epsilon x\neg Fx\text{'}),$$

since the epsilon term is not a numeral, and so is not available as a substitution instance.

On the other hand, when we are talking about the standard model we are not talking about numerals and formulae using a predicative locution. We are talking about natural numbers and propositions (i.e., what numerals refer to and what the formulae express) using an operator locution, i.e. the crucial locution is 'It is provable that p' rather than '"p" is derivable'. The distinctiveness of the operator locution is not always appreciated, leading to some of the confusion in this area, since the derivability of a formula within a formal system has also come to be called the 'proof' of that formula. I shall here reserve the word 'proof' for the propositional notion, and word 'derivation' for the formula notion. Once crucial difference, as we shall see later, is that Löb's Theorem is not applicable to the propositional operator, as with the derivability predicate, since the operator 'It is provable that' necessarily entails the operator 'It is true that'. More specifically to the present point, since we know, if we choose the standard model, that for some natural number y, $y = \epsilon x\neg Fx$, then given that for that particular number there is a proof that Fy we have a proof that $(\forall x)Fx$ via the resulting proof that $F\epsilon x\neg Fx$. In symbols, if 'Pp' says 'It is provable that p' and we have

$$P(\exists y)(y = \epsilon x\neg Fx\&(\forall y)PFy,$$

then from this there follows that $(\exists y)(y = \epsilon x\neg Fx)$, and so that $PF\epsilon x\neg Fx$, and hence that $P(\forall y)Fy$. The epsilon term is now a possible substitution instance of the variable since, even though it is not a numeral, it still refers, through our choice of the standard model, to the natural number in question.

The epsilon treatment therefore shows that one can derive no formula of the form '$y = \epsilon x\neg Fx$' when 'y' is a numeral, yet we can still know that a sentence of this form expresses a truth in the standard model. But we know this by a process quite different to any formal proof procedure. Indeed, knowing any fact about the standard model must involve some non-formal process at some stage, since knowing the standard model is a model of the formulas in question (those in Peano's Axioms, for example) must be a non-formal process. Specifically we know the sentence above expresses a truth simply by interpreting the symbolism in a certain way, specifically the standard way. A Turing machine cannot specify what model its results are to be applied to, but we can. The point substantiates Searle's view about the limits of mechanism [7], [8]: only humans have the power to interpret what Turing machines might produce. Searle made this point in the extensive debate that arose from Lucas's paper 'Minds, Machines and Gödel' [2]. Lucas even suggested that the limits of mechanism demonstrated mankind's freedom of the will [3], and Penrose, following him, tried to locate some quantum structures in the brain that might account for this freedom [5], [6].

Whatever is happening in the brains of humans, though, the epsilon calculus, through the introduction of choice functions, breaches the mechanistic understanding of human language that dominated logic elsewhere in the twentieth century. We shall see this more fully in the remainder of this paper, where what I have called 'indexical' epsilon terms are taken as a simplified model for the complexities in natural languages brought about through considerations of context. Wider attention to epsilon terms would have alerted logicians to the presence of indexical terms in language, since 'ϵxFx' does not have a determinate reference unless $(\exists!x)Fx$. As a result there is no guarantee of self-reference in such an equation as "a = 'Pa"' if 'a' is an epsilon term. If you say that $\epsilon xFx =$ '$P\epsilon xFx$' then the choice of reference of the unquoted epsilon term is made. But the reference of the epsilon term within the quotes is not necessarily the same. The identity is with a sentence on the right hand side, and so is independent of its propositional interpretation.

2 The Liar

There are at least three reasons why Francesco Berto missed the above points in his recent discussion of Gödel's theorems [1]. The first was that he repeated, without qualification, improvement, or criticism much of what Graham Priest has said with regard to elementary paradoxes like that supposed to exist with sentences such as 'This sentence is false'. The second was that he also repeated what Priest has had to say about the more technical and specialised difficulties supposed to exist with sentences like 'This sentence is not provable'. But the third, and major reason concerns his attachment to the standard predicate calculus, and his inattention to Hilbert's conservative extension of it, the epsilon calculus. On page 210 of his paper Berto says:

> Gödel himself pointed out the analogies between his undecidable sentence and such paradoxes as Richard's or the Liar. But it seems clear that, whereas the Liar, 'This sentence is false', produces an antinomy, with the Gödel sentence, metamathematically read as 'This sentence is not provable', no contradiction is forthcoming.

But there is no contradiction forthcoming in the Liar case either, once one carefully separates sentences from the propositions they might be used to make. Sentences are mentioned commonly by using quotes; propositions, by contrast, are the referents of 'that'-clauses. So they are categorically different. Separating the two is particularly necessary when indexical expressions, like 'this', are involved. For 'This sentence is false' has many different uses, and so this same sentence can be used to express many different propositions. One may express, for instance, the proposition that the sentence 'The earth is square' is false thus:

This sentence is false: 'The earth is square'.

So Berto's sentence is not self-referential in itself, since its subject phrase has to be given the specific 'self-referential' interpretation in order for this to happen. And the most important thing to notice then is that the specific interpretation then involved makes its subject phrase refer to a sentence that contains a subject phrase without a determinate referent. For in

This sentence is false: 'This sentence is false',

the first 'this sentence' has a reference, but the second 'this sentence' has none, because at that second occurrence it is within quotation marks, and so is only mentioned, and not used. That means the proposition then expressed is simply false, since the referred-to sentence is neither true nor false. A host of other 'self-referential' paradoxes – including Grelling's and Russell's – yield to similar analyses [12].

3 Paraconsistency?

There is clearly no paradox with 'This sentence is not provable', on any basis. But an attempted application of the situation with formal derivability to natural language has led Priest, and now Berto, to hold that a very severe meta-paradox is derivable. For if Gödel's Theorems were applicable to natural language and not just a formal system like Whitehead and Russell's *Principia Mathematica* (PM), then that would seemingly make natural language paraconsistent. As Priest and Berto understand the matter, we would then both be able, and yet also be unable, to 'prove' the same thing, namely, a sentence of the form '$(\forall x)Fx$'. Of course application of *reductio* to this supposed result would have automatically shown there was something wrong with the assumption that natural language could be formalised in an appropriate way, but neither Priest nor Berto, as is well known, has been inclined to obey this logical law. In fact it is only recognition, on the above basis, of the difference between propositions and sentences that resolves the matter fully and thereby cancels the major theoretical argument Priest has run against the existence of classical contradiction, i.e. Boolean negation, in natural language.

For, first of all, what it is that we can *prove* in our natural language, in connection with Gödel's theorems, and what a formal system like PM cannot *derive* are categorically different [11]. One is a proposition – that all natural numbers are F, for a certain predicate 'F'– and the other is a sentence '$(\forall x)Fx$'. We get our propositional result by interpreting the formal symbolism a certain way, in terms of its standard model, as we have seen. That is a feat the system itself cannot even attempt, since it has no power over how its symbols are interpreted, being entirely formal.

But, second, there can be no replication of this separation starting not from PM but from natural language itself. So there are no Gödelian theorems for natural language. For it is easy to see that the fixed-point theorem, which is crucial to the Gödelian construction (and also Tarski's theorem), cannot arise in our natural language, since that is a language including indexicals (also empty referential terms), making the propositions expressible in such a language innumerable. It is here that my calling '$\epsilon x Fx$' an 'empty' term when there are no Fs and an 'indexical' term when there is more than one F becomes relevant. For the indexical and empty terms in natural language are like certain epsilon terms, specifically '$\epsilon x Px$' when it is not the case that $(\exists! x)Px$. Indeed it is the presence of such individual terms that distinguishes the epsilon calculus. If for every 'P' it were true that there was just one P, then this calculus would reduce to the standard predicate calculus.

Priest's argument is surprisingly weak at this point, since he simply assumes that natural language can be formulated as a standard predicate calculus, and does not address the problems that empty terms and indexicals pose to trying to get all propositions expressible in that language syntactically represented. The technical impasse, making the point especially clear in the case of the formal epsilon calculus, is that functions taking the related epsilon terms as arguments do not have calculable values, and so cannot be represented in a formal system. The upshot is that the supposition that natural language is itself paraconsistent cannot be defended using the Priest-Berto argument. In addition there is a simple proof, improving on Tarski's theorem, that the truth predicate attached to 'that'-clauses is consistent, since all operators are equivalent to predicates of 'that'-clauses (N.B., not predicates of *formulas*), and in line with this general equivalence 'that p is true' is the same as 'it is true that p'. But 'it is true that' is the vacuous modality in the logic KT, which is consistent, so truth predicated of propositions is consistent.

Likewise the provability of propositions is consistent, since as before 'It is provable that p' unconditionally entails 'It is true that p'. It is here that the fact that Löb's theorem does not apply to the proof of propositions, mentioned before, is most relevant. Of course it is crucial for this result to hold that the associated provability predicate is attached to 'that'-clauses as subjects, not formulas (or their names, or their Gödel numbers). For Löb's theorem does apply in connection with the derivation of formulas. But the fixed-point theorem on which this theorem is based does not hold for propositions, as we have seen. So, because of the close connection between Löb's theorem and Gödel's second theorem, that means that Gödel's proof of his second theorem does not hold for propositions.

What has probably led many people astray on the above two issues has been attachment to what often was their best friend, the standard predicate calculus. In twentieth-century formal logic it is presumed that a model for some predicative formulas consists in a non-empty universe where every individual term has a unique referent.

That means, as we have seen, that no individual term is an 'empty name', or an indexical, context-dependent expression. The numerals discussed at the start of this paper are a prime example; and there we also saw how different are epsilon terms. But thinking of all elementary terms as like numerals, every elementary sentence 'Pa' is taken to express a proposition, and just one such – that a is P. That means that self-reference does occur as a feature of the language, when $a =$ 'Pa', for instance. For the quoted 'a' has to have the same referent as the unquoted one on the given understanding. The result is that truth and provability cannot be defined in the language, i.e., Tarski's and Gödel's theorems hold. But that is no loss since the language so interpreted is unusable in any practical way. The formal predicative languages that have been common since the days of the logical positivists are abstractions from natural language, and it is that abstraction that has produced the limitative results. Indeed it has led to many paradoxes being produced that are not features of the richer forms of expression available in natural language.

The abstraction that has been involved has, as we have seen, removed the 'that'-clauses in natural language (trying to replace them with mentioned sentences or formulas), and it has also removed individual terms that either have no definite referent at all, or only one that is dependent on context. The two matters are intimately linked, since without such terms in the language there is little need to discriminate propositions from formulas, since every formula containing an individual term then expresses one and just one proposition. The conflation of predicates of 'that'-clauses with predicates of formulas has therefore been a central part of the subsequent difficulties, but the particular practical problem in the case of individual terms is that it overlooks the fact that we are not omniscient about what exists in the physical world. For we can sometimes be wrong when saying things like 'There is a mouse in the room' [i.e. '$(\exists x)(Mx\&Rx)$'], and even if we are right, then there might be more than one thing we could be talking about. So the individual term 'it' in any subsequent elementary remark, like 'It is on the carpet' [i.e. '$C\epsilon x(Mx\&Rx)$'], may have no determinate referent at all, or only one that has yet to be specified. As the epsilon symbolisation of these remarks shows, the epsilon calculus can handle these two further options, since the cross-reference is secured through the term '$\epsilon x(Mx\&Rx)$' arising not only in the elementary remark, but also in the epsilon equivalent of the existential remark. Many other cases have been analysed in a similar way [4], [9], [10]. But analyses cannot be formulated using any individual term in the standard predicate calculus, as we have seen. If we were omniscient, and knew beforehand the truth of any sentence of the form '$(\exists! x)Px$', we might be able to use the logical positivists' predicate language in the supposed way, i.e. in line with the standard model for the predicate calculus. But the humbling fact is that we are not omniscient in this way, making Tarski's and Gödel's theorems of academic interest only, being irrelevant to our practical use of ordinary language. Most particularly it means that natural language is not paraconsistent, as Priest and Berto have both maintained.

References

[1] F. Berto, The Gödel paradox and Wittgenstein's reasons, *Philosophia Mathematica* 17(2): 208–219, 2009.

[2] J. R. Lucas, Minds, machines and Gödel, *Philosophy* 36: 112–127, 1961.

[3] J. R. Lucas, *The Freedom of the Will*, Oxford University Press, 1970.

[4] W. P. M. Meyer Viol, *Instantial Logic: An Investigation into Reasoning with Instances*, ILLC, Amsterdam, 1995.

[5] R. Penrose, *The Emperor's New Mind*, Oxford University Press, 1989.

[6] R. Penrose, *Shadows of the Mind*, Oxford University Press, 1984.

[7] J. Searle, *Minds, Brains and Science*, Harvard University Press, 1984.

[8] J. Searle, *The Rediscovery of the Mind*, MIT Press, 1992.

[9] B. H. Slater, Epsilon calculi, *Logic Journal of the IGPL*, 14(4): 535–590, 2006.

[10] B. H. Slater, Hilbert's epsilon calculus and its successors, in D. Gabbay and J. Woods (eds), *Handbook of the History of Logic*, Vol. 5, Elsevier Science: 385–448, 2009.

[11] B. H. Slater, What Priest (amongst many others) has been missing. *Ratio* 23(2): 184–198, 2010.

[12] B. H. Slater, Consistent truth, *Ratio* 27(3): 247–261, 2014.

[13] C. Smoryński, The incompleteness theorems, in J. Barwise (ed.), *Handbook of Mathematical Logic*, North Holland: 821–865, 1977.

Hartley Slater
Department of Philosophy
The University of Western Australia
35 Stirling Highway, Crawley (Perth), WA 6009, Australia