The metalinguistic perspective in mathematics*

Michael D. Potter, Fitzwilliam College, Cambridge, U.K.

In this paper I consider three technical arguments in logic which have been held by various authors at diverse times to have consequences for the philosophy of mathematics. The first is Frege's permutation argument, which Davidson (1979) has used to show that proper names have a determinate reference only in a relative sense. The second is the Löwenheim-Skolem theorem, which (like its close relative, the theorem on the existence of non-standard models of arithmetic) apparently shows that the true sentences of a reasonably sophisticated first-order theory are not of themselves sufficient to fix the concepts of that theory. The third is the paradox of the set of all sets, which has been held by Lear (1977) and others to threaten a realist conception of set theory.

It is the last of these arguments which interests me most and which was my starting point in writing this article. What led me to consider the other two was the realization that all three arguments have a common feature: they can be understood as being conducted in a metalanguage which does not coincide with the object language. The question I intend to address here is whether the metalanguage is to be understood merely as being external to the specific (mathematical) object language about which the argument attempts to draw philosophical conclusions, or whether it rather has to be understood, if the argument's philosophical force is to be maintained, as external to *all* language. If the latter, there may be good reason to suppose the argument to be incoherent.

Frege's permutation argument is, of the three I shall be considering, the most mathematically straightforward and the most general in respect of the languages to which it may be applied. Suppose first that we have available to us an interpretation of some object-language. Suppose second that f is a non-trivial permutation of the domain of the interpretation. Define a new interpretation as follows: keep the domain unchanged; where the first scheme makes a name refer to an element a of the domain, let the second scheme make it refer to $f(a)$; and where the first scheme makes an n-ary

*Acta Analytica, **11** (1993) 79–86.

predicate refer to an n-ary relation R, let the second scheme make it refer to $\{(f(x_1), \ldots, f(x_n)) : (x_1, \ldots, x_n) \in R\}$. Plainly, a sentence is true in the second interpretation if and only if it is true in the first.

Now this argument appears to show that no ascription of reference to the names in our language is better than any other at grounding an account of the ascription of truth-values to the sentences of the language. Since, if we believe the context principle, that is the only rôle that an ascription of reference to names has to play, we are led to the conclusion that there is no reason to prefer any one referential scheme over any other derived from it in this way. Language, it seems, floats free of the world: it is not tied to it as we had hitherto imagined.

The first thing to note about this application of the permutation argument is that it is impossible for us to conceive of it as applying directly to our own language. The reason for this is, roughly speaking, that our language contains its own semantics. More precisely, the notion of reference in English is constrained by the disquotational scheme which requires that 'α' should refer to α, where α stands schematically for any name in English.

So for the argument to have any force we must first conceive of ourselves as applying it to another language, on the interpretation of which *our* disquotational scheme places no constraint. We then come to realize that if we can apply the argument to the language of another, then he or she is equally free to apply it to ours. So, at least, one might think. However, the argument does not sustain this grand conclusion. To do so, it would have to attribute to the metalanguage speaker access to precisely the privileged notion of reference which it attempts to deny to the object-language speaker. When it is shorn of this privilege, all that remains is an argument for indeterminacy of translation, which is of no relevance to the present discussion.

In order, therefore, for the permutation argument to cast genuine doubt on the notion that proper names have determinate reference, we must regard it as being formulated in an absolute metalanguage which is not itself susceptible to the argument. But if I imagine that I do have access to a language which is absolute in this sense, then it must be *my* language, '*the* language which I understand' (Wittgenstein 1933, 5.62). Hence the object-language, to which we are conceiving of the permutation argument as being applied, cannot be our language (or even a fragment of it). But in that case the conclusion of the argument is of no interest to us.

One is left, though, with a lingering doubt: even if the permutation argument does not, as was originally claimed, show that names do not successfully refer, nothing I have said so far shows that they do refer. An argument that they do based on an appeal to causal chains misses the point since the permutations which trouble us are ones which respect all the constraints to which speakers of the object language in question are subject, and causal links may very well be supposed to be among those constraints.

Nor can an appeal to simplicity succeed, at least in the case of language whose expressive power is very limited. Of a language with two names 'a' and 'b' and one binary predicate symbol 'R', there seems to be nothing to choose between an interpretation which makes 'a' mean Janet, 'b' mean John and 'xRy' mean that x is below y, and an interpretation which makes 'a' mean John, 'b' mean Janet and 'xRy' mean that x is above y.

Now the example I have just given is an artificially restricted one, and it may seem implausible that there might be two equally simple and direct interpretations of a language whose syntactic complexity approached that of our own. However, implausibility is not impossibility, and it is perfectly correct that there is nothing except artificiality that leads us to reject an account of French according to which 'Londres' means Dover and 'Douvres' means London, provided that we make attendant changes to the translations of all the predicates which might, in French, be applied to those names. This is of course very artificial and if all that is being attempted is the argument for indeterminacy of translation to which I alluded earlier, then much more plausible examples may be constructed. When it is treated as an argument for the inscrutability of reference, however, the mistake comes at precisely the point where we switch from considering the case where the metalanguage is our language and the object language is that of another community to the case where the rôles are reversed.

The only intelligible answer to the question 'Why do names refer?' is 'Because that is their linguistic rôle'. In other words, the explanation must be, as I have already urged, that there is no standpoint from which the question of reference for the names of the language can even be asked except for the standpoint from which we seek to attribute meaning to the sentences of the language: 'If everything in the symbolism works as though a sign had meaning, then it has meaning.' (Wittgenstein 1933, 3.328)

So the only sane resolution of the apparent challenge to our notion of determinate reference for names posed by the permutation argument is to deny the existence of an absolute position external to all language. The degree to which we are willing to countenance a view of our own language as being relative must remain flexible, depending on the purpose to which we intend to put such a view, but whatever degree of relativity we choose to ascribe, it is hypocritical to deny just that degree of relativity to the metalanguage also.

Before I turn to the Löwenheim-Skolem theorem, it is worth noting briefly the similarities between the rôle of that theorem in philosophical discussion and that of the existence of non-standard models of arithmetic. Both drop out of the completeness proof for first-order logic. Both suggest that the use we make of mathematical sentences is insufficient to fix the concepts we mention in those sentences. Both therefore appear to threaten the moderate realist position which holds that the truth-conditions for the sentences of mathematics are fixed by our mathematical practice, broadly interpreted. Both function by taking a model of the theory in question as given (the 'intended' model) and constructing from it a new 'unintended' model with different properties. However, in both cases the properties which differ between the models are inexpressible in the first-order mathematical language we started by considering. Because of these close similarities, I will restrict myself in this paper to considering Skolem's paradox; I hope it will be clear how what I say would apply to a discussion of the significance of non-standard models of arithmetic.

Skolem's paradox trades on the fact that nothing in first-order mathematical practice allows us to synthesize the second-order universal quantifier. We have an argument to show that for any list allegedly enumerating the real numbers we can construct a real number not in the list. The Löwenheim-Skolem argument constructs a model such that there is a way of counting the real numbers, although of course not in the model. Why should we be persuaded by this argument into thinking that the real numbers might 'really' be countable? Only because we make the same mistake as we did at first in considering the permutation argument: we are mesmerized by the metatheorist into regarding her as speaking with absolute authority; we conceive of her as having reached round the back of our language to set-theoretic reality itself. If instead we regard the language in which Skolem's argument is couched as being just another ordinary (mathematical) language like any other, then the force of the argument disappears entirely.

Putnam (Putnam 1980) has argued that the Löwenheim-Skolem theorem threatens the moderate realist, but not the intuitionist or the Gödelian platonist. If we disallow him any appeal in the metalanguage to the Gödelian resources which he plainly will not countenance in the object-language, then no threat to moderate realism from the Löwenheim-Skolem theorem remains. Of course, this is not to say that the moderate realist has no questions left to answer: once he has convinced us that an appeal to the context principle is sufficient to license us in ascribing reference to the names occurring in our mathematical language, he has still to answer the anti-realist challenge to provide an account of the truth-conditions for the sentences of the language. But this is a dispute in the resolution of which we have no reason to expect the Löwenheim-Skolem theorem to play any part.

However, the resolution of the threat of Skolem's paradox to realism which I have sketched here leaves a nagging worry that we still have not explained the mismatch which the paradox throws up between the meanings of the word 'countable' in the object-language and the metalanguage. It will be helpful to postpone consideration of this worry until after I have discussed the last of our three arguments, namely the paradox of the set of all sets. This argument differs from the other two in that the form in which it is generally stated makes no apparent use of a metalanguage at all. If we understand the unbounded universal quantifier, then we can quantify over all the things that there are. In particular, since sets are — according to the realist — among the things that there are, we can quantify over all sets. If we can quantify over them, then there seems *prima facie* to be no good reason why we should not refer to a set of all of them, or to a set of all those of them which are non-self-membered. And that way — as we all know — paradox lies.

To see how the paradox of the set of all sets is linked to the concerns we are discussing here, let us suppose that you, as an object-language speaker, have found a solution to the paradox which satisfies you. You have said, for example, that it is part of grasping the iterative notion of set to grasp that set formation is an indefinitely extensible operation, involving a creative tension between the extending operation of taking power sets and the completing operation of taking unions. You understand — or so you claim — why it is possible to quantify over all sets but not to talk of a set of all sets. I, on the other hand, wish to conduct a model-theoretic study of your mathematical practices. I therefore talk about your intended model of set theory. I conceive of this model, of course, as an ordered pair consisting of a set and a relation on that set. The set in question ('set' in the metalanguage sense) has as its elements all the sets ('set' in the object-language sense) which you intended your universal quantifier to quantify over. Yet *you* must, on pain of contradiction, refrain from regarding it as a set. But if you don't regard it as a *set*, what can you regard it as? Calling it a proper class is plainly just postponing the evil hour. It seems that you are prevented from referring to it, or (hence) conceiving of it, at all. Moreover, I did not set out, as one might argue I did in formulating the permutation argument or Skolem's paradox, purposefully to subvert your intentions; I was just trying to do some model theory. There is indeed the consequence of what I have done that my account of your use of the universal quantifier is at odds with yours, but I did not set out to achieve that. I am, on the face of it, at a loss to explain what I have done wrong, and have therefore reinstated the paradox.

The threat to a realist view of set theory is clear. I can refer to a set of which you cannot conceive. That set therefore exists for me but not for you. A realist cannot countenance the notion of existence fragmenting in this manner. Yet the alternative of regarding the application of model-theoretic techniques to set theory as illegitimate is unacceptable. Model-theoretic results are as impeccable *mathematically* as the results of any other area of mathematical enquiry. If there is a flaw, it must be in our interpretation of those results. Moreover, it would clearly be undesirable to have a solution which depended on the particular features of the set-theoretic example in question. It is not that model-theoretic results have one meaning when applied to Peano Arithmetic or geometry and quite another when applied to set theory, since that view leaves out of account the generality of model-theoretic methods; it makes set theory a special case, and hence once more paradoxical.

If the answer is to be generally applicable, and is not to bring into doubt the validity of model-theoretic methods *in themselves*, it must be that there is something which the metalanguage loses simply by not being the object language. The short answer is just that the object language is *your* language and the metalanguage is *my* language, but that sounds more like an excuse than an explanation. There must, if we are to give an adequate explanation of our use of a metalanguage, be something we can identify in the practice of a linguistic community which the metalanguage perspective does not leave room for. This is the ability of members of such a community to grasp indefinitely extensible concepts. What is distinctive about the metalinguistic perspective is that it regards the object-language as syntactically fixed, its interpretation the only thing about it which is yet to be determined. It therefore falsely imprisons us, as speakers, within that language, so that when we are presented with an object which we can recognise as being a set and therefore as coming within the range of our quantifiers, we are left apparently speechless, without the linguistic resources to say that it comes within that range. Nevertheless, the fact remains that we *can* say so. It must follow that a perspective from which language appears to be closed is of necessity a distorted one.

It may be instructive to compare here a remark of Gödel on Turing's argument that the human mind functions like a machine (Gödel 1986–2003, vol. II, p. 306, his emphasis): 'What Turing disregards completely is the fact that *mind, in its use, is not static, but constantly developing,* i.e., that we understand abstract terms more and more precisely as we go on using them, and that more and more abstract terms enter the sphere of our understanding.' If we view the human mind as 'not static, but constantly developing', it follows that we must also regard language in the same light.

Now I placed some stress on the requirement that the solution to the set-theoretic paradox should not be a particular one, but should apply quite generally to all uses of the metalinguistic perspective. It is therefore encouraging to note that the solution I have just suggested proves helpful in resolving the worry about Skolem's paradox which I left unresolved. That was the worry that the object-language and metalanguage meanings of the word 'countable' might be different. We can see that in order to arrive at that conclusion it was necessary to make use of precisely the interpretation of model-theoretic reasoning that we have now learnt to regard as suspect. The proof of the submodel form of the Löwenheim-Skolem theorem works only because it treats the object-language as closed. It takes a fixed model of that language and throws away large parts of it, preserving only what is necessary in order to keep unchanged the truth-values of all the sentences of the language. But, as I have been arguing, it is a mistake to see the relationship between language and the model in this way.

References

Davidson, D. (1979), 'The inscrutability of reference', *Southwestern J. Phil.,* 10: 7–19

Gödel, K. (1986–2003), *Collected Works*, Oxford University Press

Lear, J. (1977), 'Sets and semantics', *J. Phil.,* 74: 86–102

Putnam, H. (1980), 'Models and reality', *J. Symb. Logic,* 45: 464–482

Wittgenstein, L. (1933), *Tractatus Logico-Philosophicus*, corrected edn, Kegan Paul