(Against) Michael Potter on Gödel and Carnap

Sharon Berry

November 26, 2008

1 Introduction

In 'Reason's Nearest Kin' Michael Potter raises some problems for Carnap's conventionalist philosophy of mathematics. The first problem is a dilemma: if Carnap's mathematical conventions are recursively enumerable then (by the Incompleteness Theorem) they cannot suffice to let us answer all questions in arithmetic, so there will have to be some portion of of mathematics which is not conventional but requires "empirical" investigation. On the other hand, if the conventions are not recursively axiomatizable then in what sense can we, finite creatures be said to follow them? I will argue that neither horn of the dilemma is as problematic as Potter seems to think.

Second, Potter argues that, if math is conventional, we run the risk of adopting inconsistent conventions (i.e. conventions that allow the derivation of 0=1 and hence every other well formed symbol string). If we were so unlucky as to be working with inconsistent mathematical conventions then our mathematical terms would fail to acquire meanings, so all our mathematical talk - perhaps all our talk in general- would be meaningless. But, says Potter, it wildly implausible to say that we now run -even a very small- risk of everything we say being meaningless. So, conventionalism leads to the unacceptable conclusion that everything you are now saying might be meaningless. I will argue that Carnap can defend himself against this criticism as well. The real problem for Carnapian conventionalism lies elsewhere, in territory pointed out by Quine.

2 Carnap and Incompleteness

Carnap proposes a conventionalist philosophy of mathematics (and a priori reasoning in general). According to this story we lay down (or implicitly accept) framework principles which then determine the meaning of our words. When we are working within a given system of framework principles these give us a grip on how to answer any questions of mathematics or science that might arise (though experience may also be necessary). But if we are choosing between two different incompatible ways of expanding our framework (e.g. should we accept the proposition-language which tells us to say "there is a proposition which ..." under certain circumstances, or a nominalist language which tells us to say there are no propositions and all existing objects are in space and time?) there is only a pragmatic question of deciding (in light of the formal properties each framework) which would be most useful to adopt.

This gets a little complicated in the case of mathematics because you will need to use the mathematics in your current background theory to figure out what formal properties your candidate framework extensions have. Thus if you ever add to your framework in a way that makes your mathematics strictly stronger you will not be able to prove in advance whether your candidate new system has the one formal property which you probably care about the most: consistency. This is because if you could prove that the stronger system was consistent this would prove that your current system is consistent but - if your current system is recursively axiomatizable and consistent [1] - G/"odel's Incom-

[1] and strong enough to capture arithmetic

pleteness theorem says you can't do that.

So there's a way in which each addition to our mathematics which increases what it can prove is a leap in the dark. You can't prove in advance that your addition doesn't add **so** much strength that your new framework will let you 'prove' "0=1" and everything else. This is not to say that you can never hope to eventually prove the framework is consistent. But, by a cruel twist of mathematical necessity, you can only do this in circumstances where it won't do anything to reassure you that the new framework is consistent. Perhaps you go from the original framework F1 to the strenghtening F2, and then some day later you go to a further strengthening F3 which allows you to prove certain facts about what's derivable in F2. In this case you will be able to "prove" that your leap in the dark was made safely i.e. the framework F2 you accepted was consistent. But remember that F3 is strictly stronger than F2. So if F2 were *inconsistent* then F3 would be too - so of course it would prove "F2 is consistent" - along with "0=1" and "3=5+5" !

Perhaps the right response to this situation just to pick a good looking mathematical framework (say one that looks to be simple and hasn't let anyone derive a contradiction yet) and stick with it until and unless you should ever actually find yourself proving 0=1. Then at least we can purge mathematics from the necessity of making these distressing arational-looking Kirkeguaardian leaps in the dark whenever our mathematical system needs to be strengthened. Just pick, say, the framework of talking about numbers and sets and leave it at that. But here Incompleteness causes problems again. For, if your framework of talking about numbers is recursively axiomatizable - i.e. if you could program a computer to check whether proof proceeds in accordance with the framework - then it won't be able to derive everything that is intuitively true about the numbers. For example, there will be sentences of the form "there is a number with such and such an arithemtical property" where you framework lets you derive that 1 doesn't have the property, and 2 doesn't and so on for every **particular** number but you cannot derive the above statement that **all** numbers have that property. So if you dig in your heels and banish existential angst from mathematics by just choosing one framework and sticking at that, there will be some things that your preCarnapian intuitions say must be either true or false but your framework does not allow you to prove.

So now you have the technical and historical background for Potters' criticism.

3 The first horn of Potter's dilemma

Potter's first criticism is a dilemma. Is your mathematical framework recursively axiomatizable? The first horn goes like this: If Carnap says his mathematical framework is recursively axiomatizable then (as we saw above) there will be some claims (like every number has this property which is easily checkable in the framework) which a) intuitively are either true or false but b) can't be proved or disproved by employing the rules which make up the mathematical framework. In itself this isn't a problem for Carnap. The idea of frameworks is supposed to apply to all parts of our cognitive life, including empirical science. So Carnap definitely accepts that there should be questions like "Are all ravens black?" which can be asked within a framework, but which merely considering the rules of for the scientific framework doesn't let you figure out.

But the problem is that Carnap takes all truths to be either analytic or empirical. And (very characteristically of his turn of mind) he gives sharp definitions for these terms. First we define "analytic" sentences as ones which can that are derivable from the framework principles. This fits with the idea that the framework principles for a language fix the meaning of its terms, and that analytic truths are supposed to be 'true in virtue of meaning'. "Synthetic"/"Empirical" sentences are those which aren't analytic (and whose negations also aren't analytic). Then we define "analytic" *terms* as ones that **can only figure** in sentences which are derivable (or whose negations are derivable) from the framework principles and "empirical" terms as those which lack this property. This captures the idea that "...is red" might be a piece of empirical vocabulary even if "there is a red ball or there is not a red ball" is analytic.

These definitions mean that (if he takes the framework principles for mathematics to be recursively axiomatizable) Carnap will classify mathematical language as "empirical" and certain mathematical sentences as "empirical". Potter thinks its unacceptable to say that basic sentences of arithmetic (the ones that *can* be proved in the framework) are analytic but more complex sentences whose truth is determined by these sentences are empirical.

I claim that this move is not (in any readily apparent way) unacceptable or even implausible. Everything depends on what one means by 'empirical'. If by 'empirical' we meant 'commits one to the existence of new objects' or 'requires sense investigation to discover' then it perhaps it is strange to think that unquantified statements in arithmetic are analytic but quantified statements (whose truth is intuitively determined by the totality of these unquantified arithmetical statements) are empirical. Even this is not clear. Those who accept Quine's criterion would be happy to say that particular unquantified arithmetical claims don't commit one to the existence of objects but (some) [2] quantified claims about all the numbers do. And -even without considering Incompleteness - there are cases where you would need to use a computer to produce a certain proof in accordance with your framework. Hence empirical observation about e.g. the laws of electricity would be requisite for you to learn some truths which are derivable from by your mathematical framework. So you

[2] the Quineans would probably say 'all'

might (actually) need to use empirical methods to determine the even claims that follow from your analytic framework principles in the most direct way. Is it so mysterious that more some more complicated quantified statements whose truth is determined by the meanings which your framework principles fix in such a way that one **necessicarilly** needs to use empirical methods to discover them?

But neither of these claims is really to the point, since the problem only arises if we don't take Carnap as his word that he is *stipulating* what he means by "analytic" and "empirical". His definition is what gives us the odd sounding conclusion that some mathematical claims are "empirical" - so we can't first use this definition and then substitute in some other more intuitive definition to make his view look implausible. The claim that some mathematical statements are empirical simply amounts to the claim that some mathematical statements (like most scientific statements) cannot be arrived at purely by following the relevant framework principles. There is no kind of conflict between this and the claim that other, more basic, mathematical statements *can* be derived merely from the framework principles.

We can picture it like this: laying down some stipulations fixes the meaning of your terms in such a way that other things (e.g. blah is not provable) become true, though you may not be able to figure them out just by following the rules.

4 The second horn of Potter's dilemma

Turning now to the second horn of Potter's dilemma, let us suppose that the framework principles for math are supposed to allow rules which are not recursively checkable such as 'infer '$\forall$ (x) F(x)' if 'F(n)' is true for each particular number n'. In this case, Potter objects that it is problematic to suppose that we finite beings could count as following such a rule. How could we be 'implicitly grasping' a rule that's not recursively axiomatizable? If a computer couldn't be programed to follow this rule how could we?

The answer is, I think, that there's already some abstraction involved when we say that a person is following a rule. We don't follow mathematical rules in the way that say, the motions of our bodies obey the laws of physics. We count someone as following a given rule for how to add two numbers even if they are disposed to make mistakes - i.e. even if they don't always act in accordance with the rule. Their behavior just has to be close enough to look like an attempt to follow the rule (and maybe satisfy other criteria like not being produced by a look up table, being the behavior of a creature with conscious experiences etc.)

So, when you think about the sense in which we are following **recursive** rules, (i.e. we aren't literally disposed to behave in accordance with them but act like we are "trying" to follow them) it doesn't seem like very much of an abstraction to say that a person acts like they are trying to follow a non-recursive rule. So, in the case above, we see someone who mostly writes down $\forall$ (x) F(x) in cases where indeed F(n) is true for each particular number n. But being a finite being they can't be sure of ding this by running through every number n. So sometimes they will write down $\forall$ (x) F(x) in a case where there is some n such that (if specifically prompted with the number n) they would say F(n). In these cases, whenever they do happen to be so prompted (or just wind up calculating whether F(n) for some other reason) they then slap their foreheads and go back and erase the $\forall$ (x) F(x).

Here you might plausibly say that someone is trying to follow an infinitary rule. Of course the infinitary rule doesn't correctly describe their actual behavior but neither do finitary ones like the rules for how to add two numbers.

Thus I claim that neither horn of Potter's dilemma poses much of a problem for the Carnapian.

5 The threat of inconsistentency

Potter's second objection to Carnap turns on the threat of inconsistency. We might have an inconsistent collection of framework principles i.e. one that allowed us to derive 0=1 and thence every other string in our language. If so these framework principles would determine the meaning of each of the words in the framework so as to ensure the truth of each of these sentences - leaving us with an essentially meaningless language in which every possible string expressed a truth. Insofar as there is the threat (however small) that we may actually be using inconsistent framework principles there is the threat that our whole language is meaningless in just this way. But (says Potter) we can be absolutely certain that our language isn't meaningless in this way. Thus, a view on which there's even a tiny chance that we may have accepted an inconsistent framework and hence everything we say may be meaningless is unacceptable.

Now, people clearly did actually accept an inconsistent axioms for set theory (and hence, what Carnap would count as an inconsistent framework). That is a fact of recorded history. So Potter's objection can't be that Carnap's conventionalism ignores some magical faculty of rational intuition which swoops in and prevents us from accepting inconsistent axioms. We did accept inconsistent axioms and no such faculty swooped in.

Rather, I take it, Potter has in mind that intuition that, when you say to yourself "here I am reading a boring paper on philosophy of math" there's a sense in which you can be absolutely sure this sentence can't meaningless (or at least that **some** of your sentences aren't meaningless). On the other hand, reflection on history - both the history of ideas and one's own personal history of accepting fallacious arguments an concepts that turn out to be incoherent - suggests that you *might* have accepted some inconsistent principles for a priori reasoning/'framework principles'. Now at first glance this tension -between your feeling that it's obvious that your sentences mean something and the possibility that you have accepted an incoherent concept- isn't so much a problem for Carnap as a problem for everyone. So I suggest that we first think about what we *want* to say about the situation and then ask whether Carnap is blocked from agreeing with this.

So what **do** we want to say about e.g. Frege's talk of extensions prior to the discovery of Russell's paradox? It seems to me that there's a sense in which Frege's thoughts essentially involve the incoherent naif concept of set or extension. So it wouldn't be *precisely* right to translate him as thinking something about, say, sets of objects falling under a concept in the modern (ZF centric) sense of the word 'set'. Nor would it quite capture Frege's state of mind to associate his thoughts with some claim like 'there are some things, the extensions, which satisfy such and such axioms (including Basic Law 5) and one of them...' So it seems to me there's a very strict way in which we should say that when Frege said to himself "I am not in the extension of the concept horse" he was thinking something incoherent which we cannot associate with any coherent proposition or set of truth conditions. On the other hand, as Frege sat there, he would have had the same feeling of understanding and indubitably expressing a claim as we do when we say "I am not a member of the set which contains all and only horses" or "I am not a horse". And the role of this sentence in his behavior and other inferences would have been rather similar to the role of the above sentences in ours. So there's a looser way in which we might get at **something** real about Frege's mental state by translating him as thinking any of these three things: 'There is a thing called the extension of the concept horse, which satisfies certain axioms and it does not include me' 'I am not a member of the set of horses' 'I am not a horse'. Which one we pick will, naturally, depend on our purposes.

In addition to this, I think we should note that we don't take Frege's acceptance of one inconsistent axiom (and hence, in a sense, of inference rules which would allow him to derive every sentence in his language) to propagate out and strip all his other sentences of meaning. So we have no qualms about saying that Frege meant and said something, in the most fullblooded sense when he wrote "The wind howls outside my door".

So, this is what I think we want to say about people who wind up accepting inconsistent axioms. There's one sense in which these axioms **do** constrain the meaning of certain words that figure in the axioms so as to make these words (strictly speaking) meaningless in that person's idiolect. And there's another, looser, sense in which we can make sense of such a person's mental life by ascribing various meanings to sentences in which these unfortunate words occur. Neither of these facts is incompatible with the phenomenology of saying to yourself "obviously I am not in the extension of the concept horse, and obviously the claim I just made isn't totally meaningless". Furthermore, we aren't inclined to let the acceptance of some inconsistent axioms as strip all the claims in the person's language of meaning (or non-trivial truth conditions) even though there's a sense in which the person in question accepts rules which allow them to generate contradiction and thence every sentence in their language.

Now, how much of this can Carnap accept? What should we say about people who make the pragmatically bad choice of adopting a framework that allows them to derive every sentence in their language? To my knowledge, Carnap never explicitly addresses this question (how to translate people who accept inconsistent frameworks?) and has few relevant commitments. However, I don't see why he would have to dispute the phenomenology in question, or the practice of translating people using incoherent concepts using various nearby coherent concepts.

However, accepting this answer raises the related worry that the principle of tolerance might commit us to saying that Frege wasn't **wrong** about anything - even particular claims like "these axioms are consistent" - but rather that he was just working in a different framework, whose word "consistent" doesn't mean the same thing as the word "consistent" does in ours. This **would** be an unintuitive commitment of Carnap's view.

Now it seems to me that Carnap does not need to say this. I mean, think of the kinds of things which make us want to say that Frege was wrong about whether his system was consistent rather than just expressing a different thing than we do by this word - his dismay when he heard from Russell, his immediate abandonment of Basic Law 5 etc. These aren't the kind of behaviors which make us look at someone and say that their expressions mean whatever it takes to make their axioms come out right. Rather, they look like just the kind of behaviors mentioned in the section above which would make us say that someone's actual behavior is an imperfect attempt to follow an infinitary rule. So, if one takes the second horn of Potter's dilemma, one can parry the point on behalf of Carnap by saying that what these behaviors show was that Frege's axioms were not **really** his framework principles since he was willing to reject these axioms in response to certain other considerations (whatever principles governed that were his *real* framework principles). In this case his real rule was something more like (what we would express as): you can write down "S is consistent" if for every number n "n doesn't code the proof of 0=1 from S"

However, I think this brings us to something which really is a serious problem for Carnap. For, if we adopt the solution, it starts to look mysterious what the real un-give-up-able framework principles are, and whether these principles would really look like framework principles rather than some kind of post-Quinean reflective equilibrium.

6 Conclusion

In this paper, I have argued that the Carnapian conventionalist can address a number of criticisms of his view, including two main ones posed by Michael Potter. Carnap's account is (contra Potter) not sunk either by the fact that mathematical stipulations will be either non-recursive or incomplete, nor by the fact that framework stipulations could be syntactically inconsistent. Carnap's theory can easily be expanded to account for these phenomena in an intuitive way. However, it must be admitted that once these minor changes have been made, Carnap's view winds up looking very much like Quine's.