On Some Subsystems of Second-Order Arithmetic – 1

Induction, more or less

Peter Smith
Faculty of Philosophy, Cambridge

November 25, 2007

Certain writers, after first dealing their material a rough blow, then say that it naturally falls into two parts . . .

G. C. Lichtenberg

0 Introduction

The first main topic of this paper is a weak second-order theory that sits between first-order Peano Arithmetic PA_1 and axiomatized second-order Peano Arithmetic PA_2 – namely, that much-investigated theory known in the trade as ACA_0. What I'm going to argue is that ACA_0, in its standard form, lacks a cogent conceptual motivation.

Now, that claim – when the wraps are off – will turn out to be rather less exciting than it sounds. It isn't that all the work that has been done on ACA_0 has been hopelessly misplaced: that would be a quite absurd suggestion. The mistake, if that's what it is, has been a relatively small one. Still, we really ought to try to put things into conceptual good order here. That's part of what philosophers are for.

Here's the structure of my main claim. On the one hand, interesting work on ACA_0 actually only uses *part* of the strength of the theory: or as we might put it, the interesting work is actually carried on in a cut-down theory I'll call ACA!. This theory, I'll be claiming, *does* have a good conceptual motivation – it is in fact the theory that the putative conceptual grounding for ACA_0 *actually* underpins. On the other hand, I'll be arguing that original-strength ACA_0 inductively inflates. I mean, to put it more carefully, that anyone who accepts ACA_0 as a cogent theory can have no reason not to accept a certain significantly stronger theory, with a stronger induction principle. This stronger theory is standardly known as plain ACA. So, my claim comes to this: you can either go for the cut-down theory ACA!; or you can go for the much richer theory ACA. What you can't do is – I mean, what you can't have a stable conceptual motivation for doing – is to rest content with the intermediate strength ACA_0 in its standard presentation. Yet in much of the literature, in particular in Simpson's encyclopedic book *Subsystems of Second-Order Arithmetic* (1991), neither ACA! nor full ACA gets so much as a mention, and the conceptually unstable theory ACA_0 gets all the glory.

Why is my claim at all interesting? For at least two reasons. As I'll note first, ACA_0 plays a key role in the project of 'reverse mathematics' explored in *SoSOA*, which we can usefully think of as (in part) an exercise in calibrating the infinitary presumptions of various theorems of classical analysis. If we are going to go in for this project at all, then we really ought to be clear and careful about the presumptions of the calibrating theories (in particular, as we will see, we should be alert for cases where fully accepting a theory requires taking aboard more than is explicitly asserted by the theory). Second, the status of ACA_0 and related theories is tied up with the fate of Daniel Isaacson's Thesis that first-order PA_1 marks a boundary to the set of truths expressible in L_1 (the language of first-order arithmetic) that can be established using purely arithmetical ideas plus logic. The basic issue is this: it might seem that ACA_0 is what we get when we put together purely arithmetical ideas with a broadly logical idea (the idea that arithmetical predicates have extensions); but I'm going to argue that taking on a thorough-going commitment to ACA_0 in fact commits you to endorsing full ACA. And that latter theory isn't conservative over PA_1. So all this would threaten the Thesis, if arithmetic plus logical ideas were indeed to take us as far as ACA. But the argument will be that in fact those ideas only can take us as far as $\mathsf{ACA!}$ and thus that the Thesis lives to fight another day. The fate of Isaacson's Thesis is evidently an important and interesting question quite apart from any concerns we might have for the project of reverse mathematics.

This paper then has a supplementary topic. For I will also consider a different family of extensions to PA_1, this time not second-order theories but theories that add a truth-predicate to first-order arithmetic, and which then differ among themselves about the amount of induction that they permit for the new predicate. Two standard such theories are often labelled $\mathsf{PA}_1 \cup \mathsf{T}$ and $\mathsf{T}(\mathsf{PA}_1)$; and I'll be arguing – analogously to before – that the seemingly minimalist theory $\mathsf{PA}_1 \cup \mathsf{T}$ in fact inductively inflates to $\mathsf{T}(\mathsf{PA}_1)$, again meaning that someone who *fully* accepts $\mathsf{PA}_1 \cup \mathsf{T}$ has no good conceptual grounds for not endorsing the stronger $\mathsf{T}(\mathsf{PA}_1)$ with a stronger induction principle. This again looks, prima facie, to be bad news for Isaacson's Thesis, since $\mathsf{T}(\mathsf{PA}_1)$ like ACA is not conservative over PA_1: but again the sting can be taken out of the argument. And it will turn out that the connection between our two topics is even closer than that parallel makes it sound.

I begin with some general scene-setting; ACA_0 makes it first appearance in §3; truth-theories are discussed in §6.

1 Calibrating infinitary commitments

Suppose we are interested in the project of reconstructing chunks of classical analysis – say, the applicable theory of piecewise continuous functions – in set theoretic terms. For simplicity, we'll take natural numbers as already given as urelemente. Then, according to the usual story, we construct the integers as equivalence classes of ordered pairs of naturals, which takes us three steps up the set-theoretic hierarchy if we implement pairs in the now standard way. Then we construct the rationals as equivalence classes of ordered pairs of integers, which takes us another three steps up the hierarchy. All very familiar, but – of course – also all wild overkill! We just don't *need* to take powersets six times starting from the naturals in order to deal with rationals: after all there are no more rationals than numbers. It will do just as well to set up an effective bijection between naturals and rationals, and just deal with rationals via their natural-number codings.

Where we *do* begin to need some set-theoretic oomph is at the next step, where we construct the reals, say as Cauchy sequences of rationals (we can take these proxy reals to be sets of natural-number codings for pairs which code a sequence place and a code-for-a-rational). Still, you might say, that this first level of commitment to sets of numbers won't take us very far: since as soon as we start wanting to talk about *functions* defined over the reals, won't we need to go up at least one more level to get ourselves pairs of reals? Not so; or at least not so, if we stick to thinking about piecewise continuous functions of the traditional kind, so we are not countenancing functions in the modern stretched sense of quite arbitrary correlations between reals. Nice smooth functions can be characterized by what happens to them at rational arguments. And, to cut the story short, we can handle such functions by more coding tricks to keep them at the same level as the reals (see the extended treatment in *SoSOA*). In summary, then, a second-order arithmetic, which can talk about just numbers and sets of numbers, can indeed give us a framework for constructing a theory of the reals and sensibly behaved functions over the reals.

And now the following project suggests itself. We can ask: *just how strong a theory of second-order arithmetic is needed to prove a proxy for this or that classical theorem of analysis?* In particular, just how strong a comprehension axiom do we need, telling us which sets there are, in order to prove various theorems?

Now this project might be motivated by sceptical thoughts of the kind that go back to Weyl's *Das Kontinuum* and beyond: the classical notion of arbitrary infinite subsets of the naturals is incomprehensible – or so the sceptical claim goes – and that forces us to be more careful about which sets we *do* countenance. However, the reconstructive project remains interesting even if we don't have a sceptical agenda. Even if we are cheerfully maximalist about which sets of numbers there eventually are, it is still of great interest to seek to *calibrate* theorems of real analysis by seeing which assumptions about sets are needed to prove them in a second-order framework, working up through stronger and stronger assumptions to see what becomes provable at each different level. We'll be particularly interested to map the cases where going up a level means bringing into play new concepts. For a simple example, some classical theorems are constructively provable – so we'd expect that we can prove them in a second-order arithmetic which only countenances recursively decidable sets of numbers. Proving other theorems will take stronger set-existence assumptions, where endorsing the assumptions would require taking aboard non-constructive ideas.

The business of calibrating the strength of the assumptions which are actually needed to prove proxies for various classical theorems in second-order arithmetic is a core theme of Simpson's *SoSOA*. Simpson argues for the special status of five particular subsystems of second-order arithmetic – in order of increasing richness, these are RCA_0, WKL_0, ACA_0, ATR_0, and $\Pi^1_1\text{-}\mathsf{CA}_0$. And the fundamental claim is that most familiar classical theorems in fact partition into five classes, those can already be proved in the weakest system RCA_0, those that require WKL_0 to prove them, those that first become available at the level of ACA_0, and so on.

Now, the technical facts here are not up for dispute, of course. But their significance is a different matter. And one question to press concerns the real conceptual commitments of these five canonical subsystems. What new ideas are needed in advancing from one theory to the next? Simpson in fact makes some brief remarks about this, but – being no philosopher – he arguably gets things wrong. Thus, RCA_0, or a close relative, does seem to have some philosophical interest, but not the interest that Simpson claims for it (it isn't constructively acceptable). WKL_0 is conceptually very puzzling. And, my topic here, there are even issues about ACA_0, despite the fact that – at least prima facie – it appears to be an especially natural subsystem to consider, and moreover it arguably can cope with pretty much all the analysis we actually need for physics. (You can see, then, why *this* sort of subsystem of arithmetic is, or at least should be, of interest to those who like indispensability arguments: ACA_0 gives us more or less what's indispensable to science, and so will be – according to that familiar line of thought – at about the limit of what we need to accept as true as opposed to a mathematical fiction. From that perspective, its commitments and conceptual grounding will therefore be of particular importance.)

2 The form of a second-order arithmetic

a) Before introducing ACA_0, I need to say something – which will have to be very brief indeed – about the framework in Simpson's book: for him, a 'subsystem of second-order arithmetic' is an arithmetic framed in a two-sorted *first-order* language, the two sorts of quantifiers respectively running over numbers and sets of numbers, with a restricted set comprehension principle.

Put baldly like that, as in Simpson's own rather take-it-or-leave-it presentation, this might momentarily strike you as a puzzling characterization. For haven't we all learnt that the differences between a genuinely second-order theory and a (many-sorted) first-order theory run deep? But two reminders.

First, for our purposes, we can keep all second-order quantification monadic (because we will be dealing with arithmetics which can cope with an effective coding function between pairs and numbers, and we can think of dyadic numerical relations as properties of codes-for-pairs, and so on).

Second, we are interested here in *formalizable* arithmetics with a recursive notion of proof: so our monadic second-order quantifiers will naturally be thought of as governed by analogues of the rules for the first-order quantifiers, though of course the resulting axiomatized logic will not be complete for 'full' second-order consequence – see e.g. (Shapiro, 1991).

Given that the second-order part of our logic *is* monadic, it then need be no more than a matter of suggestive notational variation to rewrite second-order sentences using the symbolism for set membership rather than the symbolism for property application. In particular, we can write the induction axiom and the comprehension schema which tells as what sets there are as

$$\mathsf{Ind} \quad \forall \mathsf{X}((\mathsf{0} \in \mathsf{X} \wedge \forall \mathsf{x}(\mathsf{x} \in \mathsf{X} \rightarrow \mathsf{Sx} \in \mathsf{X})) \rightarrow \forall \mathsf{x}\, \mathsf{x} \in \mathsf{X}),$$

$$\mathsf{Comp} \quad \exists \mathsf{X} \forall \mathsf{x}(\mathsf{x} \in \mathsf{X} \leftrightarrow \varphi(\mathsf{x})).$$

where φ holds the place for some appropriate kind of open sentence. So now we indeed will have two sorts of *individual* quantifier in play, governed by the usual rules. (Note, though, that the easy transition to talk about sets leaves it open just how we are to interpret this talk ontologically. If you fuss about these things, you might want to press the case that we can and should move back to interpreting the second type of quantifier as a plural quantifier, and we needn't treat it as a running over sets thought of as individuals at all. But frankly, I think it is just distracting from the serious issues to fuss about these ontological niceties. What distinguishes the various subsystems of second-order arithmetic is the strength of their various infinitary assumptions – and whether we read those assumptions set-wise or plural-wise is not of the essence. So, if only because of the familiarity of this way of speaking, we'll stick here to set talk.)

With that – too sketchy though it really is – by way of motivation, let's introduce some terminology. L_2 will be a two-sorted first-order language for arithmetic, with $\mathsf{0}$ a constant of the first sort, $\mathsf{S}, +, \times$ as functions of the first sort, and $\in$ a relation between items of the first sort and the second sort. Then we'll say that a *neat* system of second-order arithmetic is an axiomatized formal theory built in the language L_2, with a standard two-sorted classical first-order logic (plus identity rules for the first sort), whose axioms are those of Robinson Arithmetic Q, plus the induction axiom Ind, plus all those instances of the comprehension schema Comp for wffs $\varphi(\mathsf{x})$ belonging to some specified class C – where it is decidable which wffs belong to C. Different neat theories are then determined by fixing on different classes C of permissible instantiations of the comprehension schema.

The *strongest* neat theory is of course the standard theory PA_2 (a.k.a. Z_2) – i.e. the usual formal axiomatization of second-order Peano Arithmetic – where we allow *any* $\varphi(\mathsf{x})$ belonging to our two-sorted arithmetical language to occur in the comprehension schema. And the various theories in Simpson's hierarchy are also either (equivalent to) weaker neat theories, or are such theories with additional axioms.

3 Introducing ACA_0

a) Unless we say otherwise, the schema Comp allows the substituted open sentence $\varphi(\mathsf{x})$ to embed universal set-quantifiers. And so, in general, instances of the comprehension schema can involve definitions of sets which require for their understanding a grasp of quantification over .... Well, over *what*?

We could say: quantification over *all* arbitrary subsets of the numbers – the appropriate answer, perhaps, if we are interpreting the strongest second-order theories. Yet, as I indicated before, we might well hesitate over the very idea of arbitrary infinite sets of numbers – sets which are supposedly perfectly determinate but are in the general case beyond any possibility of our specifying their members. We can, perhaps relatively uncontroversially, make sense of the membership of a set of numbers being determined by possession of some suitable characterizing property which gives a criterion – even if not an effective recipe – for picking out the numbers; and we can equally make sense of the membership being merely stipulated (more or less arbitrarily). However, the first idea gives us infinite sets but not arbitrary ones; and the second idea may give us arbitrary sets (whose members share nothing but the gerrymandered property of having being selected for membership) but not infinite ones – unless we are prepared to conceive of a completed infinite series of arbitrary choices.[1] Neither initial way of thinking of sets uncontentiously makes sense of the classical idea of arbitrary infinite sets of numbers.

But be that as it may: our interest now is going to be in subsystems of second-order arithmetic, so we are in any case considering taking a step back from purporting to be quantifying over all arbitrary subsets of the first-order domain. But then what are the options for understanding wffs involving set quantifiers? If we are putting on hold the thought that we understand the idea of a domain of *all* arbitrary sets of numbers, we can't help ourselves either to the idea of restricting *that* domain. So should we proceed, so to speak, from the bottom up? We might try saying: the set-quantifiers in an allowed instance of Comp quantify over just those kosher sets defined by instances of Comp. However, that looks rather uncomfortably impredicative: the totality of kosher sets will be defined in terms of quantifications over itself, and we are given no independent handle on the totality. So how then can we understand those embedded set-quantifiers?

Well, there's one obvious quick way to avoid any such tangles. *Simply require substitutions for $\varphi(\mathsf{x})$ in* Comp *not to have any embedded set-quantifiers.* If we do so, and otherwise cleave to neatness, then we get the arithmetic defined as follows:

> ACA_0 is the neat arithmetic whose axioms are those of Robinson Arithmetic Q, plus Ind, plus those instances of Comp where $\varphi(\mathsf{x})$ lacks bound set-variables.

We'll say that a theory like this which bans set-quantifiers from substitutions for $\varphi(\mathsf{x})$ in Comp has *Arithmetical Comprehension* – hence the label ACA_0 (I'll explain the significance of the subscript in a moment).

Note, however, that banning *bound* set-variables in $\varphi(\mathsf{x})$ still allows *free* set-variables to occur. (Or allows the presence of parameters formed from a different vocabulary than the variables, if that's the way we prefer to set up our logic.) We'd better pause next to explain why we should want to allow them.

[1] And surely, '[t]he notion that an infinite set is a "gathering" brought together by infinitely many individual arbitrary acts of selection, assembled and then surveyed as a whole by consciousness, is nonsensical.' (Weyl, 1918, Ch. 1, §4)

b) Given a choice between either (i) avoiding impredicativity by banning bound set-variables in $\varphi(\mathsf{x})$ or (ii) going the whole hog and banning *all* set-variables, bound or free, what principled reason is there for choosing between the resulting theories?

Well, let $\mathsf{ACA_{pf}}$ – 'pf' for parameter-free – be the theory which you get by indeed banning all set-variables from instances of $\varphi(\mathsf{x})$ in Comp; in other words, $\mathsf{ACA_{pf}}$ is the neat theory that requires instances of $\varphi(\mathsf{x})$ to be open wffs which also already belong to the language L_1 of first-order arithmetic. Now, endorsing theory $\mathsf{ACA_{pf}}$ requires us to hold that the open sentences of the L_1 do indeed express cogent conditions which determine sets as their extensions – 'arithmetical' sets, as we'll call them (otherwise, of course, we shouldn't be endorsing even the restricted comprehension principle saying that there are such sets). So, if Y is such an arithmetical set, then for any number x there is – according to us – a fact of the matter whether $x \in Y$. But in that case, an open sentence of the language of first-order arithmetic augmented with a way of saying that some number is in Y *also* expresses a fully determinate condition according to us. But then what conceptual reason can we possibly have for supposing that conditions $\varphi(\mathsf{x})$ expressed in the language of first-order arithmetic *do* determine sets as their extensions, while equally determinate conditions $\varphi(\mathsf{x}, \mathsf{Y})$ expressed in the augmented language which allows parameters acting as temporary names for those arithmetic sets *don't* have extensions? After all, if we arithmetically define a condition in terms of membership of some arithmetic set, then it will be equivalent to some purely arithmetical condition – so of course it has an extension!

'Hold on! You just said that if we accept arithmetical sets, we should equally accept sets defined with parameters *acting as temporary names for arithmetic sets*. Fair enough. But the specification of $\mathsf{ACA_0}$ allows the wffs substituted into the comprehension schema to have parameters – free set-variables – without any explicit restriction built in. So doesn't *that* overshoot?' No: it remains the case that the only sets that $\mathsf{ACA_0}$ actually knows about are the arithmetic sets, so from its perspective (so to speak) parameters will be temporary names for such sets. More formally, it is in fact not hard to show that the smallest model of $\mathsf{ACA_0}$ where the domain of the first sort of quantifier is the natural numbers has just the arithmetical sets as the domain of the second sort of quantifier.[2] So accepting $\mathsf{ACA_0}$ indeed commits us to no more than the arithmetical sets that we are committed to by $\mathsf{ACA_{pf}}$, just as we wanted.

Hence, if we are going to seriously countenance arithmetical sets at all and start quantifying over them, as in theory $\mathsf{ACA_{pf}}$, it seems we should also be prepared to countenance sets arithmetically defined in terms of particular arithmetical sets. And the natural formal reflection of this idea is to allow free set-variables/parameters in our comprehension principle after all, giving us $\mathsf{ACA_0}$.

Two quick comments about this. First, I'm not denying that there aren't technically interesting differences between $\mathsf{ACA_{pf}}$ and $\mathsf{ACA_0}$. But my claim is that there isn't a philosophically stable stance which would make $\mathsf{ACA_{pf}}$'s comprehension principle acceptable without making $\mathsf{ACA_0}$'s principle acceptable too.

Second, I have laboured this point about $\mathsf{ACA_0}$'s allowing set parameters in the comprehension principle because it usually seems to be passed over without real comment. There is, however, a relevant discussion by Azriel Levy, writing about the relation of VNB to ZF (see Fraenkel et al. 1973, ch. 3, §7, slightly reworked in Levy 1976).[3] – for the relation between ZF and VNB is very close to that between $\mathsf{PA_1}$ and $\mathsf{ACA_0}$. In each case a single-sorted first-order theory is extended by adding a second sort of variable, running over sets/classes, whose use is governed by a predicative comprehension principle Comp where substitutions for $\varphi(\mathsf{x})$ lack quantifiers of the second sort. Levy's central argument for allowing class parameters when substituting into Comp in the case of VNB turns out to be parallel to the argument for allowing set parameters into $\varphi(\mathsf{x})$ in the case of $\mathsf{ACA_0}$.

[2]For a proof, (see Simpson, 1991, p. 317, Corollary VIII.1.11).

[3]Tim Storer reminded me of this.

c) So, the claim is, ACA_0 seems to be a very natural stopping point on the retreat *downwards* from a commitment to impredicative versions of second-order arithmetic. But that's not the only way of looking at things. For ACA_0 can also be seen as an equally natural stopping point on the way *upwards* from first-order Peano Arithmetic.

Of course, that's more or less immediate if we are looking for a neat theory (in our official sense) which fully includes PA_1. For such a theory will have to be at least as strong as $\mathsf{ACA}_{\mathsf{pf}}$ if we are to be able to prove PA_1's instances of the first-order induction schema for arbitrary formulae of L_1; and we've already argued a theorist who is prepared to accept $\mathsf{ACA}_{\mathsf{pf}}$ should be happy with full ACA_0 too.

Still, it's worth proceeding a little more slowly here. Imagine again, then, that we are interested in the project of reconstructing a significant amount of the classical theory of the reals based on a theory of sets of numbers. What would be a *minimal* commitment to numerical sets, to get us started on that reconstructive project?

Well, if we countenance recursive sets of numbers, then that should set us en route giving a constructive theory of recursive reals. However, it is well known that constructive analysis must lack many of the characteristic theorems of classical analysis. So let's take the next step, of countenancing recursively enumerable – but perhaps non-recursive – sets of numbers. Indeed, I suppose one might argue that it is taking *this* step which first embroils us with something like the classical idea of 'actual' infinities. For the truths about recursive sets might plausibly be thought of as disguised truths about the finitary decision procedures for fixing their membership. On the other hand, accepting that there are non-recursive (even if r.e.) sets involves taking aboard the thought that there are infinite sets, facts about whose membership can transcend any possibility of discovery by finitary means.

Now, it is familiar that the r.e. sets are those which are the extensions of Σ_1 wffs of L_1. And a two-sorted arithmetic which embodies a commitment to r.e. sets can naturally be presented as one that restricts instances of comprehension to Σ^0_1 wffs of L_2 (where a Σ^0_1 wff is equivalent to a wff starting with one or more existential quantifiers of the first sort, followed by a kernel Δ^0_0 wff which contains no quantifiers of the second sort and at most bounded quantifiers of the first sort).

Call the neat theory with comprehension for Σ^0_1 wffs '$\Sigma\mathsf{CA}_0$'. Then, if we are going to tangle with the 'actual infinite' in a classical way at all, then $\Sigma\mathsf{CA}_0$ would seem nicely to formalize a minimal commitment. However, it is an immediate and easy result that $\Sigma\mathsf{CA}_0$ is actually equivalent to ACA_0. In other words, Σ^0_1 comprehension entails unrestricted arithmetic comprehension. The formal proof goes by showing that $\Sigma\mathsf{CA}_0$ plus Σ^0_k comprehension implies Σ^0_{k+1} comprehension, for $k \geqslant 1$, and then the result that $\Sigma\mathsf{CA}_0$ implies full arithmetic comprehension follows by induction.[4] Informally, the idea is that $\Sigma\mathsf{CA}_0$ gives us the computably enumerable Σ_1 sets; and then – since we now believe that there are determinate facts of the matter about the membership of those sets – we should believe in the Σ_2 sets which we can enumerate by a computable procedure that can call upon an oracle to tell us the facts about membership of the Σ_1 sets; then – since we now believe that facts about the membership of those Σ_2 sets are also determinate – we should believe in the Σ_3 sets which we can enumerate by a computable procedure that can call upon an oracle to tell us about the facts membership of the Σ_2 sets; and so on ever upwards.

[4]To establish the inductive step, remember that we can construct a primitive recursive coding function for pairs, which we can represent in first-order arithmetic. So write $[\mathsf{x},\mathsf{y}]$ for a wff that represents the p.r. function $[x, y]$ that maps x, y to code number for the ordered pair $\langle x, y\rangle$. Now assume Σ^0_k comprehension and suppose that $\varphi(\mathsf{x})$ is Σ^0_{k+1}. Then $\varphi(\mathsf{x})$ can be written as $\exists \mathsf{y}\psi(\mathsf{x},\mathsf{y})$, where ψ is Π^0_k. $\neg\psi$ is then Σ^0_k so we can apply Σ^0_k comprehension to yield the set Y of all codes $[\mathsf{x},\mathsf{y}]$ such that $\neg\psi(\mathsf{x},\mathsf{y})$. Then by Σ^0_1 comprehension, we can get the set X such that $\mathsf{x} \in \mathsf{X} \leftrightarrow \exists \mathsf{y}([\mathsf{x},\mathsf{y}] \notin \mathsf{Y})$. Plainly, $\mathsf{x} \in \mathsf{X}$ iff $\exists \mathsf{y}\psi(\mathsf{x},\mathsf{y})$ iff $\varphi(\mathsf{x})$. So we have Σ^0_{k+1} comprehension.

d) So, putting the downward and upward considerations together, they suggest then that there is indeed a certain prima facie philosophical naturalness about ACA_0. It is the neat system of arithmetic that could, it seems, very well appeal if you accept PA_1, and on the one hand think you can make sense of the idea of some infinite sets of numbers – namely, the sets given by the (broadly) logical idea of the extensions of first-order open wffs – but on the other hand you don't (yet) want to commit to the idea of arbitrary unspecifiable subsets of the natural numbers or to run into tangles of impredicativity. And I take it that this combination of ideas at least gets to the starting line as a conceptually cogent position. (But we'll have ask in in a moment whether ACA_0 is the ideally *best* way of reflecting these ideas in a formal system: that, indeed, is the principal topic of this paper!)

e) Finally before pressing on, let's note a few fundamental facts about ACA_0.

Theorem 1 ACA_0 *is conservative over* PA_1*: i.e. if* φ *is a sentence of* L_1 *such that* $\mathsf{ACA}_0 \vdash \varphi$*, then* $\mathsf{PA}_1 \vdash \varphi$.

Theorem 2 *Unlike* PA_1, ACA_0 *is finitely axiomatizable.*

Theorem 3 ACA_0 *exhibits exponential speed-up over* PA_1.

The first of these is perhaps what you already expect. You might say, by way of motivation, 'What ACA_0 adds to PA_1 is just an explicit commitment to the logical extensions that can already be picked out by wffs of PA_1's language L_1. Why should going explicit about those sets change what we can prove about the numbers themselves?' But that's dangerous talk, of a kind that can lead us astray! Still, conservativeness is indeed easily proved. And our second and third theorem again are probably expected, given the parallel we noted between ACA_0 vs. PA_1 and VNB vs. ZF.

Now, those three results are mildly interesting: but ACA_0 *matters* because of results like the next three. Given definitions of the reals as Cauchy sequences, and decently natural related definitions,

Theorem 4 ACA_0 *can prove that polynomials, exponentials, trig functions, etc. in reals are continuous functions, and that their differentials are as expected.*

Theorem 5 ACA_0 *can prove that the Riemann integral* $\int_a^b \varphi(x)dx$ *exists, for* $\varphi(x)$ *continuous in* $a \leqslant x \leqslant b$*, and that it has the expected properties for familiar functions.*

We also have a more complex but crucial result, about the solution of ordinary differential equations:

Theorem 6 *Given the initial value problem*

$$x' = f(x,t), \quad x(0) = 0$$

where x *is a function of* t*,* x' *is the differential with respect to* t*, and* f *is continuous through a neighbourhood of* $(0,0)$ *then there is a neighbourhood of that point and a function* φ *such that* $x = \varphi(t)$ *is a differentiable solution in that neighbourhood: and that's provable in* ACA_0.

I'm not even going to gesture towards proofs of these. But of course, they are typically proved by taking appropriate instances of comprehension; instantiating with respect to set parameters new to the argument; combining with parametric instances of induction, and proceeding from there. Of course, it is the availability of such results, showing that much classical analysis can be reconstructed even in the weak predicative system ACA_0, which makes its exploration all worth while.

4 ACA_0 vs ACA – and ACA!

a) ACA_0 is a neat subsystem of full second-order arithmetic ('neat' in our official sense). We've seen that the theory does indeed seem to have some prima facie conceptual attractions. We've claimed that it can indeed be used to reconstruct swathes of classical applicable analysis. So far, then, so very good.

However, there remains a lurking conceptual problem, that seems to have gone unnoticed. And one useful way to throw light on the key issue here is to consider the relation between ACA_0 and a pair of alternative theories which also have Arithmetic Comprehension: the constrast will prove very instructive.

Simpson says that the subscript '0' in the label 'ACA_0' indicates *restricted induction.* But 'restricted' compared with what? He doesn't spell this out in *SoSOA*, but the answer is: compared with ACA. This is the theory you get by starting from ACA_0, but then replacing the single induction axiom Ind with (the closures of) all instances of the induction schema

$$\{\varphi(0) \land \forall x(\varphi(x) \to \varphi(Sx))\} \to \forall x\varphi(x),$$

where φ can be an arbitrarily complex L_2 wff, i.e. can contain set-quantifiers ad libitum.[5]

Now, allowing unrestricted instances of the L_2 induction schema is evidently equivalent to allowing unrestricted use of the corresponding inductive inference rule:

$$\frac{\varphi(0) \qquad \forall x(\varphi(x) \to \varphi(Sx))}{\forall x\varphi(x)}$$

Indeed, in some presentations the whole discussion basically proceeds in terms of systems of arithmetic with such an inductive inference rule rather than with induction axioms. But inference *rules* are normally taken, by default, to apply theory-wide; and so ACA seems on such an approach to become the obvious theory of choice with Arithmetical Comprehension (and theories with restricted induction like ACA_0 – if visible at all – will look to be artificially hobbled). Given the apparent naturalness of treating induction as a rule, should we then – despite the considerations in the previous section – in fact prefer ACA to ACA_0?

b) What's to chose? Trivally, ACA extends ACA_0. Less trivially, ACA proves more purely arithmetical sentences than ACA_0; in other words, while ACA_0 is conservative over PA_1 for L_1 wffs, ACA isn't. Just for a start, ACA can prove $\mathsf{Con}(\mathsf{PA}_1)$.

Why so? Well, the short answer is that the numerical property of Gödel-numbering a true closed wff φ of L_1 can be defined in second-order terms. We know from Tarski's theorem about the undefinability of truth that there is no purely arithmetic wff, i.e. no first-order L_1 wff, which is a truth-predicate for L_1 wffs. But there *is* an L_2 wff $\mathsf{T}(x)$ such that, provably in ACA, for each L_1 sentence φ, $\mathsf{T}(\ulcorner\varphi\urcorner) \leftrightarrow \varphi$ (where, as usual, $\ulcorner\varphi\urcorner$ is the numeral for a Gödel coding for φ). In fact, a Σ^1_1 wff suffices, i.e. the predicate $\mathsf{T}(x)$ can take the form of an existential set-quantifier followed by a kernel wff involving only numerical quantifiers. And given its distinctive liberality with induction, ACA can prove general results about the applicability of this Σ^1_1 predicate. In particular, it can invoke the usual type of induction over the length of proofs, this time applied to $\mathsf{T}(x)$, to show that all closed theorems of PA_1 are true, and hence establish that PA_1 is consistent.[6]

[5] This notational convention is in fact adopted more generally. Thus a second-order arithmetic T_0 – labelled with the subscript – is a neat theory (or at least a theory with a neat core) which deals with induction via the induction axiom Ind. T is then the corresponding theory which instead allows all instances of the induction schema.

[6] We don't need to delve too far into the technical details here. But perhaps it is worth saying just a little more.

We can start by defining an L_2 predicate $\mathsf{P}(X, n)$ which in effect says that X is the set of (Gödel

c) So technically the theories ACA_0 and ACA peel apart and the latter theory is arithmetically stronger.[7] But which theory is *conceptually* the more attractive one?

You might ask: what philosophical stance could make it a stably motivated position to be simultaneously mean with comprehension (only stipulating that arithmetical predicates determine sets as their extensions) yet very generous with induction (allowing that any L_2 predicate, even though it might not officially determine some property-in-extension, can still be sufficiently kosher to use in cogent inductive arguments)? But in fact, there does seem to be a knockdown argument *in favour* of ACA's generosity with instances of the induction schema (or equivalently, in favour of using the induction rule unrestrictedly).

> Instances of the induction schema are *conditionals*:
>
> $$\{\varphi(0) \wedge \forall x(\varphi(x) \to \varphi(Sx))\} \to \forall x\varphi(x)$$
>
> So they actually only allow us to derive some conclusion $\forall x\varphi(x)$ when we have *already* established the corresponding premiss (i) $\varphi(0)$ and can also establish (ii) $\forall x(\varphi(x) \to \varphi(Sx))$. But if we can already establish (i) and (ii) then trivially we can (iii) case by case derive each and every one of $\varphi(0)$, $\varphi(S0)$, $\varphi(SS0)$, ... However, there are no 'stray' numbers which aren't denoted by some numeral; so that means (iv) that we can show of each and every number that φ holds of it. But then what more can it possibly take for φ to express a genuine property that indeed holds for every number, so that (v) $\forall x\varphi(x)$ is true? In sum, it seems that we can't possibly overshoot by allowing instances of the Induction Schema for *any* open wff φ of L_2 with one free variable. The only *usable* instances from our generous range of axioms will be those where we can in fact establish the antecedents (i) and (ii) of the relevant conditionals – and in those cases, we can't go wrong in accepting the consequents (v) too.

Which takes us back to the earlier thought that, rather that ACA overshooting because its induction axioms outrun the instances of comprehension, it is ACA_0 which is more problematic because it unnaturally hobbles the scope of induction.

d) We first introduced ACA_0 as a theory which just makes an explicit commitment to the properties-in-extension that can already be expressed by open wffs of PA_1's language L_1 (so no wonder, perhaps, that it is conservative over PA_1). Yet now we seem to have a compelling argument for inflating the inductive commitments of the theory, an argument

---

numbers for) true L_1 sentences involving no more than n logical operators. So there will be a clause in the definition that in effect says that for any atomic sentence of the form $t_1 = t_2$ for closed terms t_1, t_2, $\ulcorner t_1 = t_2 \urcorner \in X \leftrightarrow t_1 = t_2$. Then there will be clauses which say e.g. that so long as $(\alpha \wedge \beta)$ has no more than n logical operators, then $\ulcorner(\alpha \wedge \beta)\urcorner \in X$ iff $\ulcorner\alpha\urcorner \in X$ and $\ulcorner\beta\urcorner \in X$; and (roughly) for any wff like $\forall x\varphi(x)$ with no more than no more than n logical operators, $\ulcorner\forall x\varphi(x)\urcorner \in X$ iff for all m $\ulcorner\varphi(\overline{m})\urcorner \in X$ (where $\overline{m}$ is the numeral for m). Such clauses do not involve set-quantifiers, so the partial truth-predicate $P(X, n)$ involves no more than numerical quantifiers.

Then, the key step, we can define a full L_2 truth-predicate which applies to (codes for) L_1 wffs: $T(x) =_{\text{def}} \exists X \exists n(x \in X \wedge P(X, n))$. This truth-predicate is evidently Σ^1_1. (For more details, see Takeuti 1987, pp. 184–187.)

Now, with $T(x)$ thus defined in second-order terms, even ACA_0 can prove, case by case, each instance of $T(\ulcorner\varphi\urcorner) \leftrightarrow \varphi$, where φ is a sentence of L_1. This means, in particular, that ACA_0 can prove $T(\ulcorner\varphi\urcorner)$ for each axiom of PA_1. ACA_0 can also show case by case e.g. that if $T(\ulcorner\varphi\urcorner)$ and $T(\ulcorner\varphi \to \psi\urcorner)$ then $T(\psi)$. So, case by case, ACA_0 knows that the axioms of PA_1 are true, and that the application of a rule of inference preserves truth. But ACA_0, lacking induction for Σ^1_1 predicates, can't generalize on these claims. Hence this weaker theory *can't* prove consistency by explicitly showing that *all* PA_1 theorems are true.

[7] In fact, the difference is quite considerable. To repeat, ACA_0 is conservative over PA_1; but ACA can not only prove more arithmetical claims than PA_1, but also prove more arithmetical claims than $\mathsf{PA}_1 + \mathsf{Con}(\mathsf{PA}_1)$, or $\mathsf{PA}_1 + \mathsf{Con}(\mathsf{PA}_1) + \mathsf{Con}(\mathsf{PA}_1 + \mathsf{Con}(\mathsf{PA}_1))$, etc. And while the proof-theoretic ordinal of ACA_0 is just that of PA_1, i.e. ε_0, the proof-theoretic ordinal of ACA is much bigger, $\varepsilon_{\varepsilon_0}$.

which pushes us on from accepting $\mathsf{ACA_0}$ to accepting the much richer ACA. Can we resist the argument?

I think that the blunt answer is 'no'. In other words, if you do *fully* buy $\mathsf{ACA_0}$, the inflationary argument is compelling. Why so? Well, suppose $\varphi(\mathsf{x})$ is an L_2 open wff with just 'x' free (but which may embed quantifiers of the second type). Then, given that $\mathsf{ACA_0}$ applies first-order logic across the board and hence to any such open wff, we trivially have $\mathsf{ACA_0} \vdash \forall\mathsf{x}(\varphi(\mathsf{x}) \vee \neg\varphi(\mathsf{x}))$. So, as far as $\mathsf{ACA_0}$ is concerned, given the classical interpretation of the connectives, any number either determinately satisfies or doesn't satisfy the condition $\varphi(\mathsf{x})$. You can't fully endorse $\mathsf{ACA_0}$ without allowing the cogency of such conditions (however many set-quantifiers they may embed). But if $\varphi(\mathsf{x})$ is indeed determinately possessed or lacked by any number, what principled reason could there be for refusing to accept inductive arguments using $\varphi(\mathsf{x})$, given our argument above? There isn't one.

Still, while that might be the blunt answer, there's evidently more to be said. For what is clearly emerging is that $\mathsf{ACA_0}$ somewhat over-generates: in allowing it to form arbitrary open sentences embedding set quantifiers and giving it a full two-sorted first-order logic, we have rounded out the theory in a way that is technically neat but which arguably encodes commitments beyond those shaping the original guiding conception that the theory was officially aiming to capture. So let's go back to the drawing board again, run through the construction of $\mathsf{ACA_0}$ a bit more carefully, and see where it in fact runs beyond the core idea of a theory with Arithmetic Comprehension and the natural second-order induction principle.

e) To dramatize the story, this time let's imagine trying to take along with us a theorist Kurt, who has a taste for formalizing his knowledge of arithmetic and who accepts $\mathsf{PA_1}$, but who – given worries about predicativity and about the very idea of arbitrary infinite sets – is properly cautious about accepting claims about numerical sets.

Despite his caution, however, even before we do get him to accept the existence of any particular sets, Kurt should be willing at least to grant a promisory note to apply induction to whatever particular sets there indeed are. So let X temporarily name some kosher set: then, whichever set that is, Kurt should agree to grant

$$(\mathsf{0} \in \mathsf{X} \wedge \forall\mathsf{x}(\mathsf{x} \in \mathsf{X} \to \mathsf{Sx} \in \mathsf{X})) \to \forall\mathsf{x}\,\mathsf{x} \in \mathsf{X}$$

At this point, though, note that Kurt hasn't *any* clear idea of what sets of numbers there are, so he certainly hasn't yet fixed the domain of possible values for X. In one sense, then, he doesn't yet understand propositions starting with a universal set-quantifier. So cautious Kurt, let's suppose, balks at endorsing the quantified induction axiom Ind. (Compare, if you like, the Hilbertian finitist who allows parametric instances of first-order claims, but balks at propositions starting with a universal number-quantifier.)

Second, now suppose for the sake of argument we can also get Kurt to buy at least the arithmetical sets (for he accepts L_1 open wffs express determinate conditions, and such determinate conditions – do they not? – have logical extensions). So, at least when $\varphi(\mathsf{x})$ lacks those still-not-fully-understood set quantifiers, Kurt is happy to introduce a new set-parameter Y into his reasoning, where $\forall\mathsf{x}(\mathsf{x} \in \mathsf{Y} \leftrightarrow \varphi(\mathsf{x}))$. Further, by the same arguments as before, if $\varphi(\mathsf{x}, \mathsf{Y})$ lacks set-quantifiers but already involves a parameter Y picking out a kosher set, then Kurt should still be happy enough to accept that there is a set of things that satisfy *this* condition φ which is arithmetically-definable-by-reference-to-some-other-kosher-set. In other words, we can reasonably suppose that Kurt's version of comprehension – like $\mathsf{ACA_0}$'s official version – can also allow set-parameters to appear in the relevant $\varphi(\mathsf{x})$.

Now, we *could* leave things there, with Kurt eschewing all set-quantifications. But equally he could, taking on no further commitments, countenance sentences starting with initial existential set-quantifiers, albeit now treated as pseudo-propositions ('pseudo' because they do not have contentful negations for him). Then he could retain the *form* of $\mathsf{ACA_0}$'s comprehension schema, $\exists X \forall x(x \in X \leftrightarrow \varphi(x))$, and he would argue from its instances using the usual elimination rule for dealing with existential set-quantifiers, and he could also have a corresponding introduction rule for prenex existentials (generalizing on some parameter that temporarily names a kosher set for him). Of course, this game-playing with existentials still doesn't give Kurt any fix on what counts as the totality of sets of numbers, so he still doesn't grasp the content of *universal* quantifications over sets of numbers: so, you might well say, he cannot fully grasp the content of genuine existentials either, and his play with the prenex existential form is mere show.

But still, for convenience, let's allow Kurt the latter device of using (pseudo)existentials to keep his practice looking close to that of the $\mathsf{ACA_0}$ theorist. But now note that, whatever we can directly derive in $\mathsf{ACA_0}$ by combining instantiations of the quantified induction axiom Ind with instantiations of the comprehension schema Comp, Kurt can now show using comprehension and his parametric instances of induction (while still not grasping the content of universal set-quantifications). In particular, Kurt can follow through parallel derivations of chunks of classical analysis inside $\mathsf{ACA_0}$; it's just that he sticks with parametric claims instead of ever concluding with universal set-quantifications, and though he will echo the words of prenex existential claims he (so to speak) crosses his fingers as he speaks.

To put that more formally, consider the restricted version of $\mathsf{ACA_0}$ which (i) bans set-quantifiers from appearing in wffs except as prenex existential quantifiers, (ii) has parametric induction and Arithmetic Comprehension, and (iii) has the obvious rules for dealing with the prenex existential set-quantifiers. Call this restricted theory '$\mathsf{ACA!}$'. *Then Kurt, so far, can be seen as working in* $\mathsf{ACA!}$, *where this theory is as competent as* $\mathsf{ACA_0}$ *at least at proving interesting proxies for classical results.*

f) Now, formally, the extra step in going from the more restricted $\mathsf{ACA!}$ to the original version of $\mathsf{ACA_0}$ is the step of allowing expressions containing set-quantifiers quite unrestrictedly to count as well-formed. This liberality doesn't of course buy us anything very interesting inside the official boundaries of $\mathsf{ACA_0}$, *since we still aren't allowed use these new expressions in instances of the comprehension schema.* And note, Kurt hasn't yet been given any general way of understanding such wffs with set-quantifiers, including universals.[8] Adapting a remark of Michael Dummett's,

> To give sense to the term 'set of numbers' – a sense adequate for its use in universal generalizations about sets of numbers – more is required than that Kurt knows what a set of numbers is and has a sharp criterion for what it is to count as one such set. He also needs to 'grasp' the domain, i.e. the totality of objects to which the term 'set of numbers' applies, in the sense of being able to say, in general, what sets of numbers there are. (Compare Dummett, 1993, p. 438.)

Without such a grasp, Kurt will rightly balk at going the extra step and making claims involving set-quantifiers (or at least, those which aren't equivalent to wffs with prenex existential set-quantifiers).

In sum, then, the situation is this. The conceptual commitments which previously we supposed to underpin $\mathsf{ACA_0}$ in fact only take our cautious theorist Kurt as far as

[8] Look at it this way. Kurt has agreed that at least the extensions of purely arithmetical predicates are boojums. But *that* of course doesn't give him a grasp of 'All boojums are . . . ' or 'Some boojums are . . . '. And we are assuming that Kurt isn't happy with e.g. the idea of boojums 'constructed' as it were by an infinite sequence of arbitrary choices. So as yet he stands in need of an explanation of what universal quantification over boojums amounts to. The point here is an elementary one, not to be confused with e.g. more sophisticated Dummettian worries about indefinite extensibility.

ACA! – though that is as far as he needs to go in order to have a theory which does as well as ACA_0 as far as deriving proxies for theorems of classical analysis is concerned. The extra step which would take Kurt from ACA! to ACA_0 requires a further conceptual commitment to the idea that the totality of numerical sets is determinate, and hence to the idea that the content of quantifications over this totality are in general well-defined. Now, if Kurt does take that step, then he can endorse 'idle' implications of ACA_0 such as an arbitrary instance of $\forall \mathsf{x}(\varphi(\mathsf{x}) \vee \neg\varphi(\mathsf{x}))$ where φ embeds set-quantifiers – but, we argued, if he *does* accept that, and so takes φ to express a determinate condition, Kurt has no reason to resist applying induction to a wff like φ and advancing to accept the theory ACA. (As far as considerations of conceptual grounding are concerned, ACA_0 is thus unstably poised between ACA! and ACA.)

Given, though, that Kurt *doesn't* take the extra conceptual step beyond what commits him to ACA!, what should his attitude be to instances of induction that deploy arbitrary predicates that embed set-quantifiers? Well, he needn't positively reject them as being *false*, i.e. as having true antecedents and false consequents. For indeed how could he that? – if the antecedents of an instance are true, then as we argued the consequent is true too. Rather, the point is that Kurt doesn't have a grasp of what it takes for wffs involving set-quantifiers to be true in general; so he isn't in general in a position to grasp instances of induction involving such wffs; and in particular he isn't in a position to establish the antecedents of such an instance of induction. For example, suppose φ is Σ^1_1 (cf. the truth-predicate introduced in fn. 6). Then Kurt won't be in a position to establish the proposition $\forall \mathsf{x}(\varphi(\mathsf{x}) \to \varphi(\mathsf{Sx}))$ needed if he is to make use of the relevant instance of induction, for that generalization is equivalent to one with a prenex universal set-quantifier. In short, Kurt won't find himself entitled to get to the starting line for understanding and using an instance of induction beyond those available in ACA! (and indeed in official ACA_0). And *that* is why his position then doesn't inductively inflate.

5 On Isaacson's Thesis

What I've been discussing here is, of course, intimately related to an issue about Daniel Isaacson's Thesis that PA_1 marks a boundary to the set of truths expressible in L_1 that can be established using purely arithmetical ideas plus logic – see (Isaacson, 1987).

I have written elsewhere about the core Thesis, separating it from more contentious claims that Isaacson makes, and offering some defence (Smith, 2007, pp. 204–211): I won't repeat those discussions here, but take it as read that the Thesis is important, interesting, and plausible. The issue now is that, at first blush, it might seem that we can get to accept ACA_0 by bringing together purely arithmetical ideas with a logical conception of determinate arithmetical conditions as having extensions, extensions that are available to be talked about and generalized over. So, ACA_0 should, it might seem, be within the grasp of the logically reflective arithmetician. And that is, so far, a thought that is in harmony with Isaacson's Thesis, since ACA_0 as we noted is conservative over first-order PA_1.

But now the inflationary argument threatens the Thesis. For the argument presses someone who has *fully* bought ACA_0 to allow induction over any determinate condition that ACA_0 recognizes (whether or not ACA_0 itself asserts that that condition has an extension) – which means allowing induction over any property expressible in L_2. And *that* takes us to acceptance of full ACA, which as we have noted is no longer conservative over PA_1.

If Isaacson is to hold the line and defend his Thesis, he therefore has two options. Either could argue that, while ACA_0 is within the grasp of the logically reflective arithmetician, the argument for inductive inflation can be resisted: but that's a very hard line to defend, I'd suggest, giving the compelling argument for generosity with induction. Or he has to say that, on second thoughts, the cautious logically reflective arithmetician (our Kurt) is at most entitled to is a weaker theory like ACA!, which doesn't inflate.

It will clear from the preceding section that I'm recommending the second option as giving a defensible way out for Isaacson to protect the Thesis.

6 PA_1 and truth

(a) It will, perhaps, through some more light tangentially across the issue that I've been talking about if we turn to consider an analogous case – though in fact it will turn out that the relation between our cases is closer than mere analogy.

Start again with the theory PA_1, and this time take some sensible scheme of Gödel-coding. Relative to that coding, we can now arithmetize the syntax of PA_1 in very familiar ways. For a start, we can define predicates of L_1, At and Sent, such that $\mathsf{At}(\ulcorner\varphi\urcorner)$ and $\mathsf{Sent}(\ulcorner\varphi\urcorner)$ hold just when φ an atomic sentence or a sentence of L_1 respectively.

Next, there's an arithmetical relation that holds between two numbers x and y just when, for some sentence φ, $x = \ulcorner\varphi\urcorner$ and y is the unique number such that $y = \ulcorner\neg\varphi\urcorner$. Again, we can express this functional relation using a relational expression $\mathsf{Neg}(\mathsf{x},\mathsf{y})$ in L_1. But in fact it is just a little more elegant to extend L_1 by adding a new function expression $\dot{\neg}$, and we can then conservatively extend PA_1 by adding the corresponding axiom $\dot{\neg}(\mathsf{x}) = \mathsf{y} \leftrightarrow \mathsf{Neg}(\mathsf{x},\mathsf{y})$. Then we'll be able to prove any instance of $\dot{\neg}(\ulcorner\varphi\urcorner) = \ulcorner\neg\varphi\urcorner$.

Similarly – cutting through the intermediate stages – we can conservatively introduce a new function expression '$\dot{\wedge}$' such that $\dot{\wedge}(\ulcorner\varphi\urcorner,\ulcorner\psi\urcorner) = \ulcorner(\varphi\wedge\psi)\urcorner$, and also another function expression '$\dot{\forall}$' such that $\dot{\forall}(\mathsf{j},\ulcorner\varphi\urcorner) = \ulcorner\forall \mathsf{v}_\mathsf{j}\varphi\urcorner$ (where v_j is the j-th variable of L_1 is some canonical listing).

Finally, there is a another primitive recursive functional relation $Sub(j,k,x,y)$ which holds when y is the Gödel code of the result of substituting the numeral for k in place of any occurrences of the j-th variable of L_1 in the wff with code x. Again we can conservatively add a corresponding functional expression $\mathsf{sub}(\mathsf{j},\mathsf{k},\mathsf{x})$ to PA_1. Suppose that the third variable in the listing is z: then, for example, we'll now be able to prove any instance of $\mathsf{sub}(\overline{3},\overline{5},\ulcorner\varphi(\mathsf{z})\urcorner) = \ulcorner\varphi(\overline{5})\urcorner$.

b) So much for arithmetizing syntax; now let's consider the structure of an arithmetized theory of truth for PA_1 using that syntactic apparatus. Since it is primitive-recursively decidable whether an atomic sentence of L_1 – i.e. a quantifier-free equation – is true, there is also a numerical predicate T_{at} constructible in L_1 such that $\mathsf{T}_{\mathsf{at}}(\ulcorner\varphi\urcorner)$ holds just when φ is a true atomic sentence.

But, by Tarski's theorem again, we can't define a generally applicable truth-predicate in L_1. So, to proceed further, we now need to augment PA_1 – already conservatively extended by adding those function expressions for arithmetizing the syntactic operations, and appending the natural axioms for them – in two ways. (To fend off possible confusion, then, what we are talking about here is something quite different from the *explicit* definition of truth in *second-order* terms that was in question in fn. 6 above. *Here* we are talking about adding axioms to *implicitly* define a new predicate added to the *first-order* language L_1.)

First, we expand PA_1's language again, this time with a new monadic predicate T, to get the language L_1^+. Then second, we add a set of axioms Th which tell us how to work out the applicability of T, Tarski-style. So the axioms will tell us that the atomic sentences are true as you'd expect, a negated sentence is true when the unnegated original isn't, a conjunction is true just when both conjuncts are true, and a universal quantification $\forall\xi\varphi(\xi)$ is true just when every instance $\varphi(\overline{\mathsf{k}})$ is true. Putting that more formally, we'll have the axioms Th:

1. $\forall \mathsf{x}(\mathsf{At}(\mathsf{x}) \rightarrow (\mathsf{T}(\mathsf{x}) \leftrightarrow \mathsf{T_{at}}(\mathsf{x}))$

2. $\forall \mathsf{x}(\mathsf{Sent}(\mathsf{x}) \rightarrow (\mathsf{T}(\dot{\neg}(\mathsf{x})) \leftrightarrow \neg \mathsf{T}(\mathsf{x}))$

3. $\forall \mathsf{x} \forall \mathsf{y}(\mathsf{Sent}(\dot{\wedge}(\mathsf{x},\mathsf{y}))) \rightarrow (\mathsf{T}(\dot{\wedge}(\mathsf{x},\mathsf{y})) \leftrightarrow (\mathsf{T}(\mathsf{x}) \wedge \mathsf{T}(\mathsf{y})))$

4. $\forall \mathsf{x} \forall \mathsf{j}(\mathsf{Sent}(\dot{\forall}(\mathsf{j},\mathsf{x})) \rightarrow (\mathsf{T}(\dot{\forall}(\mathsf{j},\mathsf{x})) \leftrightarrow \forall \mathsf{k}\, \mathsf{T}(\mathsf{sub}(\mathsf{j},\mathsf{k},\mathsf{x})))$

(and similarly, of course, for any other connectives and quantifiers we take as primitive in L_1). So what we've done here is to bolt on to $\mathsf{PA_1}$ the arithmetization of one natural theory of truth for L_1, in a way that – prima facie – shouldn't upset even a deflationist/minimalist about truth. As we'd expect, the composite theory $\mathsf{PA_1} \cup \mathsf{Th}$ proves each biconditional $\mathsf{T}(\ulcorner\varphi\urcorner) \leftrightarrow \varphi$ for sentences φ. And as we'd also expect – though it isn't entirely trivial to demonstrate it – $\mathsf{PA_1} \cup \mathsf{Th}$ is conservative over $\mathsf{PA_1}$: it proves no new arithmetical truths (see (Halbach, 1999b, p. 186), (Halbach, 1999a, Theorem 3.1)).

c) But now let's reflect. Merely bolting the arithmetical truth-theory onto $\mathsf{PA_1}$ as it were 'externally' leaves us, in particular, with just the same induction axioms as we started off with. However, you might well say, why stop there? It would seem that the same line of argument that we sketched above for being generous with induction will apply again, and will motivate extending the induction axioms from instances involving just the original L_1 predicates to instances involving the new predicate T too, i.e. instances involving L_1^+ predicates. So let $\mathsf{T(PA)}$ be the theory we get taking $\mathsf{PA_1}$ plus Th plus the closures of all instances of the first-order induction schema *for the new predicates that can be formed in* L_1^+. Then, the argument seems to be, $\mathsf{T(PA)}$ should be as compelling a theory as $\mathsf{PA_1} \cup \mathsf{Th}$.

However a hard-core deflationist about truth who accepts $\mathsf{PA_1}$, and as yet thinks he can happily endorse $\mathsf{PA_1} \cup \mathsf{Th}$, will certainly balk at the idea that he must willy-nilly be driven to extend his commitments $\mathsf{T(PA)}$ (see for example (Field, 1999)). For – by his lights – the arithmetization of a mere theory of truth (commenting from the sidelines, as it were, usefully enabling him e.g. to endorse sentences in bulk, but having no more substantive content) had better not enable us to prove new arithmetical truths (had better not affect what is happening on the arithmetical field of play). But as is well known, $\mathsf{T(PA)}$ is no longer conservative over $\mathsf{PA_1}$: for a start, $\mathsf{T(PA)} \vdash \mathsf{Con}(\mathsf{PA_1})$. (The proof-strategy for deriving $\mathsf{Con}(\mathsf{PA_1})$ is the obvious one: $\mathsf{T(PA)}$ has the generalizing resources to prove that all $\mathsf{PA_1}$'s axioms are true, that inference preserves truth and hence all theorems are true. And proving $\mathsf{Con}(\mathsf{PA_1})$ *is* only the start.)

Yet, given the key argument, how *can* anyone resist inflating $\mathsf{PA_1} \cup \mathsf{Th}$ to $\mathsf{T(PA)}$? Perhaps our imagined deflationist might be tempted to say: 'As a deflationist, I don't accept that truth is a genuine property. Correspondingly, "T", the arithmetization of the truth-predicate, doesn't express a genuine numerical property. So we can't use it in inductive arguments.' But *this* is just arm-waving, unless augmented by an independent account of the otherwise decidedly murky notion of 'not expressing a genuine property'. We evidently need to proceed a bit more carefully.

d) Let's consider again our cautious theorist Kurt who accepts $\mathsf{PA_1}$. Suppose we now offer him the axioms Th as a partial characterization of the uninterpreted new predicate T we are inviting him to add to his language (recall that 'At', 'Sent' are just complicated L_1 predicates, and the defined functions $\dot{\neg}$, $\dot{\wedge}$, $\dot{\forall}$ and sub are all definitionally introduced via complicated L_1 relations, so ex hypothesi he can understand all the purely arithmetical constructions that occur in those axioms). If we give Kurt any particular number which happens to be the Gödel number of an L_1 sentence φ, he will then in principle be able to prove the corresponding theorem $\mathsf{T}(\ulcorner\varphi\urcorner) \leftrightarrow \varphi$. But if we give Kurt some number which *isn't* the Gödel number of a sentence φ, then the axioms don't tell him even in principle whether the predicate T applies or not – for the axioms so far do not speak to that case at all. The axioms for T thus, in a rather straightforward way, so far fail to pin down a determinate property.

Now, there's an obvious repair for the obvious failing. We can deal with the 'waste' cases by simply adding a new axiom

5. $\forall \mathsf{x}(\neg \mathsf{Sent}(\mathsf{x}) \to (\mathsf{T}(\mathsf{x}) \leftrightarrow \bot))$.

But this only takes Kurt so far. It is now true that, given any particular number n, then if $n = \ulcorner\varphi\urcorner$ for some sentence φ, Kurt can prove $\mathsf{T}(\overline{\mathsf{n}}) \leftrightarrow \varphi$, or else Kurt can prove $\mathsf{T}(\overline{\mathsf{n}}) \leftrightarrow \bot$. What he still *can't* do, since he doesn't yet have induction axioms for T, is prove anything *general* about T to the effect that, for every n, if $\overline{\mathsf{n}} = \ulcorner\varphi\urcorner$ for some φ, then $\mathsf{T}(\overline{\mathsf{n}}) \leftrightarrow \varphi$, or else $\mathsf{T}(\overline{\mathsf{n}}) \leftrightarrow \bot$. So, while working from inside $\mathsf{PA}_1 \cup \mathsf{Th}$, *Kurt still doesn't know whether* $\mathsf{T}(\overline{\mathsf{n}})$ *has been defined for all numbers n.*

In *this* sense, then, even with the extra axiom in play, cautious Kurt still doesn't yet know whether $\mathsf{T}(\mathsf{x})$ expresses a fully determinate property of numbers. So, for a start, he won't take himself to be entitled in general to employ universal quantifier introduction applied to complex expressions involving $\mathsf{T}(\mathsf{x})$, for he doesn't yet know which such quantifications even have a determinate truth-value. Hence, in particular, he won't in general be in a position to establish the quantified premiss needed to make use of an instance of the induction scheme involving a predicate embedding $\mathsf{T}(\mathsf{x})$. In other words, Kurt is not entitled to make use of the extended instances of induction allowed in $\mathsf{T}(\mathsf{PA})$. Which is why a flat-footedly cautious Kurt – so far – is in no position to inductively inflate $\mathsf{PA}_1 \cup \mathsf{Th}$.

e) Note, it is again *not* being suggested that Kurt positively reject the instances of the induction schema that embed the predicate T as being *false*. How can he do that? As the argument for inductive generosity reminds us, *if* the antecedents of such an instances are true, the consequent has to be true too. Rather, as before, the point is a rather more subtle one: Kurt doesn't find himself entitled to get to the starting line for using such an instance.

So now we can see in retrospect that talk of the theory $\mathsf{PA}_1 \cup \mathsf{Th}$ in fact glosses over a rather critical issue. In bolting the axioms Th onto the theory PA_1, were we intending these new axioms to interact merely with sufficient logical rules governing the elimination of quantifiers and the use of conditionals to enable the extraction of the information packaged in those axioms? Or were we intending that the whole weight of first-order logic can be brought to bear on axioms from either pool – including the universal quantifier introduction rule so that we can trivially prove, for example, $\forall \mathsf{x}(\mathsf{T}(\mathsf{x}) \vee \neg\mathsf{T}(\mathsf{x}))$? If we are just thinking about what we can get out of $\mathsf{PA}_1 \cup \mathsf{Th}$ by way of 'truth' assignments to particular formulae, the issue doesn't arise. But now the question has been raised, we see that the second alternative in fact overgenerates in enabling us to deduce more than we are strictly entitled to in just being given the axioms Th as characterizing the new predicate T. A cautious Kurt really need, and perhaps should, only use quantifier elimination and the conditional rules on Th (of course, it is an old point that the needed working logic of a truth-theory can be, perhaps ought to be, weak: see e.g. (Evans, 1976)). It is this sort of inferentially sanitized version of $\mathsf{PA}_1 \cup \mathsf{Th}$ that, so to speak, properly reflects the conceptual position that doesn't inductively inflate.

Of course, there is nothing to stop Kurt throwing caution to the winds and starting to 'think outside the box'. He can stand back from PA_1, think about his practice, commit himself explicitly to the thought that every sentence of L_1 is either true or false, reflect this thought using an arithmetized truth-predicate which he takes to be fully defined, so induction must apply to it – and so he comes to endorse $\mathsf{T}(\mathsf{PA})$. We certainly don't want to *ban* Kurt from such reflections or suppose that he must make any kind of mistake if he takes on these further thoughts. *But the key point to emphasize is just that they* are *further thoughts, not commitments already necessarily accepted in flat-footedly endorsing* PA_1 *in the first place.* So the situation is, after all, again consistent with Isaacson's Thesis – i.e. it does indeed take new thoughts to get us beyond PA_1. PA_1 plus (in a broad sense) 'logical commentary from the sidelines' takes us to the cut-down version of $\mathsf{PA}_1 \cup \mathsf{Th}$ which is a conservative extension.

Finally, before leaving these matters, it is worth noting that there is much more than a mere analogy between the case of inflating to ACA on the one hand and case of inflating to T(PA) on the other. The two themes in this paper tie together because ACA *and* T(PA) *are in a good sense the same theory!* Putting it more carefully, you can intertranslate the theories in a way which leaves their purely arithmetic content fixed (the idea is essentially that you can replace talk of a set X in ACA with a term in T(PA) standing in for the numerical code for an open formulae of L_1 with X as its extension, and then a wff of the form $\mathsf{t} \in \mathsf{X}$ gets replaced in T(PA) with a claim saying that a certain wff is true). However, it isn't at all clear what the significance of this intertranslatability is for further reflections about the real commitments of ACA. Does it show, for example, that we can in some sense deflate the ontology of ACA (and presumably the weaker ACA_0)? Or is it the other way about, does it really show that T(PA) is more committing than we suspected?[9] Here, we will just have to leave the question hanging.

7 Conclusion

We've only scratched the surface here. Now we see that ACA_0 overgenerates with respect to its conceptual motivation, we can immediately see that the same is going to be true of the other subsystems that Simpson considers in *SoSOA*, and for parallel reasons. The systematic failing here arises – I suspect – by thinking of the subsystems of full second-order arithmetic as just that, *subsystems*, approached from the top down, constructed by cutting out some commitments of the full theory without carefully rethinking what should be left in place. Better to proceed more cautiously from the bottom up, adding conceptual commitments as you go, and that way keeping closer tabs on *exactly* what it is assumed that we have conceptual command of at each level as we move upwards. But following *that* programme in more detail is the project for other days, and eventually for another book.[10]

[9] Compare: It is familiar that first-order arithmetic is intertranslatable with the theory HF of hereditarily finite sets – see (Fitting, 2007, Ch. 1). What deep morals about the commitments of PA_1 or HF should we draw?

[10] I've been particularly helped by comments and encouragement, at various stages, by Tim Button, Luca Incurvati, Dan Isaacson, Alex Paseau, Michael Potter, and Tim Storer.

References

Dummett, M., 1993. *What is mathematics about?*, pp. 429–445. Oxford: Clarendon Press.

Evans, G., 1976. Semantic structure and logical form. In G. Evans and J. McDowell (eds.), *Truth and Meaning*, chap. 199-222. Oxford: Clarendon Press.

Field, H., 1999. Deflating the conservativeness argument. *Journal of Philosophy*, 96: 533–540.

Fitting, M., 2007. *Incompleteness in the Land of Sets*. London: College Publications.

Fraenkel, A. A., Bar-Hillel, Y., and Levy, A., 1973. *Foundations of Set Theory*. Amsterdam: North-Holland Publishing Co., second edn.

Halbach, V., 1999a. Conservative theories of classical truth. *Studia Logica*, 62: 353–370.

Halbach, V., 1999b. How innocent is deflationism? *Synthese*, 126: 167–194.

Isaacson, D., 1987. Arithmetical truth and hidden higher-order concepts. In The Paris Logic Group (ed.), *Logic Colloquium '85*. Amsterdam: North-Holland.

Levy, A., 1976. The role of classes in set theory. In G. H. Müller (ed.), *Sets and Classes: On the Work by Paul Bernays*. Amsterdam: North-Holland Publishing Co.

Shapiro, S., 1991. *Foundations without Foundationalism: A Case for Second-Order Logic*. Oxford Logic Guides 17. Oxford: Clarendon Press.

Simpson, S. G., 1991. *Subsystems of Second Order Arithmetic*. Berlin: Springer.

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

Takeuti, G., 1987. *Proof Theory*. Amsterdam: North-Holland.

Weyl, H., 1918. *Das Kontinuum*. Leipzig: Veit. Translated as Weyl 1994.

Weyl, H., 1994. *The Continuum*. New York: Dover.