A Conversation about Nonstandard Models of Peano Arithmetic

John Baez and Michael Weiss

This file records an e-conversation between John Baez (JB) and me (MW). It also appears as a series of posts on my blog, diagonalargument.com.

1 Post 1

JB: I've lately been trying to learn about nonstandard models of Peano arithmetic. Do you know what a "recursively saturated" model is? They're supposed to be important but I don't get the idea yet.

MW: What books and/or papers are you reading? I used to know this stuff, indeed my thesis (1980) was on existentially complete models of arithmetic. When I looked at it a couple of years ago, I was amazed at how much I'd forgotten. Talk about depressing.

Anyway, I'll toss out a few vague ideas, to see if they help. Maybe this will be the push I need to get back to Kaye's book [6], or even Kossak & Schmerl [9]. I picked them up a few months ago, hoping to revisit my youth, but I didn't make it past the prefaces.

As Hodges puts it, model theory is "algebraic geometry minus fields". If you have an algebraic number r in a extension field K/F, it's natural to look at all the polynomials in $F[x]$ which have r as a root. It turns out that this is a principal ideal, generated by the minimal polynomial.

Sort-of generalize to an arbitrary model M of a theory T. Let $(r_1, \ldots, r_n)$ be an n-tuple of elements of M. Look at the set of all formulas $\varphi(x_1, \ldots, x_n)$ such that M satisfies $\varphi(x_1, \ldots, x_n)$. This is the complete n-type of $(r_1, \ldots, r_n)$ with respect to M.

Unlike the case in field theory, complete n-types are not usually implied ("generated") by a single formula, but when they are, they are called principal.

Next step is to free the notion of n-type from dependence on the model M. If we have a set of formulas $\Phi = \{\varphi(x_1, \ldots, x_n)\}$ that is consistent with T, then it's an n-type. (Sort of like a polynomial in $F[x]$, looking for a root.) The type is realized in a model M if there is a n-tuple $(r_1, \ldots, r_n)$ such that M satisfies all the $\varphi(x_1, \ldots, x_n)$. It's omitted if there is no such n-tuple in M.

The omitting types theorem says that any countable collection of non-principal n-types in a countable language can all be omitted by some model of T.

You can imagine that n-types tell us a lot about possible isomorphisms and automorphisms of models. The poster child: the theory DLO of dense linear orderings without endpoints. This is because there are essentially only very simple types. Pretty much the only thing you can say about $(r_1, \ldots, r_n)$ is what permutation of the subscripts puts them in monotonically increasing order, and for which subscripts i and j we have $r_i = r_j$. Anything else about the n-tuple is implied by this. Especially significant is that no quantifiers need apply.

And of course, any two countable DLOs are isomorphic, and there are lots of automorphisms of a countable DLO. If you look at the back-and-forth argument, you'll see it relies on the limited repetoire of types.

Also, to hammer the analogy with field theory, the polynomials r satisfies tells us all about the isomorphisms and automorphisms.

OK, now specialize to PA. Key thing here is the overspill lemma. This allows you to code info about subsets of the standard $\mathbb{N}$ into single elements of a non-standard M. For example, the set of standard prime divisors of a non-standard r in M.

If the set you want to code is totally arbitrary, well, you can't even express it in the language of PA, so we want to look at sets that are definable by a formula. Then we can apply the overspill lemma. Say $\varphi(x)$ is the formula for the set. Consider the formula $\psi(y) \equiv (\exists z)(\forall x < y)[\varphi(x) \leftrightarrow p_x|z]$ Here p_x is the x-th prime. Since $\psi(n)$ holds for arbitrarily large finite n's (indeed all finite n's), overspill says that it also holds for some non-standard n. So there is a z such that $\varphi(x)$ is true iff $p_x|z$, for all $x < n$. In particular it holds for all finite x, and so z codes the set via its prime divisors.

More generally, it would be nice to look at sets of n-tuples defined by an n-type. But we need some way to describe the n-type in the language of PA—we can't describe an arbitrary set of formulas in the language of PA. So we look at recursive n-types, i.e., sets of formulas whose Gödel's numbers form a recursive set.

If we're lucky, we'll be able to import a lot of classical model theory, developed in ZF (or even naive set theory) into PA, because we can code it all in PA. And this approach will help us understand things like extensions of one model by another, automorphisms, etc.

A couple more notes. First, the prenex normal form hierarchy shows up all the time in logic. A variant especially adapted to PA (the *arithmetic hierarchy*) is to let Δ_0 formulas allow arbitrary bounded quantifiers, like $(\forall x < y)$.

It turns out that a relation is recursively enumerable iff it is Σ_1. So that's a nice bridge between the model theory of PA and computability theory.

It also turns out that, because of the MRDP theorem, you don't even need to allow bounded quantifiers at the ground level. Well, that's a bit sloppy. What I mean is that Σ_1 is the same as $\exists_1$, formulas you get from putting a string of existential quantifiers in front of a quantifier-free formula.

2 Post 2

JB: The only books I know are Kaye's *Models of Peano Arithmetic* [6] and Kossack and Schmerl's more demanding *The Structure of Models of Peano Arithmetic* [9], and I'm trying to read both. But I have a certain dream which is being aided and abetted by this paper:

> Ali Enayat, "Standard models of arithmetic" [2].

Roughly, my dream is to show that "the" standard model is a much more nebulous notion than many seem to believe.

One first hint of this is the simple fact that for any sentence that's not decidable in Peano Arithmetic, there are models where this sentence is satisfied and models where it's not. In which camp does the standard model lie? We can only say: "one or the other". And the same is true in any recursively enumerable consistent extension of PA. There are always undecidable sentences in any such extension, and there are always models—infinitely many, in fact—where the sentence is satisfied, and infinitely many where it is not!

So, choosing "the" standard model among all the pretenders—or even this lesser task: finding it up to elementary equivalence—is a herculean task, like finding ones way through an infinite labyrinth of branching paths, where at each branch deciding which way to go requires brand-new insights into mathematics, not captured by our existing axioms. I'm not convinced it makes sense to talk about the "right" decision in every case.

Another hint is the Overspill Lemma, used ubiquitously in the study of nonstandard models. Roughly:

> **Overspill Lemma:** In a nonstandard model of Peano arithmetic, any predicate in the language of arithmetic that holds for infinitely many standard natural numbers also holds for infinitely many nonstandard ones.
>
> **Corollary:** There is no way to write down a predicate in the language of arithmetic that singles out the standard natural numbers.

So, if our supposed "standard" model were actually nonstandard, we still couldn't pick out the nonstandard numbers. *The aliens could be among us, and we'd never know!*

But you know all this stuff, and probably have different interpretations of what it means. I don't mainly want to argue about that; I mainly want to learn more, so I can prove some interesting theorems.

So, thanks for your great introduction! I especially love the idea of types, which I just learned. Your explanation was great. Here's more, in case anyone is reading:

> Types (model theory), Wikipedia.

You didn't quite get to "recursively saturated models", so I pestered Joel David Hamkins (a logician I know, now at Oxford) and got this nice explanation, which fits nicely with yours:

> Dear John,

Saturation is about realizing types. The *type* of an element a in a model M is the set of all formulas $\varphi(x)$ such that M satisfies $\varphi(a)$. You can allow parameters in a type. For example, if you consider the reals as an ordered field, then any two elements a and b have different types, since there is some rational p/q between them, and a, say, will realize the formula $a < p/q$, while b does not. This formula $x < p/q$ is expressible in the language of fields.

Does $\mathbb{R}$ realize all types that it could? No, because you can write down a collection of formulas $\varphi_n(x)$ asserting that $0 < x$ and $x < 1/n$. This is expressible in that language, and it is consistent with the diagram of the reals, meaning that this type could be realized in some elementary extension of $\mathbb{R}$. The type asserts that x is a positive infinitesimal number. So the reals are not saturated.

Indeed, since that type was very easy to describe, the reals are not even recursively saturated (I usually call it computably saturated). To be recursively saturated, a model M must realize every computable type (the formulas, as a set, are a computable set of formulas) that is consistent with its diagram.

In general, saturated models are very thick—they have lots of points of all the types that one could in principle have in a model of that theory. Usually, countable models are not saturated (except when the language is trivial in some way). To get around this, the idea of recursive saturation works very well with countable models, and this is what is going on in Kaye's book.

Hope this helps, and let me know if I can explain anything more about it...

Kind regards,

Joel

JB: It's so much easier to learn stuff by talking with people than by fighting through a thicket of logic symbols! But it's also good for me to keep reading the books.

3 Post 3

MW: There's also the book by Hájek and Pudlák, but I don't have a copy of that. Thanks muchly for the Enayat paper, which looks fascinating.

What you and Enayat are calling the "standard" model of arithmetic is what I used to call "an ω", i.e., the ω of a model of ZF. Is that the new standard terminology for it? I don't like it, for philosophical reasons I won't get into. (Reminds me of the whole "interpretations of QM" that books have to skirt around, when they just want to shut up and calculate.)

Leaving ZF out of it, a friend in grad school used to go around arguing that 7 is non-standard. Try and give a proof that 7 is standard using fewer than seven symbols. And of course for any element of a non-standard model, there is a "proof" of non-standard length that the element is standard. I think he did this just to be provocative. Amusingly, he parleyed this line of thought into some real results and ultimately a thesis.

JB: Fun! It reminds me a bit of this:

> I have seen some ultrafinitists go so far as to challenge the existence of 2^{100} as a natural number, in the sense of there being a series of "points" of that length. There is the obvious "draw the line" objection, asking where in 2^1, 2^2, 2^3, ..., 2^{100}, do we stop having "Platonic reality"? Here this "..." is totally innocent, in that it can easily be replaced by 100 items (names) separated by commas. I raised just this objection with the (extreme) ultrafinitist Yessenin-Volpin during a lecture of his. He asked me

> to be more specific. I then proceeded to start with 2^1 and asked him if this was "real" or something to that effect. He virtually immediately said yes. Then I asked about 2^2, and he again said yes, but with a perceptible delay. Then 2^3, and yes, but with more delay. This continued for a couple more times, till it was obvious how he was handling this objection. Sure, he was prepared to always answer yes, but he was going to take 2^{100} times as long to answer yes to 2^{100} as he would to answering 2^1. There is no way that I could get very far with this.
>
> Harvey M. Friedman, "Philosophical Problems in Logic"

I have as rather distant acquantainces a couple of famous ultra-finitists, namely Yessenin-Volpin and Edward Nelson. I think they're on to something but I think it's quite hard to formalize.

Both of them were real characters. Yessenin-Volpin was a dissident in the Soviet Union, locked in a psychiatric hospital for what Vladimir Bukovsky jokingly called "pathological honesty". Edward Nelson was another student of my Ph.D. advisor, Irving Segal. He did ground-breaking work on mathematically rigorous quantum field theory. Much later he wrote a fascinating paper on internal set theory, which I'd like to talk about sometime. In his final years he thought he had proved the inconsistency of Peano arithmetic! Terry Tao caught the mistake in the proof, and Nelson quickly admitted his error. That was an embarrassing incident, but he came out of it fine. Crackpots never admit they're wrong; he was not like that!

4 Post 4

MW: I wrote: "I don't like calling the ω of a model of ZF a *standard model*, for philosophical reasons I won't get into."

JB: I like it, because I don't like the idea of "the" standard model of arithmetic, so I'm happy to see that "the" turned into an "a".

MW: Well, on second look, I see Enayat uses "ZF-standard" in the body of the paper. I'm fine with that.

Anyway, back when I was in grad school, I wondered whether there were models of true arithmetic that are not ω's. Answer: yes. I wrote up a short note, just for myself. I'm sure the result is well-known, maybe even in one of the three books we've mentioned. (Also, at this remove I no longer remember if I came up with the argument on my own, or if my advisor gets the credit.)

JB: What's "true arithmetic"?

MW: Just the theory of the standard model, i.e., all closed formulas satisfied by $\mathbb{N}$. Now, if you don't believe the term "the standard model" means anything, then I'd have to write a much longer and more confusing definition.

JB: What's $\mathbb{N}$? (Most of the time I act like I know, but when I'm doing logic I admit that there are lots of different things one could mean by it.)

I don't think "the standard model" means anything unless it's defined. So please give me your definition!

If I pick a set theory, say ZFC or whatever, I can prove in there that's there's an initial model of PA (one with a unique embedding in any other), and I'm happy to call that the "standard model of PA in ZFC". Then I can talk about the theory consisting of all closed sentences in this model. All that's fine with me.

MW: Hmmm. . . It looks like we're not going to be able to avoid *all* philosophy. Fair warning: my philosophy of math is a mish-mash of intuitionism, formalism, and platonism. And I'm not using those terms in the usual way!

But I think this is a topic for the next post. After that, I promise we'll get back to Real Math.

5 Post 5

MW: John, you wrote

> Roughly, my dream is to show that "the" standard model is a much more nebulous notion than many seem to believe.

and you gave a good elucidation in post 2 and post 4. But Id like to defend my right to "true arithmetic" and "the standard model $\mathbb{N}$".

Maybe being nebulous isn't so bad! It doesn't wipe out your ontological creds. Like, say, look at that cloud over there:

> HAMLET: Do you see yonder cloud that's almost in shape of a camel?
>
> POLONIUS: By the mass, and 'tis like a camel, indeed.
>
> HAMLET: Methinks it is like a weasel.
>
> POLONIUS: It is backed like a weasel.
>
> HAMLET: Or like a whale?
>
> POLONIUS: Very like a whale.

No one's saying there's no cloud. Or that you can only talk about the cloud if you do so in the language of ZF set theory!

JB: I dont mind nebulous philosophy. But if youre going to say "true arithmetic" and "the standard model $\mathbb{N}$" in mathematical discourse, I think you need to define them. (That's for the "hard core" of mathematical discourse, where one is stating theorems and conjectures, and proving them. It's fine, indeed essential, to say a lot of vaguer stuff while doing mathematics.)

Imagine this:

> POLONIUS: Let G be a Lie group. If G is connected and simply connected, it's determined up to isomorphism by its Lie algebra.
>
> HAMLET: Sorry, could you remind me—whats a "Lie group"?
>
> POLONIUS: What?? Every mathematician worthy of the name has a clear intuitive concept of a Lie group! How can you possibly ask such a question?

MW: That reminds me of an incident from grad school. My friend Mark and I went to ask one of our favorite profs, Bert Walsh, something about Riemann surfaces. He began by saying, "Well, you take your Dolbeault complex..." At the time, neither Mark nor I knew a Dolbeault complex from a dollhouse. So Mark asked him, "Whats a Dolbeault complex?" To which Prof. Walsh replied, "Mark, its exactly what you think it is!"

Since this is a Philosophy-with-a-capital-P post, let me sling the jargon: do you mean that $\mathbb{N}$ is nebulous *epistemically*, or *ontologically*?

Epistemology—what we *know*—sure that's nebulous! There's a heck of lot we don't know about $\mathbb{N}$. Are there odd perfect numbers? An infinite number of prime pairs? The ABC conjecture! Even something like the Riemann ζ hypothesis can be "coded" into the language of arithmetic with a little work. (Or a *lot* of work, if you really mean "written" and not just "convince yourself it *could* be written".)

But ontologically—are there lots of different $\mathbb{N}$'s, or just one? The very term "non-standard models of arithmetic" has a double-edged tinge to it. If there are *non-standard* models, then there must be a *standard* model!

You gave a sort-of justification for the phrase, "the standard model", in post 4:

> If I pick a set theory, say ZFC or whatever, I can prove in there that's there's an initial model of PA (one with a unique embedding in any other), and I'm happy to call that the "standard model of PA in ZFC". Then I can talk about the theory consisting of all closed sentences in this model. All that's fine with me.

But do we *have* to "pick a set theory, say ZFC or whatever", to give the proof?

I said my philosophy of math was an incoherent mish-mash of intuitionism, formalism, and platonism. Let's start with intuitionism. I think I read somewhere that the taproot of Brouwer's philosophy (or maybe Poincaré's) consists in this: mathematical intuition comes first, axioms later.

That's all I'm taking from those guys. But it's important. I'm betting you saw the proof that there is, up to isomorphism, only *one* model of Peano's axioms, long before you learned about first-order theories and non-standard models and all that. (It's basically in Dedekind's famous essay, "Was sind und was sollen die Zahlen?", paragraph 79.)

Of course, "Peano's axioms" here isn't the same as "PA". The induction scheme is a single axiom quantifying over all subsets of $\mathbb{N}$, or as logicians like to say, it's formulated in second-order arithmetic. (Or you can go all the way to ZFC, if you like.)

Here's another way to put it: when you say, "pick a set theory", I say, "OK, what if I pick the intuitive set theory of Dedekind and Cantor?"

I know what you're thinking. "Has he totally forgotten about Russell's paradox? Has he even *studied* axiomatic set theory? Does he really believe that $2^{\aleph_0}$ has a single correct value, even if we may never know what it is?"

I'll admit that oxygen can get a little scarce far up in V. Do I believe that Woodin cardinals really exist? When the air becomes hard to breathe, I might want to take refuge in Hilbert's gambit, and just claim it's all a game with symbols. (That's my formalism ingredient. Again, a very small part of Hilbert's program, but important.)

The Hilbert gambit may be philosophically unimpeachable, but you know, it just isn't that much *fun!* G. H. Hardy famously refused to accept that Hilbert really believed it. When watching *Game of Thrones*, do you tell yourself the whole time, "Those aren't real dragons, they're just CGI."?

Let's take the "existence" of non-standard models of PA in the first place. From a strictly formalist standpoint, we'd have to say: "here's a proof in ZFC that $\exists N(\ldots)$", where the ellipsis is a formalization of "N is a model of the PA axioms that is not isomorphic to ω". Of course nobody does that. We carry out the proof that such beasties exist in an informal set theory, convinced (with good reason!) that it could (under duress) be turned into a ZFC proof.

And that's the platonism part of my philosophy. The outermost layer will always be raw mathematical intuition. But when I'm swimming inside ZFC, it's just easier, and *more fun*, to imagine that the ZF universe really exists. I pretend that the axioms are laws of this universe. The laws tell me there's a unique ω, and (equipped with the usual paraphenalia), *that's* what I mean by "the standard model". Who knows, maybe it's really true.

(Does *The Matrix* have any scenes of someone playing video games? Or Greg Egan's *Permutation City*? Can't remember.)

Now for the kicker. Say we go with the Hilbert gambit. You need some raw intuition about the natural numbers just to make sense of it! Anyone who will swallow "well-formed formula" and "proof tree" as sufficiently clear concepts, but balks at "natural number", has a strange perspective, IMO. The Hilbert gambit buys you something, philosophically, when we're talking about full-bore set theory. Not so much with arithmetic.

Put it this way. You wrote:

> for any sentence that's not decidable in Peano Arithmetic, there are models where this sentence is satisfied and models where it's not. In which camp does the standard model lie?... [By the Overspill Lemma] if our supposed "standard" model were actually nonstandard, we still couldn't pick out the nonstandard numbers. *The aliens could be among us, and we'd never know!*

The results you appeal to (the Completeness Theorem and the Overspill Lemma) are themselves theorems of either ZFC, or of informal set theory. But so is the uniqueness (up to isomorphism) of $\mathbb{N}$. Why regard these results differently?

On the other hand, if you meant "nebulous" epistemically, well, just about all math is nebulous in that sense. There's a lot we don't know about—fill in the blank!

6 Post 6

JB: It's interesting to see how you deploy various philosophies of mathematics: Platonism, intuitionism, formalism, etc. For a long time I've been disgusted by how people set up battles between these, like Punch-and-Judy shows where little puppets whack each other, instead of trying to clarify what any of these philosophies might actually *mean*.

For example, some like to whack Platonism for its claim that numbers "really exist", without investigating what it might mean for an abstraction —a Platonic form—to "really exist". If you define "really exist" in such a way that abstractions don't do this, that's finebut it doesn't mean you've defeated Platonism, it merely means you're committed to a different way of thinking and talking.

Indeed, I'm so tired of these Punch-and-Judy shows that I run, not walk, whenever I hear one roll into town. I like your approach better, where you seem to treat these different philosophies of mathematics, not as mutually exclusive factual claims, but as different attitudes toward mathematics. We can shift among *attitudes* and see things in different ways.

But still, I'd rather dodge all direct engagement with philosophy in this conversation, since that's not why I'm interested in models of Peano arithmetic. Or rather, my interest in models of Peano arithmetic is a highly *sublimated* form of my youthful interest in philosophy. Instead of trying to tackle hard questions like "do the natural numbers really exist, what are they, and how do we know things about them?" I find it easier and more fun to learn and prove theorems.

But I have an ulterior motive, which I might as well disclose: I want to soften up the concept of "the standard model" of Peano arithmetic. I want to push toward some theorems that say something like this: "what *you* think is the standard model, may be nonstandard for *me*."

This is pretty heretical. We've all seen this picture where the standard model is the smallest possible model, that only has the numbers it "needs to have", and any nonstandard model has extra "infinitely big" numbers, after all the "finite" ones, coming in patches that look like copies of $\mathbb{Z}$. So if your standard model looks nonstandard to me, it means some number that seems perfectly standard to you seems "infinitely big" to me. To me, it lies one of these patches that come after all the standard numbers. No matter how many 1's I subtract from it, I'll never get down to 0 as long as the number of 1's is standard *for me.*

What if for really big numbers, it's hard to tell if they're standard or non-standard?

This sounds crazy, I admit. But have you ever seriously played the game of trying to name the largest natural number you can? If you do, you'll probably wind up using the busy beaver function $\Sigma(n)$, which tells us the most 1's that can be printed by an n-state 2-symbol Turing machine that eventually halts. This function grows faster than any computable function, and makes it easy to name immensely large numbers. For example we know

$$\Sigma(7) \geq 10^{10^{10^{10^{18705353}}}}$$

but this is probably a ridiculous underestimate. We know

$$\Sigma(12) \geq 3 \uparrow\uparrow\uparrow\uparrow 3$$

which means $3^{3^{3^{\cdots}}}$, where the tower of threes is very tall. How tall? It has $3^{3^{3^{\cdots}}}$ threes in it, where this tower has $3^{3^3} = 7625597484987$ threes in it.

So $\Sigma(12)$ is fairly large, but the Busy Beaver function quickly gets much larger, and rockets into the mists of the unknowable. For example, Aaronson and Yedida showed that S(7918) can't be computed in ZFC if ZFC together with a certain large cardinal axiom is consistent:

> Scott Aaronson, The 8000th Busy Beaver number eludes ZF set theory, May 3, 2016, blog post at Shtetl-Optimized on a paper by Scott Aaronson and Adam Yedidia.

Even the game of naming really large computable numbers leads us into Turing machines and large cardinal axioms! In a thread on the xkcd blog, Eliezer Yudkowsky won such a game using an I0 cardinal. For readers not familiar with these, I recommend Cantor's Attic, which invites us to study large cardinals as follows:

> Welcome to the upper attic, the transfinite realm of large cardinals, the higher infinite, carrying us upward from the merely inaccessible and indescribable to the subtle and endlessly extendible concepts beyond, towards the calamity of inconsistency.

I find it remarkable how the simple quest to name very large natural numbers brings us up into this lofty realm! It suggests that something funny is going on, something I haven't fully fathomed.

"But come on, John", you may respond. "Just because there are very big, mysterious natural numbers doesn't mean that they aren't standard."

And indeed I have to admit that's true. I see no evidence that these numbers are nonstandard, and indeed I can prove—using suspiciously powerful logical principles, like large cardinal axioms—that they *are* standard. But that's why this paper excites me:

> Ali Enayat, Standard Models of Arithmetic.

It excites me because it relativizes the notion of "the standard model" of PA. It gives a precise sense in which two people with two versions of set theory can have different standard models. And it points out that assuming a large cardinal axiom can affect which models count as standard!

Anyway, these are my crazy thoughts. But instead of discussing them now (and I'm afraid they go a lot further), I'd much rather talk about mathematical logic, and especially the math surrounding Enayat's paper.

MW: Sounds good! We can save the philosophy (or at least what *I* call philosophy) for another day

7 Post 7

MW: Our goal for the next few posts is to understand Enayat's paper

> Ali Enayat,"Standard models of arithmetic" [2]

JB: Yee-hah!

MW: I'm going to take a leisurely approach, with "day trips" to nearby attractions (or Sehenswürdigkeiten, in the delightful German phrase), but still trying not to miss our return flight.

Also, I know you know a lot of this stuff. But unless we're the only two reading this (in which case, why not just email?), I won't worry about what you know. I'll just pretend I'm explaining it to a younger version of myself—the one who often murmured, "Future MW, just what does *this* mean?"

Enayat leads off with what TV critics like to call *table setting*. Although some critics sniff contemptuously at this, table setting is an Excellent Thing in a math paper—I wish everyone did it as well as Enayat does here. For me, it's instantly obvious why you'd care about which models of PA can be the "standard" $\mathbb{N}$ in a model of ZF. But Enayat doesn't take that for granted. He also explains why $\mathrm{PA}^{\mathrm{ZF}}$—statements that are true in all such models—is recursively axiomatizable.

Let's get some terminology and notation out of the way first. Enayat doesn't want to tie himself down to ZF in particular; you might want to add some other axioms, like AC, or Con(ZF), or SM, or some large cardinal axioms. So he uses T to stand for some recursively axiomatizable extension of ZF (maybe ZF itself).

JB: What's "SM"?

MW: SM is "there exists a model of ZF whose universe is a set (not a proper class), and whose elementhood relation is the "real" one in the "real" universe". This is the so-called Standard Model axiom. It's implied by the weakest of all large cardinal axioms, the existence of an inaccessible cardinal.

JB: Okay, interesting. We category theorists use Grothendieck universes now and then, like when studying the "category of all small categories". I believe the existence of a Grothendieck universe is equivalent to the existence of an inaccessible cardinal (or to be precise, a strongly inaccessible cardinal). It sounds like the existence of a Grothendieck universe is precisely the Standard Model axiom. I'll check that out sometime. Anyway, go on.

MW: Nowadays, *inaccessible* pretty much always means *strongly inaccessible.* You're right about Grothendieck universes: see Shulman's paper [14, §8].

If M is a model of T, Enayat says that $\mathbb{N}^M$ is a T-standard model of PA. I used to call these "omegas", since the domain of $\mathbb{N}^M$ is the ω of the model M. Note that if M is a "standard" model (in the sense I just explained), then $\mathbb{N}^M$ is just the standard $\mathbb{N}$.

JB: Hmmm. Though you say "just *the* standard $\mathbb{N}$", even under the assumptions you're making, this $\mathbb{N}^M$ seems to depend on T and on a choice of standard model M of T. Are you trying to tell me that it doesn't really depend on either of these choices?

Since I'm trying to sift through all these dependencies, let me start by reminding myself, and everyone else, that we're getting $\mathbb{N}^M$ by first choosing a recursively axiomatizable extension T of ZF, and then choosing a model M of T. Then we define the natural numbers in the usual way in the theory T, and see what set, with successor operation and zero, it corresponds to in our model. That's $\mathbb{N}^M$.

MW: Exactly. We get the same ω no matter which T and which M, so long as M is a standard transitive model. Set-theorists would say that's because ω is *absolute* for standard transitive models. Let me unpack that.

A structure for ZF is a pair (K, ϵ), where K is a subclass of V (the "actual universe of all sets"), and ϵ is a binary relation on K ("elementhood"). If ϵ is the "actual" elementhood relation $\in$, then we've got a so-called standard structure. So you get a standard structure just by looking at *all* the sets of V, and deciding which ones deserve to be in the club! $(K, \in)$ is a standard *model* iff it satisfies all the ZF axioms, natch.

Now, the sets of K could still be deceiving us, even with a standard model. Say s belongs to K, but none of its elements do. Even though s *has* elements in the "real universe" V, s looks just like the empty set inside K. To ward off discombobulations of this ilk, we demand that K be *transitive.* This means that if s belongs to K, so do all of its "real world" elements (i.e., $x \in s \in K$ implies $x \in K$), and so on down for elements of elements of s, etc.

Incidentally, the transitivity requirement for models of ZF is much like the initial segment requirement for models of PA. M is an initial segment of N (with models of PA) iff whenever $m < n \in N$, we have $m \in N$. In PA land, people tend to talk about *end extensions*: N is an end extension of M iff M is an initial segment of N. So I guess you *could* say that K is transitive iff V is an end extension of it. But set theorists don't typically talk that way.

JB: Thanks! I need to learn a lot more about absoluteness. I get the basic idea, but I don't know any of the theorems that say which things are absolute. This was borne in on me when at some point in Enayat's paper he says "routine absoluteness considerations show. . . ". I thought "Hey, wait a minute!"

I guess it's sort of obvious that concepts are more likely to be absolute when they don't refer too much to the big wide world around them. But anyway, as Yogi Berra would say, we can burn that bridge when we get to it. Continue as you see fit!

MW: Good intuition. As I put it in my Smullyan notes, to verify that a set w is really ω, we just need to crawl around inside it. We don't need to climb outside it and wander around the entire class K. Not so for the power set of ω, where we have to search high and low to make sure we've gathered *all* the subsets.

Logicians (specifically, Azriel Lévy) have developed a machinery to help determine absoluteness in set theory. Basically, peruse the quantifiers. If you hear the terms Δ_0, Δ_1, or Σ_1 being tossed around, that's what's going on.

As a side note, late in life Paul J. Cohen reminisced about his discovery of forcing. Here's a bit of what he said about his first encounter with Gdel's monograph on the consistency of AC and CH:

> …it had an exaggerated emphasis on relatively minor points, in particular, the notion of *absoluteness*, which somehow seemed to be a new philosophical concept. From general impressions I had of the proof, there was a finality to it, an impression that somehow Gödel had mathematicized a philosophical concept, i.e., constructibility, and there seemed no possibility of doing this again…

But this is a bit of a detour for us, since Enayat's paper concerns itself with the *non-standard* models of ZF.

"As above, so below." (A quote from the Emerald Tablet of Hermes Trismegistus, a alchemical sacred text.) $\mathbb{N}$ seems so cozy and familiar; the "real universe" V by contrast an enormous, almost mythological mystery.

Was Hermes right? Do the goings-on in the upper reaches of V leave their tell-tale traces in $\mathbb{N}$?

Set theorists have known for quite some time that this holds for the *second-order* theory of $\mathbb{N}$ (aka analysis). That's a way-station between PA and ZF: you're allowed to quantify over arbitrary subsets of ω, but not over subsets of the power set of ω.

The very first thing Cohen did with forcing was manufacture a model of ZF with a non-constructible set of integers—that is, a subset of ω that's not in Gödel's class L. But you can also get this from a large cardinal axiom. Silver and Solovay (independently) showed that if a specific so-called Ramsey cardinal exists, then so does a subset of ω with certain properties—properties precluding constructibility. Solovay dubbed this subset $0^{\#}$. ($0^{\#}$ would make a nice day-trip, maybe after we've talked about truth and satisfaction.)

In a way, Cohen's result also made use of a large cardinal. He relied on axiom SM; perhaps the most natural justification for SM comes from assuming there's an inaccessible cardinal. At the time, this met with some resistance from the community, so Cohen also showed how to convert his proof into a purely syntactic relative consistency argument. (Also you can circumvent SM in another way.)

What about the first-order theory PA? Now in a sense, anytime you assert *anything* about *any* recursively axiomatized theory, you're stepping into PA's jurisdiction. I don't mean that PA will always be able to *settle* the question (prove or disprove it), just that it can be expressed in the language of PA. That's what's going on with Con(ZF), Con(ZFI), etc.

ZF makes its "gravitational field" felt in PA in other ways, too. Topic for the next post.

JB: Good! That's one thing I really want to know about. I was going to mention that "as above, so below" quote from the Emerald Tablet when

I first mentioned the ramifications of large cardinal axioms on arithmetic in Post 6. I decided it was too obscure! But I love the idea that the "microcosm" of the natural numbers may mirror the "macrocosm" of the set-theoretic universe. So let's get into that!

MW: Too obscure!? Wasn't it an Oprah Book Club Selection?

8 Post 8

JB: So, you were going to tell me a bit how questions about the universe of sets cast their shadows down on the world of Peano arithmetic.

MW: Yup. There are few ways to approach this. Mainly I want to get to the Paris-Harrington theorem, which Enayat name-checks.

First though I should do some table setting of my own. There's a really succinct way to compare ZF with PA: PA = ZF−infinity!

Here's what I mean. Remove the axiom of infinity from ZF. The minimal model of this is V_ω, the set of all *hereditarily finite* sets. A set is hereditarily finite if it's finite, and all of its elements are finite sets, and all of *their* elements are finite sets, and so on down.

(Actually I was sloppy—I should have said, "replacing the axiom of infinity with its negation". I thank Sridhar Ramesh for pointing this out.)

To show that ZF−infinity (let's call it ZF$\neg\infty$) is "basically the same" as PA, you have to code things in both directions. We know how the integers are coded: von Neumann's finite ordinals. (Logicians often just say "integers" when they really mean "non-negative integers"; does this drive number theorists crazy?)

JB: If any number theorists actually talked to logicians, it might.

(Just kidding: there are actually lots of cool interactions between number theory and logic [10], with ideas flowing both ways.)

MW: In the reverse direction, suppose we've already coded the all the elements of $\{a_1, \ldots, a_n\}$ as integers $\{i_1, \ldots, i_n\}$. We just need a way to code finite sets of integers as single integers. Lots of choices here, say bitstrings, or using a product of primes:

$$\{i_1, \ldots, i_n\} \leftrightarrow p_{i_1} \cdots p_{i_n}$$

(Here p_i is the i-th prime, naturally.)

Coding between $\mathbb{N}$ and V_ω is just the appetizer. Next comes translating the formal statements between L(PA) (the language of PA) and L(ZF) (the language of ZF). This calls for another trick—something called Gödel's lemma, which uses the Chinese remainder theorem in a clever way. (Turns out you need this just to show that the "i-th prime" function can be expressed in L(PA).) And after that the main course: showing that PA can prove all the (translated) axioms of ZF$\neg\infty$, and vice versa. Kaye devotes a chapter to spelling out *some* of the details, off-loading the rest to exercises.

The point of all this? Just that you can do finite combinatorics in V_ω without much gnashing of teeth. The Paris-Harrington principle is a variant of the finite Ramsey theorem; the Paris-Harrington *theorem* says that the Paris-Harrington principle is unprovable in PA (assuming, of course, that PA is consistent!) People like to call this the first "natural" unprovable statement—"natural" because combinatorialists might give a darn, not just logicians.

JB: I understand Goodstein's theorem a million times better than the Paris-Harrington principle, and I can easily see how the natural proof uses induction up to ϵ_0, though I don't know the Kirby-Paris proof that it's unprovable in PA. I thought this came before the Paris-Harrington theorem?

MW: Goodstein proved his theorem way back in 1944 [3]. Kirby and Paris proved its unprovability (in PA) in 1982, in the paper that also introduced the Hydra game [8]. Although I don't know that you can rely on publication dates for who knew what when. Anyway, Goodstein, Hydra, and the Paris-Harrington principle hang out together like the Three Musketeers. Or the Four Horsemen of Unprovability, if you include Gentzen's result about induction up to ϵ_0, and how it's the precise "proof strength" of PA.

Jan van Plato wrote a historical article with the delightful title "Gödel, Gentzen, Goodstein: The Magic Sound of a G-String" [16]. He says:

> Goodstein had no proof of the independence of his theorem from PA, but it is clear to a careful reader of his paper that he indeed was convinced of the said independence. ...part of the reason for the early neglect of Goodsteins theorem lies in *a false modesty*. ...The same attitude is well displayed by Goodstein's subservient acceptance of each and every comment and criticism of Bernays. The latter persuaded him to suppress the claims to independence from the final version, and one can only speculate what effect a clear-cut conjecture of independence could have had on future research.

Getting back to the Paris-Harrington theorem: the original proof [12] used non-standard models of PA. Kanamori and McAloon [4] soon simplified it; the treatment in the recent book by Katz and Reimann [5] is particularly easy to follow. That's the one I want to look at here.

But since you had such fun with ordinals here (and here and here), I better add that Ketonen and Solovay [7] later gave a proof based on the ϵ_0 stuff and the hierarchy of fast-growing functions.(The variation due to Loebl and Nešetřil [11] is nice and short.) We should talk about this sometime! I wish I understood all the connections better. (Stillwell's *Roads to Infinity* [15] offers a nice entry point, though he does like to gloss over details.)

I figured you must have written about Ramsey's theorem at some point, but I couldn't find anything.

JB: No, I've never really understood Ramsey theory. More precisely, I never understood its appeal. So whenever I try to read about it, I become bored and quit. "In any party of six people either at least three of them are mutual strangers or at least three of them are mutual acquaintances." I'm sure there's *something* interesting about this. I just haven't gotten interested in it.

MW: But you do know, right, it comes in finite and infinite flavors?

JB: Yes.

MW: Ramsey was doing logic when he discovered his most famous result: he solved a special case of Hilbert's Entscheidungsproblem.

You can think of Ramsey's theorem as a supercharged pigeonhole principle. If you color an infinite set of points with a finite number of colors, at least one color paints an infinite number of points. OK, now color instead the edges of the complete graph on an infinity of points. Then there's an infinite subset of the points, such that the complete graph on that subset is *monochromatic*. That's not so obvious! Next step: again start with an infinite set of points, but this time color all the triangles (unordered triples of points). Then there's an infinite subset of points with all its triangles the same color. And so on.

For the finite version, we start with a large finite set of points, and look for a subset of a given size such that all its points, or edges, or triangles (etc.) have the same color. You'll find it if the original set of points is *sufficiently large*. How large equals "sufficiently"? That depends on (a) the cardinality of the things you're coloring (points, edges, triangles, etc.); (b) the number of colors; and (c) how big you want your monochromatic set to be.

Now here's an interesting fact. The finite version of Ramsey's theorem follows from the infinite one by a routine compactness argument. (Or you can use Kőnig's infinity lemma.) Ramsey himself gave a direct inductive proof for both versions—so the finite version is a theorem of PA. Paris and

Harrington made just a *small tweak* to the finite version. The compactness argument barely notices the change—the tweaked version (the Paris-Harrington principle) still follows from the infinite Ramsey theorem. But the direct inductive proof comes to a crashing halt!

ZF can prove the Paris-Harrington principle, an assertion purely about V_ω, but ZF$\neg\infty$ can't! Does this count as the infinite casting its shadow on the finite?

(Hmmm, I've spent so much time table-setting that now it's time to walk the dog. We'll have to get to the (model-based) proof of the Paris-Harrington theorem in the next post.)

Addendum: I thought Ketland's comment deserved some discussion. First, thanks for the reference!

Ketland's comment:

> In logic, one says theories A and B (in disjoint signatures) are definitionally equivalent if there are definitional extensions $A+$ and $B+$ which are logically equivalent. As it turns out, this definitional equivalence relationship is intimately connected to the invertibility of the translation maps between A and B. So theorems of A can be translated to theorems of B and vice versa and the translation functions mutually invert each other.
>
> For "finite ZF" and PA, the detailed definitional equivalence was only recently carefully worked out, though it's a folklore result: it involves the (1937) Ackermann encoding of sets as numbers. So a binary predicate intuitively meaning "n is an element of m" can be defined inside PA (ultimately, just using $0, S, +$, and $\times$; but in fact it needs exponentiation), and it behaves just like the membership predicate on finite sets. But, like much in arithmetic involving coding, the details are messy!
>
> The equivalence is between these two theories

1. PA
2. ZF − axiom of infinity + ¬(axiom of infinity) + the axiom of "transitive containment"

The details appeared in

Richard Kaye & Tin Lok Wong, 2007 "On Interpretations of Arithmetic and Set Theory", Notre Dame Journal of Formal Logic.

The folklore result I had in mind was "derivability equivalence". This is a corollary to Prop. 2.1 in the Kaye-Wong paper. Say $f : S \to T$ is an interpretation of theory S into theory T, and write σ^f for what we get by translating a formula σ of the S-language into the T-language. Suppose $S \vdash \sigma$. Then $T \vdash \sigma^f$. If we also have an interpretation of T into S, then S and T are equi-consistent. That's the situation with PA and ZF¬∞.

Kaye and Wong are after something stronger. Ackermann had provided inverse maps between $\mathbb{N}$ and V_ω; Kaye and Wong leverage these to get interpretations that are truly inverse: $(\sigma^f)^g = \sigma$, $(\tau^g)^f = \tau$. You don't get that with the interpretations I sketched—if you code von Neumann's 3 into $\mathbb{N}$, the result isn't 3! However, you have to replace ZF¬∞ with ZF¬∞+TC to get inverse interpretations, where TC is what they call the axiom of transitive containment (equivalent to $\in$-induction).

Skimming the paper gave me a half-baked idea. It looks like Kaye and Wong are constructing an isomorphism between two categories. I wonder if the "folklore" interpretations give a natural equivalence?

The paper says, "when the details were finally uncovered, there were surprises for both authors". (Interesting, because Kaye outlined some similar ideas in exercise 11.4 of his book.) For me, the need for the axiom of transitive containment was unexpected, although in hindsight it makes perfect sense. When describing *hereditarily finite*, I said, "and all of *their* elements are finite sets, and so on down." That's just a folksy way of saying "the transitive closure is finite." If we can't ask ω to do some heavy lifting for us, showing that there *is* a transitive closure doesn't seem to be possible with just ZF$\neg\infty$.

9 Post 9

MW: Time to talk about the Paris-Harrington theorem. Originally I thought I'd give a "broad strokes" proof, but then I remembered what you once wrote: keep it fun, not a textbook. Anyway, Katz and Reimann do a nice job for someone who wants to dive into the details, without signing up for a full-bore grad course in model theory. So I'll say a bit about the "cast of characters" (i.e., central ideas), and why I think they merit our attention.

The Paris-Harrington *theorem* says that the Paris-Harrington *principle* can't be proved in PA. The principle is a small tweak to the finite Ramsey theorem. Let's recall the Ramsey theorem, briefly. Begin with a finite set of points. Color the n-element subsets with a finite set of colors. A subset of the points is *monochromatic* if all its n-element subsets have the same color. The Ramsey theorem says we will always have a monochromatic set of points of any given size, no matter what n is or how many colors we use, *provided* we start with a sufficiently large set of points. "Sufficiently" depends, of course, on n and the number of colors and how large a monochromatic subset we insist on.

Ramsey first proved his theorem in a quest for *indiscernibles*—we'll hear more about those shortly.

Now for the Paris-Harrington principle: let the points be numbers. Demand in addition that the *size* of the monochromatic subset be greater than its *least element*. (So if the smallest element is 5, we insist on at least 6 elements.)

Paris and Harrington gave two slightly different model-based proofs of their theorem. Approach 1: show that PA proves "(Paris-Harrington principle) implies Con(PA)". So by Gödel's incompleteness theorem, PA can't prove the Paris-Harrington principle. Approach 2: show that there's a model of PA + not-(Paris-Harrington principle). Any theory with a model is consistent, so QED.

I'll concentrate on Approach 2. Before we get started, let me say just a bit about the model. Let N be a model of PA. If N does not satisfy the Paris-Harrington principle, good, we're done! Otherwise, we acquire (somehow!) an infinite set B of indiscernibles, a subset of N. Then we let M be the initial segment determined by B: all the elements of N that are less than *some* element of B. It turns out that M is a model of PA, but *does* not satisfy the Paris-Harrington principle. So now we know the role the indiscernibles play, even if we don't know what they are yet.

Cast of characters:

- Types.
- Indiscernibles.
- Overspill.
- Truth and satisfaction.
- Δ_0, bounded search, and absoluteness.
- Parameters, aka names.

Types. We talked about these in Posts 1 and 2. Or rather, we talked about complete n-types in a structure M, for an n-tuple of elements: all the formulas $\varphi(x_1, \ldots, x_n)$ that are satisfied (in M) by the tuple $(r_1, \ldots, r_n)$. Here we'll be dealing with partial types—just some of the formulas.

Indiscernibles. If two elements have the *same* complete 1-type, then they're "interchangeable", identical twins, doppelgängers—much like two roots in an extension field with the same irreducible polynomial. Two n-tuples with the same complete n-type—same idea.

Because PA structures are linearly ordered by $<$, an n-tuple $(r_1, \ldots, r_n)$ has a different type from any of its permutations. (At least if they're all distinct, which we'll assume for technical reasons.) It makes more sense to work with unordered n-tuples, $\{r_1, \ldots, r_n\}$. We'll say its type is the type of the ordered n-tuple, when the elements are ordered lowest to highest. So $r_1 < \cdots < r_n$.

A set B of indiscernibles has the property that *any two n-element subsets have the same type*. So B is like the Borg: *these* dozen Borg are just like *those* dozen Borg. Or bosons. (Except I guess you can't order Borgs, or bosons.)

That last paragraph was a little vague. To be more specific, if we have set Φ of formulas, then B is *(order) indiscernible with respect to* Φ if for every $\varphi(x_1, \ldots, x_n)$ in Φ, the truth-value of $\varphi(r_1, \ldots, r_n)$ is the same for *all* sets $\{r_1, \ldots, r_n\}$, provided only that $\{r_1, \ldots, r_n\}$ is a subset of B and $r_1 < \cdots < r_n$.

How do we get ahold of indiscernibles? Let's start with a single formula $\varphi(x_1, \ldots, x_n)$. If $\varphi(r_1, \ldots, r_n)$ is true, we color $\{r_1, \ldots, r_n\}$ blue, otherwise red. (Remember that $r_1 < \cdots < r_n$.) So we're coloring n-element sets with 2 colors, and Ramsey's theorem kicks in. With k formulas voicing their opinions, the true/false results give us 2^k colors. So we can get an infinite set B that is indiscernible with respect to any *finite* number of formulas. Stay tuned—the indiscernibles have a couple more plot twists coming up.

Overspill. You talked about this in Post 2. In a way, it's what inspired my friend David to ask, "What if seven is non-standard?"

And it's how we leap over the hurdle of the last section. We're going to need indiscernibility with respect to an *infinite* number of formulas. OK, use the *finite* Ramsey theorem for a *non-standard* number of formulas. If we set things up right, this will include all the *standard* formulas we care about.

For example, we could look at all formulas with Gödel numbers less than a non-standard c. What's a formula with a non-standard Gödel number? It's just a non-standard number satisfying a certain first-order formula, namely the one that encodes the assertion "n is the Gödel number of a well-formed formula". In other words, we use the codability of syntax into PA. We already know that PA is "basically the same" as ZF$\neg\infty$; doing syntax "inside PA" is a trivial corollary.

So do we now have all the indiscernibles we need? First we have to clear another hurdle. Our definition of indiscernibles talked about truth-values. Not just syntax, also semantics. That opens Pandora's box.

Truth and satisfaction. Before I get into this, I have to make two observations.

- Gödel's two most famous results: the *completeness* theorem, and the *incompleteness* theorem.
- Tarski's two most famous results: the *undefinability of truth*, and the *definition of truth*.

How neat is that? (Of course, you know that neither of these antitheses are really contradictory.)

I am also obligated, by virtue of having taken a course in Western Civilization in college, to include this quotation:

What is Truth?

Kaye does me one better. After citing Tarski's undefinability theorem, he says:

(thus: we can't get any satisfaction!)

I was disappointed not find an index entry for Mick Jagger.

Anyway, although the property "$\mathbb{N}$ satisfies the closed formula φ" cannot be expressed in L(PA), one *can* express satisfaction for formulas of "bounded complexity". Complexity can be bounded in various ways. Here, we need the satisfaction predicate for so-called Δ_0 formulas. Kaye spells out the tedious details (for this and much more) in a chapter that leads off with the words, "This is the chapter that no one wanted to have to write..."

<u>Δ_0*, bounded search, and absoluteness*</u>. OK. Let's say N is a model of PA. We aim eventually to construct an initial segment M of N that will also be a model of PA. "As above, so below": when does truth in M reflect truth in N?

To state a precise question: suppose $\varphi(x_1, \ldots, x_n)$ is a formula in L(PA), and $(a_1, \ldots, a_n)$ is an n-tuple in M. When does $\varphi(a_1, \ldots, a_n)$ have the same truth-value in M and N? This is issue of *absoluteness*.

Here's an important special case which guarantees absoluteness: if all the quantifiers in φ are bounded—of the form $\forall x < y$ or $\exists x < y$—then we say φ is Δ_0. In that case, $\varphi(a_1, \ldots, a_n)$ will be absolute between M and N for any n-tuple $(a_1, \ldots, a_n)$ belonging to M. That is, if M is an initial segment of N!

We talked, briefly, about Δ_0 formulas and absoluteness for ZF. There, the initial segment condition was called "transitivity". A bounded quantifier looks like $\forall x \in y$ or $\exists x \in y$. I said something like, with bounded quantifiers you can just crawl around inside the sets under discussion, you don't have to search high and low through the whole universe.

Well, it's pretty much the same with PA. If $\varphi(x)$ is already absolute between M and N, then $(\forall x<y)\varphi(x)$ is too. Note that the free variable in $(\forall x<y)\varphi(x)$ is y, so we have to show that M and N give the same truth-value to $(\forall x<a)\varphi(x)$, for any a in M. But we've bounded the search to elements of M! So what's going on in the rest of N can't make a difference.

So here's how the construction *almost* goes. Using the tricks already mentioned, acquire a subset B of N that is indiscernible for all standard Δ_0 formulas. Let M be the set of all elements of N that are less than some element of B. Why should we think that M *might* be a model of PA?

The heart of the matter: the Least Number Principle for M. That's just induction in a fancy costume. Let's say we have a formula $\varphi(x)$; the Least Number Principle says that if there's any a making $\varphi(a)$ true, then there's a least such a.

If $\varphi(x)$ is Δ_0, just say "absoluteness!" and you're done. Suppose there's some a in M making $\varphi(a)$ true, in M. By absoluteness, this same a makes $\varphi(a)$ true in N. Since the Least Number Principle holds in N, there's a *least* such a in N, which of course is no greater than the a in M we started with. Well, M is an initial segment of N, so this least such a is also in M, and it's the least a making $\varphi(a)$ in M, again by absoluteness.

But what if $\varphi(x)$ is more complicated? For example, $\forall y\exists z\,\psi(x,y,z)$? (Assume $\psi(x,y,z)$ is Δ_0.)

The trick is to use the indiscernibles in B to bound the searches. B is unbounded in M (or to sling the lingo, cofinal). Instead of saying $\forall y\exists z\ldots$, we can say

$$(\forall b_1)(\forall y<b_1)(\exists b_2>b_1)(\exists z<b_2)\ldots$$

with $b_1,b_2\in B$. Take a moment to see why: if there's a z, there's a b_2 greater than that z. Likewise, looking at all the y's amounts to the same thing as looking at all pairs $y<b_1$.

Now we use the indiscernibility of the b's. I like to think of the indiscernibles as the “sliders” in this picture:

$$\cdots\cdots x \cdots\cdots \leftarrow b_1 \rightarrow \cdots\cdots\cdots\cdots y \cdots\cdots\cdots\cdots \leftarrow b_2 \rightarrow \cdots\cdots\cdots$$

Not that you literally slide b_1 and b_2 left and right, but you can choose any two b_i and b_j with $b_i < b_j$ to replace them.

So we can just pick two *specific* b's, and not quantify over them at all! We end up with this:

$$(\forall y < b_1)(\exists z < b_2) \ldots$$

That's a Δ_0 formula, so we cry “absoluteness!” and go home.

Perhaps you smell a rat. I've gotten all this way, without using the Paris-Harrington principle. Not just that: the whole idea was that even if N satisfied the Paris-Harrington principle, M wouldn't. So far I've been giving reasons why M should reflect N. But the denouement depends on M *not* reflecting N in one crucial aspect. Time for the final plot twist.

Parameters, aka names. I've pulled a fast one. When I talked about a formula $\varphi(x_1, \ldots, x_n)$, I ignored the possibility that it might contain some constants as well as the variables x_1 to x_n. Let's say we have some constants $c_1, \ldots, c_h$. So φ really looks like this: $\varphi(x_1, \ldots, x_n, c_1, \ldots, c_h)$.

Typically, when one talks about n-types, one permits constants. Not only that, but one usually enriches the language to include a constant for every element of the domain of the structure. People often call these sort of constants *names*, and when used in a formula, *parameters*. (You may recall that when your friend Joel David Hamkins gave a précis of *type in a model* in Post 2, he allowed for parameters.)

You might have noticed I've already used parameters. When I rewrote $\forall y \exists z \ldots$ as $(\forall y < b_1)(\exists z < b_2) \ldots$, b_1 and b_2 are parameters. Also, parameters are permitted in the induction axioms. That means that the argument I just gave doesn't quite work. We need to strengthen our indiscernibles to so-called *diagonal indiscernibles*. And we'll need the Paris-Harrington principle to get these—the plain old Ramsey theorem just won't cut it.

For diagonal indiscernibles, the truth-value of $\varphi(r_1, \ldots, r_n, c_1, \ldots, c_h)$ is the same no matter what the r's, *provided* that $\{r_1, \ldots, r_n\}$ is a subset of B, that $r_1 < \ldots < r_n$, and that all the c's are less than all the r's. (And of course that φ belongs to the given set Φ of formula—Δ_0 formulas in this case.)

With the new requirement (all the c's less than all the r's) the proof that M is model of PA goes through without a hitch. The Paris-Harrington principle also has a new requirement: the number of indiscernibles must be greater than the least indiscernible. You can sort of see that the two conditions might be related. But I won't lie: an intricate trusswork of combinatorics forms the bridge between the two demands.

Time to wrap things up. Say N is a non-standard model of PA satisfying the Paris-Harrington principle. Putting all the characters above into the same scene, they conjure up a new character—a nonstandard number w. This w is a "witness" to the Paris-Harrington principle; that is, the set of numbers less than w is sufficiently large to contain the indiscernibles we need, guaranteed! Not only that, w is the least such witness. And these indiscernibles call forth the model M. (Some parameters play a role, but I'm just giving the sense of the plot.) Most important: all elements of M are less than w. (Just review how we went from w to indiscernibles to M.)

Does M satisfy the Paris-Harrington principle? Nope—if it did, it would have its own w_0, the least witness to the Paris-Harrington principle! And w_0 would be strictly less than w. But an absoluteness argument shows that "the least witness to the Paris-Harrington principle" must describe the same number in both M and N. So M is a model of PA + not-(Paris-Harrington principle), and we can ring down the curtain.

10 Post 10

JB: So, last time you sketched the proof of the Paris-Harrington theorem. Your description is packed with interesting ideas, which will take me a long time to absorb. I'd like to ask some questions about them. But for starters I'd like to revert to an earlier theme: how questions about the universe of sets cast their shadows down on the world of Peano arithmetic.

It seems a certain class of logical principles let us construct fast-growing functions $f\colon \mathbb{N} \to \mathbb{N}$: functions that grow faster than any function we could construct without assuming these principles. And these logical principles are also connected to large countable ordinals. For example, there's a simple theory of arithmetic called PRA, for 'primitive recursive arithmetic', that's only powerful enough to define primitive recursive functions. There are various tricks for coding up countable ordinals as natural numbers and defining the operations of ordinal arithmetic using these code numbers, but PRA is only strong enough to do this up to ω^ω. So, we say it has proof-theoretic ordinal ω^ω. Corresponding to this limitation (in some way I don't fully understand, but can intuit), there are functions that grow too fast to be described by PRA, like the Ackermann function.

There's a whole hierarchy of more powerful theories of arithmetic with larger proof-theoretic ordinals, nicely listed here;

> Wikipedia, Ordinal analysis.

Presumably theories with larger proof-theoretic ordinals can define faster-growing functions. What's the precise connection? I know that you can use countable ordinals to index fast-growing functions, via the fast-growing hierarchy, so there's an obvious conjecture to be made here.

Anyway, the time we reach PA the proof-theoretic ordinal is ϵ_0. There are still functions that grow too fast for PA, like Goodstein's function:

> Andrés Caicedo, "Goodstein's Function" [1]

What does "too fast for PA" mean? I mean it's a partial recursive function that PA can't prove is total.

I know Goodstein's function is somehow related to Goodstein's theorem, and you seemed to say that's in turn related to the Paris-Harrington theorem. So I'm wondering: is the Paris-Harrington principle provable in some theory of arithmetic that has proof-theoretic ordinal a bit bigger than ϵ_0?

If so, what is this ordinal?

Here's another way to express my puzzlement. The Paris-Harrington principle is connected to a fast-growing function. Wikipedia says:

> The smallest number N that satisfies the strengthened finite Ramsey theorem is a computable function. . . , but grows extremely fast. In particular it is not primitive recursive, but it is also far larger than standard examples of non-primitive recursive functions such as the Ackermann function. Its growth is so large that Peano arithmetic cannot prove it is defined everywhere, although Peano arithmetic easily proves that the Ackermann function is well defined.

What ordinal indexes the growth rate of this function, via the fast-growing hierarchy?

MW: Let me first ask you about your phrasing: "Presumably theories with larger proof-theoretic ordinals can define faster-growing functions." By "define", do you mean, "has a formula for, *and* can prove that the formula defines a total function"? I'd use "define" to mean just the first clause. I'd like to keep separate *what we can prove* from *what we can express*.

JB: I was being vague, mainly because I don't really know what I'm talking about. But okay, let's say I meant "has a formula for, *and* can prove that the formula defines a total function". That sounds like what I should have said.

MW: I think it's also time to mention the Church-Kleene ordinal ω_1^{CK}, which looks down from a lofty height on *all* the proof-theoretic ordinals. Past ω_1^{CK}, it seems we don't even have a decent way to *talk* about the ordinals (formally in L(PA)), let alone prove things about them. (Or so I've read.)

I don't know proof theory nearly as well as model theory. That's something I hope to remedy as this series continues. Let me say a few things I'm fairly confident about, plus some guesses.

Things can get finicky, especially low down among the ordinals—many of the different choices seem to "wash out" by the time we get to ϵ_0. You mention PRA, the theory of primitive recursive functions. This theory has no quantifiers! Instead, it has a function symbol for every primitive recursive function. I'd rather think about $I\Sigma_1$, which is like Peano arithmetic, except we have the induction axioms only for Σ_1 formulas. (Recall that a formula is Σ_1 if it consists of an existential quantifier in front of a formula with only bounded quantifiers.) A modeler of PA would find this an attractive theory: Σ_1 formulas are upwards absolute between end extensions.

Now, $I\Sigma_1$ and PRA both have the same proof-theoretic ordinal: ω^ω. Does this mean they're "basically the same"? My guess is yes, at least in the sense of "mutual interpretability" (like PA and ZF$\neg\infty$). Maybe they're even synonymous [17]! (Or what Ketland called "definitionally equivalent".)

But what, exactly, is the meaning of "proof-theoretic ordinal"? (Let's stick with theories in the language of PA.) This could mean: the smallest recursive ordinal α for which the theory *cannot* prove the validity of transfinite induction up to α (and hence can prove it up to any $\beta < \alpha$). (That's the definition in the Wikipedia article on Ordinal analysis.) Or it could mean: the smallest α for which transfinite induction up to α (plus some very basic stuff) proves the consistency of the theory. (That's the definition in Rathgen's "Art of Ordinal Analysis" [13], a paper you've cited in the past.) For "very basic stuff", Rathgen chooses ERA=Elementary Recursive Arithmetic, a subsystem of PRA.) I imagine you prove the equivalence of these two definitions—whenever they *are* equivalent—using arguments patterned after Gentzen's proofs for ϵ_0.

Then we have the various function scales: the Hardy, Wainer (aka fast-growing), and slow-growing hierarchies. You might want to say that the proof-theoretic ordinal is the smallest α for which the theory cannot prove that Hardy's h_α is total (and hence it can prove that h_β is total for any $\beta < \alpha$). I've seen that definition in various places; here's a stack-exchange question about it.

Here's a puzzle for you (for me too): glance at the definition of the Ackermann function. Say the two-place version. As the Wikipedia article notes, the value for $A(m, n)$ depends values of $A(m^*, n^*)$ with (m^*, n^*) preceding (m, n) in the lexicographic ordering—looks like a job for transfinite induction up to ω^2! Yet the Ackermann function outpaces every primitive recursive function, and as we've read, PRA handles induction up to any ordinal less than ω^ω.

Things are more clear-cut for ϵ_0. First off, Con(PA), induction up to ϵ_0, the Paris-Harrington principle, the generalized Goodstein theorem, a version of the Hydra theorem, and the Kanamori-McAloon principle (don't ask) are all *equivalent* over PA. The Hardy and Wainer (aka fast-growing) hierarchies "catch up" to each other, so it doesn't matter which we use to define sub-ϵ_0 growth rate. The Buchholz-Wainer theorem says that a function is provably recursive if and only if it appears in the Wainer (or Hardy) hierarchy somewhere before level ϵ_0.

So to answer your question about growth rates: the Paris-Harrington function (like the Goodstein function) belongs to level ϵ_0 of the fast-growing hierarchy (also the Hardy hierarchy). OK, what about the proof-theoretic ordinal of the theory you get by adding any of these to PA? I'm not so sure. It should be the "next" proof-theoretic ordinal after ϵ_0, but I'm pretty sure that's not ϵ_0+1: once you've got transfinite induction up to ϵ_0, it's no sweat to push it a bit further. My first guess would be ω^{ϵ_0}, except that's equal to ϵ_0! Maybe ω^{ϵ_0+1}? Maybe ϵ_1?

It might be fun/useful to think through these matters out loud sometime.

JB: In my readings I got the impression that the two definitions of proof-theoretic are *not* equivalent. I thought that the "smallest α for which transfinite induction up to α (together with some basic stuff) proves the consistency of the theory" definition was a naive one inspired by Gentzen's proof of the consistency of PA using induction up to ϵ_0. I thought the "smallest α for which the theory *cannot* prove the validity of transfinite induction up to α" definition was the one that people actually use these days.

For example, people say that ZFC has a proof-theoretic ordinal, some countable ordinal below the Church-Kleene ordinal that's so large nobody has figured out how to talk about it. But I've never seen anyone claim that you could prove the consistency of ZFC using transfinite induction up to some whopping big countable ordinal (together with some other basic stuff). If this is true I'd be very interested!

I'll ask on mathoverflow.

MW: My impressions track yours pretty closely. Although I don't know that any one definition has driven all others off the field. As long as there are thesis topics to assign and papers to write, people will investigate all conceivable definitions, and try to suss out how they relate.

I mentioned my favorite definition above, which coincidentally is the one Noah Schweber mentions in his answer to your mathoverflow question:

> The *computational* proof-theoretic ordinal $|T|_{\text{comp}}$ of T is the supremum of the computable ordinals α such that there is some

notation n for a which T proves is well-founded.

As he points out, the "consistency" definition has some rather serious problems. Pohler's Proof Theory says:

> ...it is tempting to regard the order type of the shortest primitive recursive well-ordering which is needed in a consistency proof for a theory T as characteristic for T. That this idea is malicious was later detected by Georg Kreisel...

who cooked up a way to make the "consistency" ordinal ω for any "reasonable" theory. (Noah Schweber says "pathological" and "appropriate" instead of Pohler's "malicious" and "reasonable". I'm guessing his "pathological" example is the same one that Pohler describes, but I haven't checked.)

By the way, you seem skeptical that one could prove the consistency of ZF using induction up to some countable ordinal (on top of some basic stuff). I can't say I have an intuition on this, one way or the other. Asserting the consistency of *any* recursively axiomatizable theory is, after all, a purely combinatorial (i.e., syntactic) claim. Con(ZF) is but a small pale shadow, cast down into the finite realm by the magic mountain of V.

I'm mostly curious about how the computational proof-theoretic ordinals relate to the growth-rate hierarchies. Maybe we'll revisit this when I've spent more time boning up on proof theory.

JB: Yes, let's move on to Enayat! I suspect we'll loop back to these themes at various points.

References

[1] Andrés Caicedo. Goodstein's Function. URL.

[2] Ali Enayat. Standard models of arithmetic. In *Festschrift volume for Christian Bennet*, September 2014.

[3] R. L. Goodstein. On the restricted ordinal theorem. *Journal of Symbolic Logic*, 9(2):33–41, 1944. URL.

[4] Akihiro Kanamori and Kenneth McAloon. On Gödel incompleteness and finite combinatorics. *Annals of Pure and Applied Logic*, 33:23–41, 1987. URL.

[5] Matthew Katz and Jan Reimann. *An Introduction to Ramsey Theory: Fast Functions, Infinity, and Metamathematics*. AMS, 2019. URL.

[6] Richard Kaye. *Models of Peano Arithmetic*, volume 15 of *Oxford Logic Guides*. Oxford University Press, 1991.

[7] Jussi Ketonen and Robert Solovay. Rapidly growing ramsey functions. *Annals of Mathematics*, 113:267–314, 1981. URL.

[8] Laurie Kirby and Jeff Paris. Accessible Independence Results for Peano Arithmetic. *Bull. London Math. Soc.*, 14:285–293, 1982.

[9] Roman Kossak and James H. Schmerl. *The Structure of Models of Peano Arithmetic*, volume 50 of *Oxford Logic Guides*. Oxford University Press, 2006.

[10] Angus Macintyre. The history of interactions between logic and number theory. URL.

[11] Martin Loebl and Jaroslav Nešetřil. An Unprovable Ramsey-type Theorem. *Proc. AMS*, 116:819–824, 1992. URL.

[12] Jeff Paris. A Mathematical Incompleteness in Peano Arithmetic. In Jon Barwise, editor, *Handbook of Mathematical Logic*, volume 90, pages 1133–1142. North-Holland, 1977. URL.

[13] Michael Rathgen. The Art of Ordinal Analysis. *ICM 2006*.

[14] Michael Shulman. Set theory for category theory. *arXiv*, 2008. URL.

[15] John Stillwell. *Roads to Infinity*. A K Peters, 2010.

[16] Jan van Plato. Gödel, Gentzen, Goodstein: The Magic Sound of a G-String. *Mathematical Intelligencer*, 36(4):22–27, 2014. URL.

[17] Albert Visser. Categories of theories and interpretations. URL.