Review of: "Nonlinearity and Illfoundedness in the Hierarchy of Large Cardinal Consistency Strength"

Sakaé Fuchino[1]

1 Kobe University

Potential competing interests: No potential competing interests to declare.

Joel Hamkins' NONLINEARITY AND ILLFOUNDEDNESS IN THE HIERARCHY OF LARGE CARDINAL CONSISTENCY STRENGTH is really an inspiring read. As Peter Holly also writes in his review, the second part of the paper from section 7 on is highly recommended for a broad audience in mathematics and philosophy of science who are interested in the role of modern research of set theory, in connection with the foundational questions in mathematics in particular.

This being said, I have to confess my enormous difficulty in the technical details of the first part of the paper. At the moment the difficulty is not yet totally resolved. The following is thus merely a primary report of my struggle with this paper. I shall update this report later with more comments when I obtain better understanding of the material. Meanwhile I hope this version of the report is helpful for those who have difficulties similar to mine in this paper.

My understanding of statements about relative consistency in set theory is that they should be (at least in the end) formulated in metamathematics.

In meta-mathematics there is no semantics. We can only talk about provability which is expressed in statements like "there is/is not a (concretely given) proof p such that $_0 \vdash$^p φ" where $_0$ here is some base theory (e.g. some fragment of a second order extension of PA or some weak set theory) in which we would like to code the logic. $_0$ may be chosen such that we can also formulate the notion of models, model relation etc., and prove Completeness and Incompleteness Theorems (as theorems in $_0$).I n such theory $_0$, we have at least four different notions of "truth": ① : "φ holds" ($_0 \vdash$ φ), ② :"($_0$ thinks) φ is provable in " ($_0 \vdash$ ⌜⌜ ⌝⌝┣ ⌜φ⌝), ③ : "($_0$ thinks) φ holds in a model " ($_0 \vdash$ ⊧ "φ"), ④ : "($_0$ thinks) thinks φ is provable in " ($_0 \vdash$ ⊧ "⌜⌜ ⌝⌝┣ ⌜φ⌝"). Actually we can easily extend this list ad infinitum by adding "($_0$ thinks) a model thinks that φ holds in a model N_0" ($_0 \vdash$ ⊧∃N_0(N_0⊧"φ")),... etc.

Cohen's result saying "ZFC + ¬CH is consistent", for example, is to be understood as a meta-mathematical statement: "If ZF is consistent then ZFC + ¬CH is also consistent". More precisely, this should mean that there is a mechanical procedure with which a given proof p of inconsistency in ZFC + ¬CH (if such a p ever exists) can be recast into a proof p' of inconsistency in ZF. Not all experts of logic are aware of this. I remember that now almost 20 years ago when I gave an introductory tutorial on forcing, a prominent Japanese proof theorist, who was among the audience, could not believe this statement and I had to explain how to see it in length.

Now, a statement like

“PA + Con(S) + ¬ Con(T) is consistent”

should mean in metamathematics that there is no (concretely given) proof for the (concretely given) theory PA + Con(⌜⌜ ⌝⌝) + ¬ Con(⌜⌜ ⌝⌝) where and are concretely given theories, both of which extend a concretely given theory for which the consistency of + Con(⌜⌜ ⌝⌝), or sometimes even the consistency of + Con(+ Con(⌜⌜ ⌝⌝)) is assumed and from which at least the consistence of PA + Con(⌜⌜ ⌝⌝) and that of PA + ¬ Con(⌜⌜ ⌝⌝) follow.

If I interpret it correctly, the same statement in Hamkins' narrations means “ $_0$' ⊢ Con(⌜⌜PA⌝⌝ + Con(⌜⌜S⌝⌝) + ¬ Con(⌜⌜T⌝⌝))” where $_0$' is some strong enough extension of the base theory which may be assumed to be consistent (I say “assumed” since I can not say any more due to the Second Incompleteness Theorem). Working in $_0$' he then uses the Completeness Theorem (as a theorem in $_0$') and take a model of ⌜⌜PA⌝⌝ + Con(⌜⌜S⌝⌝) + ¬ Con(⌜⌜T⌝⌝) in $_0$' (note that ⌜⌜PA⌝⌝ here is not the PA in but what $_0$' think is ⌜⌜PA⌝⌝).

It seems, when a conclusion like “ * is consistent” is obtained in Hamkins' setting, what is actually attained is sometimes not the statement “ * is consistent” in metamathematics but rather “ $_0$' ⊢ Con(⌜⌜ *⌝⌝)”. In many cases, this creates no problem for metamathematical consideration because of the following:

Lemma ℵ1. If $_0$ ⊢ Con(⌜⌜ $_1$⌝⌝), then $_1$ is consistent (in metamathematics) assuming that $_0$ is consistent.

Proof. Suppose $_1$ is not consistent. This means that there is a proof such that $_0$ ⊢^ 0≡1. This can be translated to $_0$ ⊢ ⌜⌜ $_1$⌝⌝ ⊢^⌜ ⌝ 0≡1.
By the assumption of $_0$ ⊢ Con(⌜⌜ $_1$⌝⌝), it follows that $_0$ ⊢ 0≡1. This is a contradiction to the assumption of consistency of $_0$. □

Even though we can reinterpret Hamkins' arguments as a corresponding metamathematical statement via Lemma ℵ1 above, some details of his proofs remain extremely difficult to understand for me. To explain my difficulty, let me try to analyze the following paragraph from the proof of Theorem 2:

[The last but one paragraph of the proof of Theorem 2.]:“Since σ is not refutable, it follows that PA+Con(PA)+σ is consistent, and so it is also consistent with the assertion of its own inconsistency ¬Con(PA+Con(PA)+σ). In any model of this combined theory, σ is refutable in PA+Con(PA), but since also σ is true there, there must not be any smaller refutation of τ. Since this syntactic situation will be provable in PA, it follows in light of what the sentences assert that the model thinks that PA proves that σ is true and τ is false. So from Con(PA) it follows both that Con(PA + σ) and ¬ Con(PA + τ) in this model.”

[My reading of the paragraph]: We work in $_0$. Where I assume that $_0$ implies Con(PA+Con(PA)). Since σ is not refutable in PA+Con(PA) (i.e. PA+Con(PA) ⊬¬σ), PA+Con(PA)+σ is consistent. By the Second Incompleteness Theorem (formulated in $_0$), it follows that *:=PA+Con(PA)+σ+¬Con(PA+Con(PA)+σ) is consistent. By the Completeness

Theorem, there is a model ⊧ *. By ⊧ ¬Con(PA+Con(PA)+σ), we have ⊧PA+Con(PA)┣¬σ. By the choice of σ (and since ⊧PA), this means

⊧ PA+Con(PA) ┣ ∃p(PA+Con(PA)┣^p¬τ ∧ ∀q p(PA+Con(PA)┣/^q¬σ)) ...

My reading has been stuck at this point and I could go any further from here. First when I learned from Taishi Kurahashi that corresponding formal proof (in meta-mathematics) uses the formalized Σ_1-completeness, I came to the following alternative proof. Here the Σ_1-completeness is the following statement.

Proposition ℵ2. (Formalized Σ_1-Completeness) For an arithmetical Σ_1-sentence θ, we have

PA' ⊢ ⌜⌜ ⌝⌝┣⌜θ⌝→(⌜⌜ ⌝⌝┣⌜θ⌝). □

Note that the above can be reformulated as

PA' ⊢ ⌜⌜ ⌝⌝┣⌜θ⌝→ ¬ Con(⌜⌜ ⌝⌝+⌜¬θ⌝).

PA' denotes here a theory which might be slightly stronger than PA with a second-order part which is strong enough accommodate a reasonable model theory. It is often assumed that PA satisfies this Σ_1-Completeness. But for me this assumption seems to entail a bit more than the very strict variant finitary standpoint in metamathematics. Proposition ℵ2 is a theorem for a weak second-order arithmetical system which has the same Σ_1 part as PA. However, this Σ_1 equivalence is established using model theoretic method which seems to exceed the very strict version of finitary standpoint. This is one of the things I am not yet quite sure at the moment and on which I am in an on-going discussion with Hiroshi Sakai. For simplicity, however we assume in the following PA'=PA also satisfies Proposition ℵ2.

[An alternative to my reading of the paragraph]: We consider

*:=PA+Con(PA)+τ+¬Con(PA+Con(PA)+τ)

(Note that we use τ in place of σ!). Working in *, ¬Con(PA+Con(PA)+τ) means PA+Con(PA)┣ ¬τ. Note that ¬τ is a Σ_1-formula by definition of τ given in the paper. By the Formal Σ_1-Completeness, it follows that PA+Con(PA)┣ ¬Con(PA+τ) (If we switch in the proof of the Formal Σ_1-Completeness here, a model-theoretic argument is deployed at this point).

The previous paragraphs in Hamkins' proof of Theorem 2 translates to PA+Con(PA+Con(PA))┣ Con(PA+σ). Thus, by Lemma ℵ1 above, we can conclude that PA+Con(PA+σ)+¬Con(PA+τ) is consistent. □

Theorems 3,4 also should be able to be treated in this manner.

At the moment I cannot yet correctly work out Theorem 18 which deals with what is called “cautious enumeration” of a theory. In the proof of the theorem, finite subsets of a rearranged enumeration of a theory are treated. If we talk about a finite fragment of a theory , the finiteness here can have several different meanings. It can mean concretely given finite collection of formulas in the sense of metamathematics but it can also mean finite set of the set ⌜⌜ ⌝⌝ where ⌜⌜ ⌝⌝ is the set defined in $_0$ by the same recursive definition as the one given in metamathematics to decide which formula belongs to .

The finiteness here is what the base theory $_0$ thinks is finite. If we are working in the base theory $_0$ and consider a model of some set theory there, then ⌜⌜ ⌝⌝^ and ω^ can be totally different from what $_0$ thinks are and ω, since can contain nonstandard numbers and formulas.

This subtle distinction seems to be quite relevant here since it seems that Lévy-Montague Reflection Theorem is applied in the proof. The Lévy-Montague Reflection Theorem is famous for easily producing apparently correct proof of the inconsistency of the set theory if we are sloppy with the fine difference between these notions of finiteness.

There are still many other issues I want talk about in connection with this paper but I shall write them in the next update of this review.