Cahiers du Centre de logique
Volume 17

The Paris-Harrington Theorem in an NF context

by

T. E. Forster
Cambridge University

Preface

The text which follows is a pantomime horse. The front part is a talk I wrote for the mathematics seminar at the University of Canterbury in second semester 2008 — in the expectation of a cancellation that in fact never came. It was a tutorial written partly for a general mathematical audience and partly to answer some questions of Bill Taylor's, namely:

(a) *to explain how to derive* P-H *by a compactness argument from Finite Ramsey;*

(b) *to illustrate some of the ways in which* P-H *is stronger than Finite Ramsey (tho' the proof of independence from* PA *is not covered);*

(c) *to say something about the unstratified nature of* P-H *and what its significance might be.*

(c) *was rather peripheral to the front part, but is central to the hind part. The hind part is a talk prepared for OHYAST in Brussels in october 2008, and I am grateful to the organisers of that conference for reproducing these notes in the* Cahiers. *It is the result of my finally analysing* P-H *in an* NF *context.* P-H *exploits the notion of a "relatively large" set of natural numbers. The fact that this is clearly an unstratified notion has the makings of an embarrassment for* NF *studies, since NFistes always blithely assert that all of arithmetic is stratified and can be done in* NF *without any worries. Although the compactness proof is standard I spell it out here in some detail in order to make clear how a failure of stratification might arise. The question of how stratified are the various devices that we need to exploit if we are to implement in arithmetic everything that we normally take to be part of arithmetic deserves some examination. This little essay is a report on work-in-progress.*

1. Notation

When x is a finite set of naturals $\texttt{butlast}(x)$ is x shorn of its top element.

$[X]^n$ is the set of unordered n-tuples from X.

$\alpha \to (\beta)^\gamma_\delta$ says: take a set A of size α, partition the unordered γ-tuples of it into δ bits. Then there is a subset $B \subseteq A$ of size β such that all the unordered γ-tuples from it are in the same piece of the partition.

For Paris-Harrington we will also need the notion of a relatively large subset of $\mathbb{N}$: $x \subseteq \mathbb{N}$ is *relatively large* if $|x| > \min(x)$.

2. Motivation

I am grateful to Bill Taylor for prodding me into organising and clarifying my thoughts on this question, and inviting me to give a talk about it. [1] There are many reasons why this topic is of interest. The two that seem to concern him and me most are

(a) it does not seem possible to give a straightforward proof of Paris-Harrington along the lines of — for example — the proof given by Rado of Ramsey's theorems. The compactness proof of it (which is the only one known to me) is significantly less effective even than Rado's (which itself is not completely effective since it uses excluded middle on the predicate of infinitude on sets of naturals).

1. Thanks also to Dave Turner for finding (some of the) mistakes in my Part III lecture notes on this subject.

(b) the property of *relative largeness* which plays such a central rôle in Paris-Harrington is ill-typed in at least two senses. As far as I know the only person — other than the *NFistes* — to have noticed this fact is Harvey Friedman [2], and he has not given voice to the thought — that occurs to me from time to time — that the extra strength of Paris-Harrington has something to do with this failure of typing.

3. Finite Versions of Ramsey and Compactness

There are finitary versions of Ramsey's theorem. Indeed it was a finitary version that Ramsey needed to prove his theorem about decidability of universal formulæ — he proved the infinitary version only because it was easier. (Nowadays there are much easier proofs of both finitary and infinitary versions).

The finite version we want is

Theorem 1. $(\forall mnk)(\exists j)(j \rightarrow (m)^n_k)$.

This is not the version Ramsey needed for his proof of his decidability result, but is of more interest to us here, since it is this version (not the version Ramsey needed for the decidability result) that can be spiced up to give Paris-Harrington.

It's not hard to see how one can prove $(\forall mnk)(\exists j)(j \rightarrow (m)^n_k)$ directly by careful applications of Rado's method; this method will prove $2^{2n} \rightarrow (n)^2_2$ for example — though this is far from best possible: e.g., we know $6 \rightarrow (3)^2_2$.

However it is also possible to deduce this finite version of Ramsey's theorem from the infinite version by a compactness argument. There are several reasons why Ramsey didn't do it that way. For one thing, the first appearance of compactness for predicate logic did not appear until

2. Subject: [FOM] PA Incompleteness; Sun, 14 Oct 2007 10:02:58 -0400):

"In 'relatively large', an integer is used both as an element of a finite set and as a cardinality (of that same set).

This is sufficiently unlike standard mathematics, that an effort began, at least implicitly, to find PA incompleteness that did not employ this feature, or this kind of feature."

the following year (1930) in a paper —[1]— of Gödel [3]. For another, the compactness proof is highly ineffective in that no bounds can be recovered from it; nobody in their right mind would try to do it that way unless they had an ulterior motive. We do have such an ulterior motive: since the proof of Paris-Harrington (at least the only proof known to me) proceeds by a compactness argument it is very useful to run through the compactness proof of finite Ramsey by way of a rehearsal for it. [4]

So here is the compactness proof of Theorem 1,

$$(\forall m, n, k)(\exists p)(p \to (m)^n_k).$$

Proof. — Suppose that claim is false, and that there are n, m, k in $\mathbb{N}$ such that for all $p \in \mathbb{N}$ there is a set P with $|P| = p$ and a colouring $f : [P]^m \to \{1, 2, \ldots k\}$ such that there is no set $X \subseteq P$ with $|X| = n$ and $|f``[X]^m| = 1$. Fix n, m, k, and for each p let Y_p be the set

$$\begin{aligned}\{f : f : [\![1,p]\!]^m \to [1,k] \wedge \neg(\exists X)(X \subseteq [1,p] \\ \wedge |X| \geqslant n \\ \wedge |f``[X]^m| = 1)\}\end{aligned}$$

of bad k-colourings of the m-tuples of the naturals below p. (k and m are fixed.)

(Beware: square brackets are here being used *both* to denote intervals in $\mathbb{N}$ —as in $[1,k]$— *and* to denote the set of m-sized subsets of things —as in $[X]^m$.)

For any k, the set F_k of all k-colourings of m-tuples of initial segments of $\mathbb{N}$ is countable. (Each initial segment $[1,p]$ has only a finite set of m-membered subsets and there are only finitely many ways of colouring the set of those subsets). So we can uniformly wellorder F_k. Suppose this to be done, somehow. Then, for each p, we set f_p to be the first element of Y_p in the sense of that ordering.

We are now going to define a (bad) partition π of $[\mathbb{N}]^{m+1}$ into k pieces. You are given a set $x \subseteq \mathbb{N}$ of size $m+1$ and have to decide which piece to put it into. Its last member is $p+1$ for some natural number p. $x \setminus \{p+1\}$ is now a subset of $[1,p]$ and is therefore a suitable input for f_p. $f_p(x \setminus \{p+1\})$ is

3. Theorem X: thanks to the late Torkel Franzen for the citation.

4. There are other compactness arguments to be found in the literature: for example Friedman's proof of FFF.

now a number $< k$, and that tells you which piece to put x into. (Slightly more formally, put x into the $f_{(\sup(x)-1)}(x \setminus \{\sup(x)\})$-th piece.) So π partitions $[\mathbb{N}]^{m+1}$ into k pieces. We will show that π is bad.

With a view to obtaining a contradiction suppose X to be an infinite set monochromatic for π. Let $p+1$ be a member of X (and we will want to be able to find arbitrarily large such $p+1$). Consider those $(m+1)$-tuples from $X \cap [1, p+1]$ whose last element is $p+1$. What does π do to them? It sends every such $(m+1)$-tuple x to $f_p(\texttt{butlast}(x))$, and — because X is monochromatic — all these $f_p(\texttt{butlast}(x))$ are the same, whatever x we pick up. Now *every* m-sized subset of $X \cap [1, p]$ can be turned into such an $(m+1)$-tuple by the simple expedient of sticking $p+1$ on the end, so f_p sends every m-tuple from $X \cap [1, p]$ to the same number $< k$. But that is simply to say that $X \cap [1, p]$ is a subset of $[1, p]$ that is monochromatic for f_p. Now f_p was chosen so that any set monochromatic for it was of size less than n. So $X \cap [1, p]$ is of size less than n. So — no matter how large we pick $(p+1) \in X$ — we find that $X \cap [1, p]$ has at most n members. So $|X| \leqslant n+1$ and X was not infinite, contradicting the Infinite Ramsey theorem. □

We need to make a note here of the way in which this proof is less effective than the proof of Rado's given in the previous section. It is true that Rado's proof uses excluded middle — and is therefore beyond the pale for the extremely-squeamish — but is is effective in the weak sense that, by close examination of it, we can quite straightforwardly recover bounds for witnesses to the existential quantifier. In contrast the proof we have just given does not divulge bounds in this way. The reader will not be surprised to be told that the proof we are about to give of Paris-Harrington will be similarly tight-lipped.

4. Statement and Proof of the Paris-Harrington Theorem

Theorem 2 (Paris-Harrington). *For every n, m, k in $\mathbb{N}$, there is p so large that whenever $f : [\{1, 2, \ldots p\}]^m \to \{1, 2, \ldots k\}$ there is a relatively large $X \subseteq \{1, 2, \ldots p\}$ such that $|X| \geqslant n$ and $|f``[X]^m| = 1$.*

Proof. — We argue by compactness, as above.

Suppose there are n, m, k in $\mathbb{N}$ such that for all $p \in \mathbb{N}$ there is $f : [\{1, 2, \ldots p\}]^m \to \{1, 2, \ldots k\}$ such that there is no relatively large $X \subseteq$

$\{1, 2, \ldots p\}$ such that $|X| = n$ and $|f``[X]^m| = 1$. Fix n, m, k and p and let Y be the set

$$\begin{aligned}\{f : f : [\{1, 2, \ldots p\}]^m \to \{1, 2, \ldots k\} \wedge \neg(\exists X)(X \subseteq \{1, 2, \ldots p\} \\ \wedge |X| > \min(X) \\ \wedge |X| \geqslant n \\ \wedge |f``[X]^m| = 1)\}.\end{aligned}$$

This time let Y_p be — not the set of

> colourings-that-are-bad-in-the-sense-of-lacking-a-monochromatic-set-of-size-n

but the set of

> colourings-that-are-bad-in-the-sense-of-not-having-any-monochromatic-sets-of-size-n-that-are-relatively-large.

As before, initial segments of the monochromatic set X will be monochromatic for the colourings f_p. Now sets that are monochromatic for f_p are either smaller than n or are not relatively large. By considering initial segments of X that are long enough we can take care of the first condition, so the only way they can manage to be monochromatic for f_p will be by failing to be relatively large. So, for some large j, consider the initial segment consisting of the first j elements of X. We now know that this is not relatively large, so its first element must be bigger than j. So the first element of X is at least j. But j could have been taken to be arbitrarily large. □

5. The Quantifier Prefix of Paris-Harrington

Paris-Harrington is dramatically stronger than finite Ramsey (see [2] for example) and one might well wonder whether or not there are any syntactic clues to the source of this extra strength. The feature that chiefly caught my interest in this connection is the unstratified/ill-typed nature of the property of *relative largeness*, and we will get onto that in due course.

One obvious difference between Finite Ramsey and P-H is that P-H doesn't talk about colourings of tuples from arbitrary finite *sets* but of colourings of tuples quite specifically from *initial segments of the naturals.*

However we will first get out of the way a simple observation about the quantifier prefix. First we rephrase Finite Ramsey as an assertion about colourings of tuples from a finite set:

For all n, m, j in $\mathbb{N}$
There is k in $\mathbb{N}$ so large that
For every set X of size k and
For every m-colouring χ of $[X]^j$
there is $X' \subseteq X$ with $|X'| = n$
and X' monochromatic with respect to χ.

Next we rephrase Finite Ramsey as an assertion about colourings of tuples of naturals. An *enumeration* of a set X is a bijection between X and an initial segment of $\mathbb{N}$. This enables us to extend the notion of relative largeness from sets of natural numbers to subsets of arbitrary sets Y once Y has been equipped with an enumeration: $Y' \subseteq Y$ is relatively large with respect to an enumeration e of Y iff $e``Y'$ is relatively large *tout court*. Given $X$ (as in the statement of the theorem) it is clear that once we have found $X' \subseteq X$ (as in the statement of the theorem) we can pick an enumeration $e$ of $X$ so that $e``X' = [0, n]$. X' is now relatively large with respect to e. So here is Finite Ramsey phrased as an assertion about relatively large monochromatic sets.

For all n, m, j in $\mathbb{N}$
There is k in $\mathbb{N}$ so large that
For every set X of size k and
For every m-colouring χ of $[X]^j$
there is an enumeration e of X and
there is $X' \subseteq X$ with $|X'| = n$
with X' monochromatic with respect to χ
and relatively large with respect to e.

Now that we have expressed Finite Ramsey in a syntax that is the same as that used to express P-H we are in a better position to compare them. Here is P-H.

For all n, m, j in $\mathbb{N}$
There is k in $\mathbb{N}$ so large that
For every set X of size k and
For every m-colouring χ of $[X]^j$ and
For every enumeration e of X
there is $X' \subseteq X$
with X' monochromatic with respect to χ and relatively large with respect to e.

We have set out Paris-Harrington and Finite Ramsey above in something very like Prenex Normal form. The two of them have the same matrix (the stuff after the prefix) and the prefixes

$$(\forall m, n, j)(\exists k)(\forall X)(\forall \chi)(\exists e)(\exists X') \quad \text{and} \quad (\forall m, n, j)(\exists k)(\forall X)(\forall \chi)(\forall e)(\exists X')$$

are almost the same except that the innermost part of the quantifier prefix of Finite Ramsey is $(\exists e)(\exists X')$ whereas the innermost part of the quantifier prefix of Paris-Harrington is $(\forall e)(\exists X')$. That is to say we have replaced an existential quantifier with a universal quantifier: clearly we must expect P-H to be stronger than Finite Ramsey.

However it is the subtle failure of typing exhibited by the concept of relative largeness that was my original reason for interest in this question of difference in strength, and it is to this that the hindquarters of the pantomime horse are devoted.

6. Paris-Harrington and Typing

There is a syntactic difference between Finite Ramsey and Paris-Harrington in that the latter (but not the former — at least before we doctored it to make it look more like P-H so we could compare the quantifiers prefixes) has an occurrence of the predicate *relatively large*. This is significant because this predicate is *ill-typed* — and in two quite distinct ways: there are two concepts of typing at play here. One is the kind of typing at work in typed programming languages, and that is the one we deal with first.

6.1. (data)-type-checking

In languages like `ML` there is a polymorphic type-constructor `list`. It acts on an arbitrary type α to give a type α-`list`. In turn we have `length` which will take an object of type α-`list` and output an object of type `num`. Not — the reader will observe — an object of type α-`num`. We could imagine a more strongly typed language in which `length` was instead a polymorphic object of type α-`list ->` α-`num`. A type-checker for such a language would be unable to find an α to type the concept of `relatively large` since any attempt to do so would encounter an occurs-check and would crash.

6.2. Typing *à la* Russell-Quine

The other kind of typing we consider is the kind of typing used in Russell-Whitehead and more specifically in the Quine set theories. The significance in this context is that "x is relatively large" will be unstratified for at least some implementations of `natural-number-of`.

There are many ways of implementing `natural-number-of` with a stratified formula — at least in NF(U). To each such implementation we can associate a concrete integer k which is the difference

$$(\texttt{type-of}\ `y') - (\texttt{type-of}\ `x')$$

in '$y = |x|$'. In fact:

Theorem 3. *For every concrete integer k there is an implementation of* `natural-number-of` *making '$y = |x|$' stratified with*

$$(\texttt{type-of}\ `y') - (\texttt{type-of}\ `x') = k.$$

Proof. — For $k = 1$ there is the natural and obvious implementation that sets $|x|$ to be $[x_\sim]$, the equipollence class of x — the set of all things that are the same size as x. For $k \geqslant 1$ we take $|x|$ to be $\iota^{k-1}([x]_\sim)$. (This works for all cardinals, not just natural numbers.)

For $k < 1$ we have to do a bit of work, and although the measures we use will not work for arbitrary cardinals they do work for naturals. We need the fact that there is a closed stratified set abstract without parameters that points to a wellordering of length precisely ω. The obvious example is the usual Frege-Russell implementation of $\mathbb{N}$ as equipollence classes which we have just used above with $k \geqslant 1$, but it is probably worth emphasising that we don't *have* to use the Frege-Russell $\mathbb{N}$ here; whenever we have a definable injective total function f where $V \setminus f``V$ is nonempty, with a definable $a \notin f``V$, then

$$\bigcap\{A : a \in A \wedge f``A \subseteq A\}$$

will do just as well. The usual definition of $\mathbb{N}$ as a set abstract is merely a case in point. (It may or may not be worth noting that there is no such set abstract in Zermelo or ZF!) Let's use the usual $\mathbb{N}$-as-the-set-of-equipollence-classes.

Consider $\{\iota^k(n) : n \in \mathbb{N}\}$. It is denoted by a closed set abstract and has an obvious canonical wellorder to length ω. For every inductively finite set x

there is a unique initial segment i of this wellordering equipollent to it, and the function that assigns x to that initial segment is a set. We conclude that the function $x \mapsto \bigcup^k i$ is an implementation of `natural-number-of` that lowers types by k. $\square$

Here is another proof. We can take $|x|$ to be $[y]_\sim$ for any y such that $\iota^k``y \sim x$. This gives us a natural-number-of $x$ that is $k-1$ types lower than $x$. For a natural-number-of $x$ that is $k+1$ types *higher* than $x$ take $|x|$ to be $[\iota^k``x]_\sim$. $\square$

7. Stratified and Unstratified versions of Paris-Harrington in NF

To get the compactness proof of Paris-Harrington to work we need the particular class abstract — Y_p on page 102 — to be a set. This is because we want f_p to be the first element of Y_p in the sense of a global wellordering of the union of all the Y_ps. There is a such a global wellordering all right, and of course every sub*set* of its domain will have a least element. Note the italics! Mere sub*classes* are not guaranteed to have least elements.

This holds in general, of course. In the NF context Y_p will be a set if "x is relatively large" is stratified. It will be a stratified property as long as `natural-number-of` lowers types by 1 (that is to say, $k = -1$) and not otherwise. Therefore if $k \neq -1$ we cannot run the compactness argument.

We now have a version P-H(k) of P-H for each k. Given that it is customary to use only the standard implementation of `natural-number-of`, (that takes $|x|$ to be $[x]_\sim$, the equipollence class of x) we should consider what the various P-H(k) look like once one reverts to usual practice. P-H(k) is captured by a doctored version of the syntax for P-H where "relatively large(x)" is replaced by "$|x| \geqslant T^k(\min(x))$". It remains to be determined whether the various P-H(k) for $k \neq -1$ are all equivalent. That may take a little while. For the moment we can at least establish the following, which is a fairly straightforward application of work of Friederike Körner [3].

Theorem 4

For each concrete k, NF + P-H(k) *is consistent relative to* NF.

PROOF. — In [3] it is shown that it is consistent relative to NF (in fact relative to any stratfied extension of NF) that there should be a *Körner function*, that is to say $f : \mathbb{N} \to \mathbb{N}$ such that $(\forall n \in \mathbb{N})(n \leqslant f(Tn))$.

Now let f be such a Körner function and consider the analogue of P-H that we can prove by replacing "relatively large" by '$|x| \geqslant f(T(\min(x)))$'. This allegation is stratified and can be proved by the usual compactness argument. But observe that if $|x| \geqslant f(T(\min(x)))$ then x is relatively large in the old sense (because $f(T(\min(x))) \geqslant \min(x)$), so we have proved the original unstratified version of P-H.

All right, so all this proves is that the original unstratified version of P-H is consistent relative to NF. What about the other versions, where we claim the existence of monchromatic sets x which are relatively large in the sense that $|x| \geqslant f(T^k(\min(x)))$? As it happens, the Körner function exhibited in [3] satisfies $(\forall n \in \mathbb{N})(n \leqslant f(T^k n))$ for each concrete $k > 1$ and thereby takes care of the remaining cases as well, at least where $k > 0$. The Körner function achieves this effect because in the relevant model there is n_0 such that $(\forall n > n_0)(n < Tn)$. We define the Körner function f to be $\lambda n.(n + n_0)$. If k is any concrete natural we have

$$\begin{aligned} f(T^k n) &= T^k n + T^k(T^{-k} n_0) \\ &> T^{k-1} n + T^{k-1}(T^{-k} n_0) \\ &\vdots \\ &> n + T^{-k} n_0 \\ &> n. \end{aligned}$$

The second line follows from the first because the RHS, being bigger than n_0, is $< T(\text{RHS})$ and therefore $> T^{-1}(\text{RHS})$ — which is the RHS of line 2. The third line follows from the second because RHS $> T^{-1}(\text{RHS})$ implies $T^{-1}(\text{RHS} > T^{-2}(\text{RHS})$ by the usual isomorphism property of T. And so on.

For negative values of k we use a Körner function obtained analogously in a model containing n_0 such that $(\forall n > n_0)(n > Tn)$; we define the Körner function f to be $\lambda n.(n + n_0)$ as before. □

Richard Kaye has pointed out to me that if one thinks of Paris-Harrington as an allegation about the existence of monochromatic *tuples* not *sets* then no stratification problem arises, since tuples of natural numbers are naturally coded in a homogeneous way as natural numbers.

Other versions of Paris-Harrington

The literature is full of variants of P-H where "relatively large" is replaced by "$|x| > g(\min(x))$" for various natural functions $g : \mathbb{N} \to \mathbb{N}$. How do such variants fare in the NF context? We would of course expect these P-H^g to show the same variety of strengths in the NF context as they do in their native habitat. In the NF context there is the additional complication that any such variant P-H^g of Paris-Harrington multifurcates in the same way the original ("P-H$^=$") did in Section 7. The question of whether or not these are provable in NF or consistent relative to it can be approached by means of Körner functions as in Section 7. It is a simple matter to check that for any definable arithmetic function $g : \mathbb{N} \to \mathbb{N}$ there is an Ehrenfeucht-Mostowski permutation model containing a Körner function $f : \mathbb{N} \to \mathbb{N}$ satisfying $(\forall n \in \mathbb{N})(n < f(g(n))$.

8. Concluding Random Thoughts

This pantomime horse is of course work in progress. One glaring omission from the forequarters is an exposition of the proof of independence of P-H from PA, and the hindquarters lack the corresponding discussion of how the independence is connected with the unstratified nature of P-H. I suspect there are some quite enlightening things one could say about that.

There are other details that merit attention. At present it seems to be an open problem whether or not there can be in NF a type-lowering implementation of `cardinal-of` for *all* sets. The trick used to obtain a type-lowering implementation of `cardinal-of` for finite sets will work also for wellordered sets whose sizes are sufficiently small alephs. If NCI is finite then NC is countable and therefore the size of a set of singletons^k for concrete k as big as you please. So in those circumstances we would have, for each concrete k, an implementation of `cardinal-of` making the type difference between '$|x|$' and 'x' precisely equal to k. In contrast there cannot be a type-lowering implementation of `ordinal-of` — because of Burali-Forti.

Bibliography

[1] K. Gödel. Die Vollständigkeit der Axiome des logischen Funktionenkalküls. *Monatshefte für Mathematik und Physik*, Vol. 37, pp. 349–360, 1930.

[2] J. Ketonen and R. Solovay. Rapidly growing Ramsey functions. *Annals of Mathematics*, Vol. 113, pp. 267–314, 1981.

[3] F. Körner. Cofinal indiscernibles and some applications to New Foundations. *Mathematical Logic Quarterly*, pp. 347–356, 1994.

[4] J. Paris and L. Harrington. A Mathematical Incompleteness in Peano Arithmetic. In J. Barwise, editor, *Handbook for Mathematical Logic*. North-Holland, Amsterdam, Netherlands, 1977.

Thomas Forster
Cambridge University
Department of Pure Mathematics and Mathematical Statistics
e-mail : tf@dpmms.cam.ac.uk