Weak Systems of Set Theory Related to HOL

Thomas Forster

October 23, 1994

D.P.M.M.S., University of Cambridge

Contents

1 HOL and *TST* **2**

2 Predicativity, truth definitions and consistency proofs **3**
2.1 Truth definitions 4
2.2 Consistency Proofs 6
2.3 More quantifiers 7

3 Polymorphism **7**
3.1 Extensionality 10
3.2 Automorphisms of type algebras 10
3.3 TNT: The theory of (positive and) negative types 11

4 Untyped Set Theories: KF, Mac, Z and ZF **12**
4.1 The Kaye-Specker lemma 12
4.2 KF 12
4.3 MacLane Set theory 13
4.4 Zermelo set theory 13
4.5 Zermelo-Fränkel set theory 14

5 Current developments and open problems **16**

6 References **16**

This is an later (expanded and corrected) version of a survey article appearing in the proceedings of HUG94. It is designed for the interested nonspecialist, and although it does not contain any proofs of novel results, it does contain announcements (of novel unpublished results) and proofs (of frequently underregarded trivialities). For the reader who wishes to take this material further, the chief advantage of this essay will be the bibliography, which would be quite hard for a naïve reader to assemble from scratch. I would like to thank my friend and colleague Juanito Camilleri

for the invitation which led to me writing this, and various members of the Cambridge Theory and Logic seminars for advice.

1 HOL and *TST*

In virtue of theorem 1 all the set theories that are related to HOL are related to it *via* a type system called *TST*. This system is descended from the type system of Russell and Whitehead and is due in its present form to Ramsey *op. cit.*.[1] Traditionally the initials spell *T*heory of *S*imple *T*ypes but "Typed Set Theory" is better. *TST* is expressed in a language with a type for each nonnegative integer, an equality relation at each type, and between each pair of consecutive types n and $n+1$ a relation $\in$. The axioms are an axiom of extensionality at each type

$$\forall x_{n+1} \forall y_{n+1}\ (x_{n+1} = y_{n+1} \longleftrightarrow \forall z_n (z_n \in x_{n+1} \longleftrightarrow z_n \in y_{n+1}))$$

and (at each type) an axiom scheme of comprehension

$$\forall x_1 \ldots x_n\ \exists y_{n+1} \forall z_n\ (z_n \in y_{n+1} \longleftrightarrow \phi(x_1 \ldots x_n, z_n))$$

with 'y_{n+1}' not free in 'ϕ'.

TST_k is like *TST* except that there are only k types, labelled $0, \ldots, k-1$. The "theory of negative types"[2] *TNT* (Wang [1952]) and its language are defined analogously, except that the types are indexed by $\mathbf{Z}$, the positive and negative integers. *TNT* is in some ways a more convenient theory to deal with than *TST*. The two are equiconsistent by a simple compactness argument.

For the purposes of this paper all these theories will be assumed to have the axiom of infinity

This is clearly a much simpler syntax that HOL: *TST* is what you get from HOL once you decide to implement functions and ordered pairs as sets: a lot of the type structure collapses. It doesn't even matter very much *how* you implement functions and ordered pairs as sets: *any* implementation of functions and ordered pairs as sets gives rise to an interpretation of HOL in *TST*. Accordingly we have the following:

Theorem 1 *HOL and TST are each interpretable in the other.*

It is easy to interpret *TST* in HOL because we can interpret type 0 as being any type we please and thereafter type $n+1$ is interpreted as $\alpha \rightarrow$`:bool` (where α is the type that interprets type n. The other direction is equally elementary but much more complicated. First we must note that any level of

[1] Hence Sheffer's joke: Ramified set theory (of Russell and Whitehead) became Ramsified set theory.

[2] This is actually a misnomer: strictly he should have called it the theory of *positive and* negative types.

a model of TST can be embedded in all higher levels by means of an iteration of the singleton function. For notational convenience we will write singletons functionally, so that $\iota`x =_{df} \{x\}$ and $\iota^2`x = \{\{x\}\}$ and $\iota^2``x = \{\{\{y\}\} : y \in x\}$ and so on. Then one can embed type $m$ in type $m+n$ by the function $\lambda x.\iota^n``x$. We have to reinterpret the membership relation as well. If we let[3] $RUSC(R)$ be $\{\langle\{x\},\{y\}\rangle : \langle x,y\rangle \in R\}$ we interpret $\in$ by $RUSC^n(\in)$, or $RUSC^n(\in)$ composed with some suitable iterate of ι.

To interpret HOL in TST we embed the type of individuals and the booleans in type 0. We embed `:num` in the lowest type containing an interpretation of the naturals. Now suppose we have interpreted two HOL types α and β into two TST types m and n, with $m < n$, say. We compose this interpretation for α with the iterated singleton map to get interpretations of both α and β in type n. The, by means of Wiener-Kuratowski ordered pairs we can interpret $\alpha \times \beta$ in type $n+2$. Function types are now no problem because once we have interpreted both α and β in type n every function in $\alpha \to \beta$ is an element of type $n+3$.

[*HOLE It is possible to do this in more detail!*]

This equiconsistency of HOL and TST is one of two results central to the programme of this paper. (the other is lemma 3.) Theorem 1 is the bridge between type theories and set theories and is the context in which all the rest of this paper operates.

Retaining the type structure enables you to reason in a way that is implementation independent. Thus opting for set theory is a sensible if you are out to discover what can be done if you are not squeamish about what implementation you use. Type theory captures what can be done in *all* implementations; Set theory captures what can be done if one is allowed to choose one's implementation.

2 Predicativity, truth definitions and consistency proofs

What do we mean by strength of a system? Because of the second incompleteness theorem no theory can prove the consistency of itself, and if $T \vdash Con(S)$ then $S \nvdash Con(T)$, so "$T \vdash Con(S)$" is clearly a relation we will need to consider. Another relation to consider is "there is an interpretation of S in T" which, in contrast to the last, is not asymmetrical.

Much of the early work on interpreting theories in each other goes back to the time when the canonical work on set theory was Russell and Whitehead, and is therefore informed by a type-theoretical intuition that is now, once again, the theme of the hour.

[3] This notation is Rosser's

If we can interpret T_1 in T_2 in such a way that we can prove in T_2 that every theorem of T_1 is true, then clearly $T_2 \vdash Con(T_1)$. What sort of conditions on T_2 are sufficient for us to do this? Well, we must be able to define in T_2 a truth-predicate for the expressions in the range of the interpretation. Now a truth-predicate is an inductively defined set of ordered pairs, and we can illustrate the complications which are relevant here by reference to the simplest case, namely the natural numbers $\mathbb{N}$. If we are working in a theory where we have a concept of cardinal number, and we know what 0 is, and we know what *Succ* is, we can say: x a natural number iff

$$(\forall Y)([0 \in Y \wedge (\forall y \in Y)(Succ(y) \in Y)] \rightarrow x \in Y)$$

This molecular formula capturing a property of a cardinal number contains a quantifier over—not *cardinals* but—*sets* of cardinals. This is an example of an impredicative definition.[4] This particular example is not *terribly* impredicative: we are defining a set of cardinals by quantifying over sets of cardinals not sets-of-sets-of cardinals, which would be more impredicative.

2.1 Truth definitions

I shall assume we are dealing with a one-sorted language. This is to keep things simple by preventing proliferation of indices and subscripts. To be specific, I shall spell out the details only for the simplest case where there is one binary relation ("$\in$") in addition to equality.

We will assume that our variables, rather than being x, y, z etc, are all x's with numerical subscripts.[5] This clearly makes no difference to us, *qua* language users, since it is a trivial relettering, but it does make life a lot easier for us *qua* students of the language. The subscripts are quite important. We call them *indices*. The purpose of this change in notation is to make visible to the naked eye the fact that we can enumerate the variables: it is much clearer that this is the case if they are written as "x_1, x_2 ..." than if they are written as "x, y ..." In fact I think we will also have to assume that no variable is bound more than once in any formula, and that there are no occurrences of any variable outside the scope of any quantifier that binds some other occurrence of that variable. Thus we will outlaw $((\forall x)F(x)) \vee ((\forall x)G(x))$ and $F(x) \vee (\forall x)(Gx)$ even though they are perfectly good wffs. It will make life easier later.

Naturally you expect that a notion of interpretation will crop up if we are trying to define what it is for a sentence to be true in a structure.

[4] "Impredicative" is a word due to Russell which he used to describe definitions of properties-of-widgets which make reference to objects of higher type—such as: sets of widgets, or functions from widgets to widgets. The word is unevocative to modern ears, and to understand why he used it we need to know that—at the time—Russell thought that such definitions were not legitimate, did not properly define predicates and so were *impredicative*.

[5] This is why I wanted no type subscripts!

A *finite assignment function* is a function that assigns elements of $\mathcal{M}$ to finitely many indices. Computer scientists will recognise this immediately as the logician's version of their concept of *state*. They will also recognise that partial (though not finite) assignment functions form a CPO. I have (see above) carefully arranged that all our variables are orthographically of the form x_i for some index i, so we can safely and conveniently think of our assignment function f as being defined on *indices*. [6]

Next we define what it is for a finite assignment function to satisfy a sentence p, (written "f *sat* p"). We define *sat* first of all on atomic sentences. First a word on use and mention. Notice that in

$$f \; sat \; x_i = x_j$$

we have a relation between a function and an expression, not a relation between f and x_i and x_j. This is usually made clear by putting quotation marks of some kind round the expressions to make it clear that we are mentioning them not using them. Now precisely what *kind* of quotation mark is a good question. Our first clause will in fact be something like

$$f \; sat \; `x_i = x_j' \text{ iff}_{\text{df}} \; f(i) = f(j)$$

But how like? Notice that, as it stands, it contains a name of the expression which follows the next colon:

$$x_i = x_j$$

Once we have put quotation marks round this, the i and j have ceased to behave like variables (they were variables taking indices as values) because quotation is a referentially opaque context. But we still want them to be variables, because we want the content of this clause to read, in English, something like: "for any variables i and j, we will say that f *sat* the expression whose first and fourth letters are 'x', whose third and fifth are i and j respectively (whatever i and j are in this case) and whose middle letter is '=', iff $f(i) = f(j)$". Notice (and this is absolutely crucial) that in the piece of quoted English text 'x' and '=' appear with single quotes round them and 'i' and 'j' do not. Now to achieve this, ordinary single quotes will not do. Quine invented a new notational device in [1951], which he modestly calls "corners" and which are nowadays known more usually as "Quine quotes" (or "quasi-quotes") which operate as follows: The expression after the next colon:

$$\ulcorner x_i = x_j \urcorner$$

being an occurrence of '$x_i = x_j$' enclosed in quine quotes is an expression which does not, as it stands, name anything. However, i and j are variables taking

[6] It is better practice to think of the assignment functions as assigning elements of $\mathcal{M}$ to the *indices* and write "$f`i = \ldots$", since any notation that involved the actual *variables* would invite confusion with the much more familiar "$f`x_i = \ldots$" where f would have to be a function defined on the things the variables range over.

integers as values, so that whenever we put constants (numerals) in place of i and j it turns into an expression which will name the result of deleting the quasi-quotes. This could also be put by calling it a variable name.

> *Putting quine-quotes round a compound of names of wffs gives you a name of the compound of the wffs named.*

A good way to think of quine quotes is not as a funny kind of quotation mark, for quotation is referentially opaque and quine quotation referentially transparent, but rather as a kind of diacritic, not unlike the LaTeX commands I am using to write this paper. Within a body of text enclosed by a pair of quine quotes, the symbols '$\wedge$' '$\vee$' etc. do not have their normal function of composing *expressions* but instead compose *names of expressions*. This also means that Greek letters within the scope of quine quotes are being used to range over expressions (not sets, or integers). Otherwise, if we think of them as a kind of funny quotation mark, it is a bit disconcerting to find that, as Quine points out, $\ulcorner\mu\urcorner$ is just μ. The reader is advised to read pages 33-37 of Quine [1951] where this gadget is introduced. We say $f\ sat\ \alpha$ iff$_{\rm df}$ $f\ R\ \alpha$ for every R satisfying (i) - (vii)

(i) $f\ R\ \ulcorner x_i = x_j\urcorner$ iff$_{\rm df}$ $f`i = f`j$
(ii) $f\ R\ \ulcorner x_i \in x_j\urcorner$iff$_{\rm df}$ $f`i \in f`j$
(iii) if $f\ R\ \alpha$ and $f\ R\ \beta$ then $f\ R\ \ulcorner\alpha \wedge \beta\urcorner$
(iv) if $f\ R\ \alpha$ or $f\ R\ \beta$ then $f\ R\ \ulcorner\alpha \vee \beta\urcorner$
(v) if for no g extending f does $g\ R\ \ulcorner\alpha\urcorner$ hold then $f\ R\ \ulcorner\neg\alpha\urcorner$
(vi) if there is some g extending f such that $g\ R\ \ulcorner F(x_i)\urcorner$ then $f\ R\ \ulcorner(\exists x_i)(F(x_i))\urcorner$
(vii) if for every g extending f with $i \in dom(g)$, $g\ R\ \ulcorner F(x_i)\urcorner$ then $f\ R\ \ulcorner(\forall x_i)(F(x_i))\urcorner$.
Then we say ϕ is true in $\mathcal{M}$ iff the empty partial assignment function $sat\ \phi$.

2.2 Consistency Proofs

Once we have a formal notion of truth-in-a-structure $\mathcal{M}$ and a formal notion of theorem-of-T we have the possibility of proving that all theorems of T are true in $\mathcal{M}$, and the obvious way to do this is by induction on quantifiers and connectives. There are in fact very many proofs of this kind.

Now consider two typed set theories T_1 and T_2 both with extensionality and both with an axiom scheme of comprehension at each type. T_1's axiom scheme is $\forall\vec{x}\exists y\forall z(z \in y \longleftrightarrow \phi(z,\vec{x})$ where no bound variables may appear in ϕ that are of higher type than 'y', and the axiom scheme for T_2 lacks this restriction.

One would expect that T_2 should be able to prove the consistency of T_1, and this is in fact the case. This is illustrated beautifully by the results in McNaughton and the general treatment in Wang [1952].

Another standard result with the same flavour was proved by Shoenfield and by Novak. The set theory GB is obtained from ZF by adding a scheme of class existence (so that the class of all x such that Φ exists as long as Φ contains no

bound class variables) and substituting for the replacement scheme an axiom that says that the image of a set in a class is a set. GB is consistent if ZF is. In fact Shoenfield shows that there is a primitive recursive function that will accept the Gödel number of a proof of a theorem-about-sets (in GB) and return the Gödel number of a ZF proof of the same theorem. However, if we allow bound class variables to appear in the class existence scheme we obtain a new theory, nowadays commonly called "Morse-Kelly", which is stronger than ZF. Quine [1966] is good on this point. See also Wang [1949].

One consequence of this is that, for any sensible system of type theory, one is liable to find that one can prove the consistency of any proper initial segment of it in some larger initial segment. Let us go into a little detail on this, and take the example of the construction in *TST* of a truth-definition for the theory of the bottom three types of a model of *TST*. By means of iterating the singleton relation (as in section 1) we can represent the first three types all as sets of the same level (probably level 5), and the satisfaction predicate will be a set of ordered pairs a few levels higher.

Thus we have the theorem

Theorem 2 *For all k and all sufficiently large n, $TST_{k+n} \vdash Con(TST_k)$*

This gives rise to interesting complications when we introduce *polymorphism*, which is the subject of the next section.

2.3 More quantifiers

Before we leave the subject of truth-definitions altogether we should mention Levy [1965]. In this beautiful monograph Levy makes *inter alia* the point that if we are trying to define a satisfaction relation on a set of formulæ that is not itself a recursive datatype, then the satisfaction relation we are trying to define is not itself an inductively defined set of ordered pairs, and so can be defined without appeal to a comprehension principle using quantification over objects of higher type. What he shows is the following. Let Δ_n be the class of formulæ with no more than n **unrestricted** quantifiers. Then a truth definition for Δ_n formulæ belongs to Δ_{n+1}.

This has the important consequence that if we increase the number of (unrestricted) quantifiers we allow to appear in instances of the comprehension scheme of our theory, we increase its consistency strength. See Levy [1965].

3 Polymorphism

Polymorphism is more general than we tend to think. It is really the phenomenon of theories and languages with endomorphisms and automorphisms.

There are several forms polymorphism can take. People who can read German should read Specker [1958], which is the best introduction to this topic.[7] His example is duality between points and lines in projective geometry. There is an automorphism (in fact an involution) of the language of projective geometry that swaps quantifiers over points with quantifiers over lines, and swaps "x and y intersect at z" with "z goes through x and y". Let us write this automorphism as Specker does, with an asterisk. Clearly if ϕ is an axiom of projective geometry, so is ϕ^*. Indeed * extends to a endomorphism defined on the recursive datatype of proofs, and this enables us to prove by induction on that datatype that ψ is a theorem of projective geometry iff ψ^* is. The important point is that this is **not** the same as saying that $\psi \longleftrightarrow \psi^*$ is a theorem. Specker says that this scheme of biconditionals is actually the same as adopting Pappus's theorem as an axiom).[8] Another example of an automorphism of a language is the automorphism of the language of first-order predicate calculus obtained by replacing every atomic formula by its negation. Like the projective geometry example this automorphism is an involution.

In general a theory T can have an automorphism σ (so that $T \vdash \Phi$ iff $T \vdash \sigma(\Phi)$ without this implying $T \vdash \sigma(\Phi) \longleftrightarrow \Phi$.

Elegant though these examples are, they is a little remote from our concerns here. Closer to HOL is the theory TNT, which is defined above. The language in which it is expressed has an automorphism too, like the language of projective geometry. In fact it has an infinite group of them, all generated by one which we will notate with an asterisk and which arises as follows. Simply raise every type index attached to a variable in a formula ϕ by one to obtain a new formula ϕ^*. For example, asterisk of

$$(\forall x_2 y_2)(x_2 = y_2 \longleftrightarrow (\forall z_1)(z_1 \in x_2 \longleftrightarrow z_1 \in y_2))$$

is

$$(\forall x_3 y_3)(x_3 = y_3 \longleftrightarrow (\forall z_2)(z_2 \in x_3 \longleftrightarrow z_2 \in y_3))$$

(The reason for working with TNT rather than TST at this point is to ensure that * is not an endomorphism but an automorphism, as with projective geometry)

As with projective geometry we notice that ϕ^* is a theorem of TNT iff ϕ is. As before we prove this by induction of the recursive datatype of proofs. (indeed * gives rise to an automorphism of this datatype, though this automorphism is of infinite order and is not an involution). This form of polymorphism, which is the kind we find in HOL and in the type theory of Russell and Whitehead

[7] Those who can't could consult Scott's review of it in *Mathematical Reviews*.

[8] Specker even shows how to express the conjunction of finitely many expressions of the form $\phi \longleftrightarrow \phi^*$ as another expression of that form. This depends on * being an involution and doesn't apply in the cases below.

was called by them *typical ambiguity*[9]: since the axioms (and therefore the theorems) are the same at each type, there is no need to put in the type indices.

One reaction to the fact that the theorems are the same at each type is to omit the type indices on the grounds that no information is provided by putting them in. Another is to quantify universally over the type indices. We can do this only if the type indices are variables of the language. For Russell and Whitehead and for Church they were not. Nor are they for any of the systems we consider here.

In some ways the situation is a bit like that with regard to first-order predicate calculus. In that language we cannot quantify over predicates, and so expressions like

$$(\forall F)((\forall x)F(x) \rightarrow (\exists y)(F(y)))$$

simply do not make sense. However they can be *given* a sense: the expression $(\forall x)(F(x) \vee \neg F(x))$ is a valid wff of first order logic. That means that $(\forall x)(F(x) \vee \neg F(x))$ for all choices of F. One could, if one felt like it, hang a '$(\forall F)$' in front to express this fact.

We can distinguish between polymorphisms of *theories* and polymorphisms of *models*. What we have considered so far is polymorphism of theories. We can also consider the following. Consider a model $\mathcal{M}$ of TNT. Ask yourself: if we alight on a type, can we tell which type we have alighted on? This is equivalent to the question: is there a sentence true at a unique type? For suppose there is a sentence ϕ which is true only at type n, say. Then ϕ^* is true only at type $n-1$, ϕ^{**} is true only at type $n-2$ and so on. In fact this is not a first-order property of $\mathcal{M}$, and standard model-theoretic nonsense will allow us to construct models that satisfy this without adding any axioms to TNT. Notice that even if T proves $\phi \longleftrightarrow \neg\phi^*$ T can still have models with this weak kind of polymorphism!

A much stronger kind of polymorphism is the assertion that all types look the same. If Γ is a class of formulæ in $\mathcal{L}_{TNT}$, the language of TNT, we call the scheme $\phi \longleftrightarrow \phi^*$, for $\phi \in \Gamma$, "Γ-ambiguity" or $Amb(\Gamma)$. Ambiguity for all formulæ is just Amb. The full scheme Amb is strong, and is not known to be consistent. We can prove $Amb(\Gamma)$ consistent for various natural classes Γ. I will mention only three of these, and only one will be pursued here.[10]

The historically earliest example found by anyone is the class of all formulæ that mention only two types (Grishin [1969]). The background to this is that all

[9] I know of no good reason for this term to have been replaced by 'polymorphism': people who study TST continue to use the old word. I assume this is another example of a neologism arising because people are unfamiliar with the literature. The original reference is Russell and Whitehead vol 1 end of ch 2.

[10] There is a much stronger kind of ambiguity, which is considered by Specker [1958]. A theory T is ambiguous in this strong sense if every model of T has an automorphism that respects *. Projective geometry plus Pappus's theorem has this property: every model of this theory has an automorphism that exchanges points and lines.) No version of TNT has this feature and it is not of any concern to us here.

one can say in a typed set theory with two types can be said in the first-order theory of infinite atomic boolean algebras, and this is known to be a complete theory.

For us the most important example of a Γ for which $Amb(\Gamma)$ is known to be consistent relative to TNT is the class of all formulæ of the form $(\forall x_1 \ldots x_n)\Phi$ where ϕ is built up from atomics by the usual boolean operations and restricted quantifiers in the style of Levy (*op. cit.*) **and** quantifiers $\exists x \in \mathcal{P}`y$ and $\forall x \in \mathcal{P}`y$ ($\mathcal{P}`x$ is the power set of x). This is in Kaye-Forster [1991]. In the terminology of that paper, Γ is $\Sigma_1^{\mathcal{P}}$. This result leads us directly to the material of the next section.

This section concludes with three tangential minor topics which can be safely skipped. In the next section we will see how schemes of typical ambiguity give rise to consistency results for untyped set theories.

3.1 Extensionality

Interestingly, in view of the way in which extensionality is proof-theoretically problematic, one can show that if the axiom of extensionality is weakened to allow lots of empty sets (or *urelemente*) but retained for nonempty sets (so that distinct nonempty sets have distinct members) to obtain a system which we call $TSTU$, then the axiom scheme $\phi \longleftrightarrow \phi^*$ can be added without any extra consistency strength being gained. This is in Jensen's revolutionary paper [1969]. See page 12 below.

3.2 Automorphisms of type algebras

The idea of polymorphism or typical ambiguity for a type theory is of course tied up with the idea of an automorphism of what one might call the *type algebra* of the theory under consideration. The most straightforward case is TNT. Its types are indexed by $\mathbf{Z}$ not by $\mathbb{N}$, so that the type algebra is the monad $\mathbf{Z}$. Asserting the biconditional $\phi \longleftrightarrow \phi^*$ for all ϕ has the same effect as asserting the biconditional $\phi \longleftrightarrow \phi^n$ (where ϕ^n is the result of applying n asterisks to ϕ) for all ϕ and all n. This second, more inclusive scheme is probably what one would naturally think of as an axiom scheme of polymorphism but it follows from the weaker version because the automorphism group of the type algebra $(\mathbf{Z})$[11] is cyclic. The type algebras of even quite simple elaborations of TST—consider for example the Church-style type theory with only one type constructor (namely function types) and where every type is a function type—get complicated. However it is known that the automorphism group of the type algebra of this last theory is a finitely generated simple group.[12] Note that in order to have here a concept of "higher type" which behaves like our concept

[11] Both the type algebra and its automorphism group are naturally called $\mathbf{Z}$!

[12] Conway told me this was proved by a Californian by the name of R. C. Thompson, but I have never seen it published.

of higher type in TST we will need a natural definable partial order of the type algebra.

3.3 TNT: The theory of (positive and) negative types

Theorem 2 tells us that TST proves the consistency of all its proper initial segments. This means that TNT proves the consistency of TST_k, for each k. Now we were able to infer the consistency of TNT from the consistency of TST by a simple compactness argument—any proof of an inconsistency in TNT can be reproduced inside TST_k for some k—so we know that $TNT \not\vdash Con(TST)$. If it did, we could reproduce the compactness argument inside TNT and TNT would prove its own consistency. Therefore $TNT \not\vdash (\forall k)Con(TST_k)$. Therefore TNT is ω-incomplete. Although any consistent system extending arithmetic is incomplete, the ω-incompleteness is not always this transparent. It is an open question whether or not $TNT + Amb \vdash (\forall k)Con(TST_k)$. TNT is a bit odd in other ways. Although—as we have seen—its consistency follows by a compactness argument, it has no standard model. See Forster [1989]. Although TST and TNT are equiconsistent, and TST can be interpreted in TNT (obviously!) it seems fairly clear that there is no interpretation of TNT in $TSTI$. (Certainly none that commute with *!) This shows that a result of Friedman (unpublished) to the effect that two equiconsistent *finitely axiomatisable* theories can be interpreted in one another cannot be strengthened by dropping the italicised condition.[*HOLE reference*]

We can detour here briefly to have a sketch of a proof that TNT + typical ambiguity for all formulæ implies the axiom of infinity. We can implement cardinal arithmetic at each type, and we can make various assertions about the cardinal number of the universe at each type. One thing we can prove is that $\overline{\overline{V_{n+1}}} = 2^{\overline{\overline{V_n}}}$ at each type n. Also, for any number k, we can consider the sequence $\{k, log_2(k), log_2(log_2(k)), \ldots\}$ and ask how many members of this sequence are whole numbers. Let us do this to $\overline{\overline{V_n}}$. It turns out that we need the negation of the axiom of infinity to ensure that the length of this sequence is well-defined, preferably in the form "$\overline{\overline{V_n}}$ is a natural number". Clearly, since $\overline{\overline{V_{n+1}}} = 2^{\overline{\overline{V_n}}}$ we know that the length of the sequence we obtain starting at $\overline{\overline{V_n}}$ is one less than the length of the sequence starting at $\overline{\overline{V_{n+1}}}$. In particular their lengths are of different parities (remainder mod 2). Let Φ be the assertion that the sequence obtained in this way is of odd length. So as long as we can express Φ within the theory, it looks as if we should be able to prove $\Phi \longleftrightarrow \neg\Phi^*$. It turns out that this is indeed the case, and so complete ambiguity proves the axiom of infinity.

4 Untyped Set Theories: KF, Mac, Z and ZF

4.1 The Kaye-Specker lemma

Recently Richard Kaye has proved a very useful theorem which enables us to infer the consistency of one sorted theories from typed theories with ambiguity schemes to which they are related. This is a strengthening of a theorem in Specker [1962].

LEMMA **3** Kaye, [1991].

Suppose that $\mathcal{M} = \langle M_0, M_1, M_2 \ldots \rangle$ *is a structure for the language of TST and that* Σ *is the class of formulae of the form "*$\exists x_1 \ldots x_n \Phi(x_1 \ldots x_n, y_1 \ldots y_n)$*" for* Φ *in some class* Δ *which contains all atomic formulae and is closed under conjunction and substitution of variables and contains* $\psi^*(y_1 \ldots y_n)$ *whenever it contains* $\psi(x_1 \ldots x_n)$*. Suppose further that* $\mathcal{M} \models Amb(\Sigma)$*. Then there is a structure for the signature* $\langle \in, = \rangle$ *that satisfies any* σ *of the form* $\forall y_1 \ldots y_n \Phi(y_1 \ldots y_n)$*, where the result of adding suitable type indices to* Φ *is true in* $\mathcal{M}$ *and the* $\mathcal{L}_{TST}$ *formula corresponding to* Φ *is in* Σ*.*

This is the second of the two results central to this paper. It means that whenever we can prove the consistency of $Amb(\Gamma)$ relative to TNT, we get a consistency result for a one-sorted set theory: it is the bridge between typed set theories and untyped set theories. It can now be applied to the positive results we saw in the previous section to the effect that weak versions of ambiguity were provable or consistent.

The first result, Grishin [1969], does not concern us greatly here. Applying Kaye's lemma to this gives us the consistency of a funny set theory called NF_3, but this is not of much interest.

The second result (from section 3.1), that we can consistently assume full ambiguity if we drop extensionality for empty sets, gives rise to a consistency result for a system called *NF*U. See Jensen *op. cit.*. This is much more interesting than NF_3 and is definitely one to watch. *NF*U has an able and active researcher and promoter in the form of Randall Holmes: `holmes@math.idbsu.edu`.

4.2 KF

For present purposes the most important application of Kaye's lemma to an ambiguity scheme is in Forster-Kaye [1991]. Starting from a model of $TST+Amb(\Sigma_1^{\mathcal{P}})$ we can obtain a model of the theory that Kaye and I immodestly called KF.

First, some terminology: A Δ_0 formula is one that contains no unrestricted quantifiers. Restricted quantifiers are quantifiers in the style "$(\forall x \in y)(\ldots$" and "$(\exists x \in y)(\ldots$". A formula of the language of set theory is *stratified* if it can become a formula of the language of TST by adding type indices consistently to the variables in it.

KFI(= KF+ infinity) is a theory in a one sorted language with two primitives, $\in$ and $=$. It has the following axioms

1. Extensionality: $(\forall xy)(x = y \longleftrightarrow (\forall z)(z \in x \longleftrightarrow z \in y))$
2. Empty set: $(\exists x)(\forall y)(y \notin x)$
3. Pairing: $(\forall xy)(\exists z)(\forall w)(w \in z \longleftrightarrow (w = x \vee w = y))$
4. Union $(\forall x)(\exists y)(\forall z)(z \in y \longleftrightarrow (\exists w)(z \in w \wedge w \in x))$
5. Powerset: $(\forall x)(\exists y)(\forall z)(z \in y \longleftrightarrow z \subseteq x)$
6. Infinity: *There is an infinite set.*
7. Stratified Δ_0 separation (Axiom scheme: one instance for each stratified Δ_0 ϕ): $(\forall w_1 \ldots w_n)(\forall x)(\exists y)(\forall z)(z \in y \longleftrightarrow (z \in x \wedge \phi(z)))$ (where the '$(\forall w_1 \ldots w_n)$' binds all the remaining free variables in ϕ)

4.3 MacLane Set theory

In fact a refinement extends this to a relative consistency proof of a theory trading under various names in the literature, but which I was brought up by my *Doktorvater* Adrian Mathias to call 'Mac' after Saunders MacLane, who advocated it as an adequate basis for all of mathematics. Mac is like KF except in not having the restiction to stratified ϕ in axiom scheme 7. The ϕ still have to be Δ_0 though. This was first proved by Jensen [1969] and clarified in Lake [1975]. There are also distinct proofs in Mathias [2???] and Kaye-Forster [2???]. In fact Mathias shows that all the axioms of Kripke-Platek set theory (see Barwise *op. cit.* for an axiomatisation of Kripke-Platek) can be added to Mac without gaining extra consistency strength.

4.4 Zermelo set theory

It seems that this is as far as one can naturally go without reaching much stronger systems. The next things to consider are the stronger systems formulated by allowing more quantifiers in the separation scheme but these systems have not attracted much attention and do not seem to have any proper names. The next natural step up the ladder is Zermelo set theory which is the union of all those. The be precise Zermelo is like KF except in having no restrictions whatever on the formulæ in axiom scheme 6. Zermelo is of course much stronger than Mac and KF etc. We have just seen that Mac is precisely as strong as *TST* and HOL, and it is an old result of Kemeny's [1949] that we can prove the consistency of *TST* in Zermelo. Although it is not the first set theory we meet here, Zermelo was historically the first one-sorted axiomatic set theory (if we neglect the naïve set theory of Frege), and there are two natural-looking

structures that it can be thought of as axiomatising. We need some definitions to describe them.

$$\beth_0 = \aleph_0;\ \beth_{n+1} = 2^{\beth_n}.$$

H_κ = the set of things hereditarily of size $< \kappa$. Recursively:

- The empty set is in H_κ
- If $x \subseteq H_\kappa$ and $\overline{\overline{x}} < \kappa$ then $x \in H_\kappa$

($\overline{\overline{x}}$ is the cardinality of x).

($ZF \vdash$ H_κ is a set).

$$V_0 = \emptyset;\ V_\alpha = \mathcal{P}`\bigcup_{\beta<\alpha} V_\beta$$

A set is wellfounded iff it appears in some V_α. The least α such that $x \in V_\alpha$ is the **rank** of x.

$V_{\omega+\omega}$ is then the object of interest.

Zermelo appears to be the theory of $H_{\beth_\omega}$ or $V_{\omega+\omega}$ in the sense that both these structures are natural models for it. However there are facts about small sets of low rank that can only be proved by reasoning about sets of high rank, and Zermelo does not prove the existence of sets of rank $\omega + \omega$ or greater. To reason about sets of rank $\omega + \omega$ or greater we need ZF.

4.5 Zermelo-Fränkel set theory

The difference between Zermelo set theory and Zermelo-Fränkel set theory is the axiom scheme of replacement. This axiom scheme says that anything that is the same size as a set is a set:

$$(\forall x)(\exists! y)(R(x,y)) \rightarrow (\forall w)(\exists z)(\forall y)(y \in z \longleftrightarrow (\exists u \in w)(R(u,y)$$

The motivation behind this is a belief that paradoxes are connected with big collections and so the way to avoid paradox is to ensure that large collections do not turn out to be sets. If this is true, then it should certainly be safe to suppose that anything the same size as a set is a set. It turns out that the axiom scheme of replacement has very strong consequences. In particular it has strong consequences of the kind mentioned above: in ZF we can prove termination of functions whose termination cannot be proved in Zermelo set theory.

It has been known for many years that ZF proves the consistency of Zermelo set theory. Nowadays some quite refined information is coming to light about the precise strengths of different kinds of replacement. A variant of replacement is the axiom scheme of collection:

$$(\forall x \in X)(\exists y)(\psi(x,y)) \rightarrow (\exists Y)(\forall x \in X)(\exists y \in Y)(\psi(x,y))$$

It is easy to show that collection and comprehension together imply replacement. To show that replacement implies collection assume replacement and the antecedent of collection, and derive the conclusion. Thus

$$(\forall x \in X)(\exists y)(\psi(x,y))$$

Let $\phi(x,y)$ say that y is the set of all z such that $\psi(x,z)$ and z is of minimal rank. Clearly ϕ is single-valued so we can invoke replacement. The Y we want as witness to the "$\exists Y$" in collection is the sumset of the Y given us by replacement. Notice the use of the axiom of foundation here. We use it to get a *set* of z which are ψ-related to x. This obstructs the proof of this for stratified formulæ : it is *not* the case that stratified replacement implies stratified collection. The following counterexample is due to Mathias. Consider the assertion: for every natural number n there is a set of size n consisting of infinite sets all of different sizes. This is provable in Zermelo set theory. However in, say, $Z+V=L$ we can show that there is no set which collects all these together, because the sumset of such a set would be an infinite set of infinite sets of infinitely many different sizes, and we know that Zermelo set theory does not prove the existence of such a set, since in $Z+V=L$ all sets of infinite cardinals are finite. So stratified collection is not provable in Zermelo. However Coret [1970] has shown that every stratified instance of the axiom scheme of replacement is provable in Zermelo set theory. Therefore stratified replacement does not imply stratified collection.

The motivation for Collection is a lot less obvious than the motivation for Replacement, but it is robust. Recall the concept of Δ_n formula. These contain n restricted quantifiers and as many unrestricted quantifiers as you want. The axiom scheme of collection is precisely what is need to prove a normal form theorem to the effect that all formulæ can be manipulated into a normal form where all the quantifiers are out at the front of the formula (which is standard) and that all the restricted quantifiers come after all the unrestricted quantifiers.

The first thing we notice about replacement is that it enables us to prove the existence of sets of size $\beth_\omega$ (the supremum of $\{\beth_0, \beth_1 \ldots \beth_n \ldots\}$). Consider the following recursively defined function:

$$f`0 = \mathbb{N}; f`(n+1) = \mathcal{P}`(f`n)$$

We prove easily by induction on the natural numbers that $f`n$ is defined for all n and is of size $\beth_n$. By replacement, $f``\mathbb{N}$ is a set, and $\bigcup f``\mathbb{N}$ is at least as big as any of the $f`n$.

It was noticed very early on—by Russell and Whitehead—that no construction like this seemed to be possible in their type theory.

5 Current developments and open problems

Holmes [1995] considers variants of *TST* where the types are partially ordered and whenever $\alpha \leq \beta$ there is a membership relation defined between objects of type α and objects of type β (not just when $\beta = \alpha + 1$!).

By Kaye's lemma the full scheme of typical ambiguity is equivalent to the consistency of Quine's *NF*. It is open whether this theory is consistent. It is also open whether or not the result of adding full ambiguity to *TNT* is a theory that proves $(\forall k)(Con(TST_k))$. It is also open whether or not *TNT* has an ω-model; and open whether or not *TNT* has a model in which every set is definable.

Those interested in realizability interpretations of intuitionistic systems may wish to ponder the following *aperçu* of Holmes. In situations where the innocent kind of typical ambiguity ($\vdash \phi \longleftrightarrow \vdash \phi^*$) holds one always has a simply definable bijection between proofs of ϕ and proofs of ϕ^*. But in realizability interpretations definable maps from proofs to proof are themselves proofs of conditionals so this should give us a means of inferring $\vdash \phi \longleftrightarrow \phi^*$.

6 References

Barwise, J. [1975] Admissible sets and structures, an approach to definablity theory. Springer-Verlag 1975.

Buchholz, W. and Wainer S. [1985] Provably computable functions and the fast-growing hierarchy. Logic and Combinatorics, AMS Contemporary Mathematics v 65 pp 179-198.

Church, A. [1940] A formulation of the simple theory of types. *Journal of Symbolic Logic* **5** pp. 56–68.

Coret, J. [1970] Sur les cas stratifiés du schema de remplacement. *Comptes Rendues hebdomadaires des séances de l'Académie des Sciences de Paris série A* **271** pp. 57–60.

Forster, T.E. [1989] A second-order theory without a (second-order) model. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* **35** pp. 285–6

Forster, T.E. and Kaye, R.W. [1991] End-extensions preserving power set. *Journal of Symbolic Logic* **56** pp. 323–28.[13]

Forster, T.E. and Kaye, R.W. [2???] More on the set theory KF. unpublished typescript 24pp.

[13] Errata. p 327. Line 11 should read 'and $a \in M$ such that $M \models \overline{\overline{\pi}}`a = \overline{\overline{\mathcal{P}}}`a$'. Line 13 the expression following '$M \models$' should be '$\overline{\overline{\pi}}`a = \overline{\overline{\mathcal{P}}}`a$'. Line 26 '(not just $\pi`a = \mathcal{P}`a$)' should read '(not just $\overline{\overline{\pi}}`a = \overline{\overline{\mathcal{P}}}`a$)'. Line 28 '$\pi`a$' should read '$\overline{\overline{\mathcal{P}}}`\pi`a$'.

Friedmann, H. [1992] The Phenomena of Incompleteness. AMS Centenniel publications vol II pp 49-84.

Grishin, V.N. [1969] Consistency of a fragment of Quine's NF system. *Soviet Mathematics Doklady* **10** pp. 1387−90.

Holmes, M.R. [1995] The equivalence of *NF*-style set theories with "tangled" type theories; the construction of ω-models of predicative *NF* (and more) Journal of Symbolic Logic *to appear*

Jensen, R.B. [1969] On the consistency of a slight(?) modification of Quine's NF. *Synthese* **19** pp. 250−63.

Kaye, R.W. [1991] A generalisation of Specker's theorem on typical ambiguity. *Journal of Symbolic Logic* **56** pp 458-466

Kemeny, J. [1949] Type theory vs. Set theory. Ph.D.Thesis, Princeton 1949

Lake, J. [1975] Comparing Type theory and Set theory. *Zeitschrift für Matematischer Logik* **21** pp 355-6.

Levy, A. [1965] A hierarchy of formulæ in set theory. *Memoirs of the American Mathematical Society* no 57, 1965.

Mathias, A.R.D. [2???] Notes on MacLane Set Theory. unpublished typescript.

McNaughton, R. [1953] Some formal relative consistency proofs. *Journal of Symbolic Logic* **18** pp. 136−44.

Mostowski, A. [1950] Some impredicative definitions in the axiomatic set theory. *Fundamenta Mathematicæ* v 37 pp 111-124.

Novak, I.L. [1950] A construction of models for consistent systems. *Fundamenta Mathematicæ* **37** pp 87-110

Quine, W.v.O. [1951] Mathematical Logic. (2nd ed.) Harvard.

Quine, W.v.O [1966] On a application of Tarski's definition of Truth. in Selected Logic Papers pp 141-5

Ramsey, F.P. [1925] The foundations of Mathematics *in* The foundations of Mathematics, Routledge and Keegan Paul 1931 pp 1−−61.

Rosser, J.B. and Wang, H. [1950] Non-standard models for formal logics JSL 15 pp 113-129

Russell, B. A. W and Whitehead, A. N. [1910] Principia mathematica.

Scott, D. S. [1960] Review of Specker [1958]. *Mathematical Reviews* **21** p. 1026.

Shoenfield, J. R. [1954] A relative consistency proof *Journal of Symbolic Logic* **19** pp 21-28

Specker, E. P. [1958] Dualität. *Dialectica* **12** pp. 451–465.

Specker, E. P. [1962] Typical ambiguity. In *Logic, methodology and philosophy of science*. Ed E. Nagel, Stanford.

Wang, H. [1949] On Zermelo's and Von Neumann's axioms for set theory. Proc. N. A. S. **35** pp 150-155

Wang, H. [1952] Truth definitions and consistency proofs. *Transactions of the American Mathematical Society* **72** pp. 243–75. reprinted in Wang: Survey of Mathematical Logic as ch 18

Wang, H. [1952a] Negative types. *MIND* **61** pp. 366–8.