SOME FORMAL SYSTEMS FOR THE UNLIMITED THEORY OF STRUCTURES AND CATEGORIES

by

Solomon Feferman(1)

Abstract. In the informal unlimited theory of structures and (particularly) categories, one considers unrestricted statements concerning structures such as that the substructure relation on all structures of a given kind forms a partially ordered structure, or that the collection of all categories forms a category with arbitrary functors as its morphisms. These sorts of propositions are not accounted for directly by currently accepted means. The aim of the present work is to give a foundation for the theory of structures including such unlimited statements - more or less as they are presented to us - by means of certain formal systems. The theories studied here are based on an extension of Quine's idea of stratification. Their use is justified by a consistency proof, adapting methods of Jensen. These systems are successful for the basic aim to a considerable extent, but they suffer a specific defect which prevents them from being fully successful. Some possible alternatives are also suggested.

§1. Introduction. The following are examples of informal statements in the unlimited or "naive" theory of structures and categories:

(1) The collection PO of all partially ordered structures itself forms a partially ordered structure under the substructure relation S. Symbolically: (PO,S) $\in$ PO.

(2) The collection WO of all well-ordered structures is well-founded under the end-extension relation E: (WO,E) $\in$ WF.

(3) The collection Cat of all categories forms a category, with the collection Funct of all functors as its morphisms, together with the usual composition, domain, and codomain operations: (Cat, Funct, $\circ$, D_0, D_1) $\in$ Cat.

(4) For any categories A,B the collection Funct (A,B) of all functors F: A $\rightarrow$ B forms a category, with the collection Nat(A,B) of all natural transformations between these functors as its morphisms, together with the usual composition, domain and codomain operations on such: (Funct(A,B), Nat(A,B), $\circ$, D_0, D_1) $\in$ Cat.

These sorts of statements cannot be accounted for directly in currently (generally) accepted mathematics, i.e. as formulated in systems such as ZF or its immediate extensions. But they do not have an unreasonable or "cooked-up" look. Each of them arises as a natural continuation of ordinary mathematical talk about structures (in particular, categories). Indeed, it seems unnatural in category theory to keep from making statements such as (3), (4) without restriction.

We do know more or less systematic means for paraphrasing statements of this kind which can be formulated in currently understood terms; these certainly serve to secure the applications of general theories of structures.[2] Still one feels that there should be a foundation which gives a more direct account of such statements, simply as they are presented to us. This paper studies such foundations in terms of certain formal systems.

As seen from (1), (2) the foundational issues here are quite old. Moreover, a number of attempts have been made over the years to obtain satisfactory formal systems permitting some such instances of self-application[3]; none of these has been clearly successful. The interests behind that work were primarily logical or philosophical, having to do with vague general ideas about properties. The work in universal algebra and category theory has brought a different and rather more specific interest from the direction of mathematics. The aims of a foundation for unlimited statements in these subjects can be formulated much more definitely. Then one is led more directly to proposed solutions and it is easier to test their adequacy.

A subject is given foundations by means of a formal system when it is shown how to formulate and develop the subject within the system and the system itself is shown to be justified. This may be done either directly, by showing that the principles of the system are evident for certain understood concepts, or indirectly by means of reduction to previously accepted principles. Zermelo's foundations for informal set theory provide an example of the former, his axioms being evident for the cumulative hierarchy. An example of the latter is given by the axiomatization of projective geometry and its interpretation by use of Euclidean spaces. That is typical of a subject involving objects beyond the limits of ordinary experience. The type of foundation pursued here for the informal unlimited theory of structures is of this second kind. We take it that the treatment of ordinary structures is already accounted for in formal systems S such as ZFC.

Thus our general aims here are to find a formal system S^* for which:

A(i) the unlimited theory of structures, particularly categories, can be formulated and developed in S^*;

A(ii) S^* contains reasonably strong S in which the theory of ordinary structures can be developed; and

A(iii) S^* can be shown consistent by currently accepted means.

We may also have subsidiary aims, such as that S^* should be simple and that the notions and axioms of S^* should have some kind of intuitive plausibility.

We concentrate on A(i) in §2, beginning with consideration of choice of language for S^* to formulate mathematical properties of arbitrary structures and then principles to establish statements such as (1)-(4) above. These lead us to a simple system S_1^* whose main existence axioms involve an extension of Quine's idea of stratification [Q1]. S_1^* is relatively weak; it is expanded in §3 to a system S^* which contains ZFC. The main part of §3 is devoted to a proof of the consistency of S^*. This is carried out by an adaptation of Jensen's methods in [J] to establish the consistency of a (weakened) form of Quine's system NF. Thus S^* satisfies A(ii) and A(iii).

The aim A(i) is examined first for S_1^* at the end of §2 and then for S^* and particularly category theory at the end of §3. While these theories fulfill A(i) to a certain extent, they suffer a serious defect in this respect. I believe the present attempt still has value in illustrating the character of the solution to the foundational problem which is sought here and the specific considerations used to test a proposed solution. But despite the partial success, an improved solution may have to be based on a quite different idea.

There are three Appendices. Appendix I gives a theorem on the existence of models of $L_{\infty\omega}$ sentences with indiscernibles satisfying certain prescribed conditions, needed for the consistency proof of S^* in §§3.2-3.3. This requires little modification of previous work, but the statement is new and perhaps of independent interest. Appendix II describes an alternative and simpler way to realize statements such as (1)-(3) above in certain theories TP_1 and TP_2 of structural properties relative to any universe; these develop a suggestion due to Kreisel. However there are problems for (4) in these systems and consequent difficulties for aim A(i) generally. Appendix III examines an idea to get around the defect above of S_1^* and S^*, leading to another system S_2^*. It is not known if S_2^* is consistent; this raises an interesting problem.

§2. A weak system S_1 for the unlimited theory of structures.

§2.1. Structures and their properties. Usually, when properties of structures are described syntactically, one specifies the signature (similarity type) σ of the structures considered and deals with properties expressed by sentences of an associated language L_σ. We want a language L_1^* within which we can deal conveniently with variable structures of arbitrary finite signature. A simple way to do this is first to identify structures with n-tuples of collections (n variable);

(1) $$X = (X_1,\ldots,X_n$$

Specifically, given a (many-sorted) structure having basic domains $A_1,\ldots,A_k$, relations $R_1,\ldots,R_\ell$, operations $F_1,\ldots,F_m$ and individuals $c_1,\ldots,c_p$, we regard each F_i as identified with its graph and replace each individual by its singleton.[4] The structure

(2) $$\mathfrak{A} = (A_1,\ldots,A_k, R_1,\ldots,R_\ell, F_1,\ldots,F_m, \{c_1\},\ldots,\{c_p\})$$

is then of the form (1). To say that X is a structure of the signature $\sigma_{\mathfrak{A}}$ of $\mathfrak{A}$ is easily done in the language L_1^* with basic symbols as follows:

(3) (i) variables $A,B,C,\ldots,X,Y,Z$

(ii) relations $=,\in$

(iii) binary operation $(\;,\;)$.

The last is thought of as a pairing operation, from which n-tupling is defined inductively by

(4) $$(X_1,\dots, X_{n+1}) =_{Def} ((X_1,\dots,X_n), X_{n+1})$$

The logical symbols of L_1^* are those of the predicate calculus, basically $\sim$, $\vee$, $\exists$, from which $\wedge$, $\rightarrow$, $\leftrightarrow$, $\forall$ are defined as usual, as well as unique existence $(\exists!X)(\dots)$ and restricted quantifiers $(\exists Y\epsilon X)(\dots)$ and $(\forall Y\epsilon X)(\dots)$. We also abbreviate $(\exists X_1)\dots(\exists X_n)(\_\_)$ to $(\exists X_1,\dots,X_n)(\_\_)$ $(\exists Y_1\epsilon X)\dots(\exists Y_n\epsilon X)(\_\_)$ to $(\exists Y_1,\dots,Y_n\epsilon X)(\_\_)$, etc.

Then for example the following expresses in L_1^* that X is a structure of signature $\sigma_{\mathfrak{A}}$ where $\mathfrak{A} = (A_1,A_2,F)$ with $F: A_1 \rightarrow A_2$:

(5) $$(\exists X_1,X_2,X_3)\{X = (X_1,X_2,X_3) \wedge (\forall U\epsilon X_3)(\exists Y_1\epsilon X_1)(\exists Y_2\epsilon X_2)[U = (Y_1,Y_2)]$$
$$(\forall Y_1\epsilon X_1)(\exists!Y_2\epsilon X_2)[(Y_1,Y_2)\epsilon X_3)\}$$

To formulate this more generally, we make the following abbreviations in $L_1^*(m \geq 1)$:

(6) (i) $(Y \subseteq X_1 \times \dots \times X_m) \leftrightarrow_{Def} (\forall U\epsilon Y)(\exists W_1\epsilon X_1)\dots(\exists W_m\epsilon X_m)[U = (W_1,\dots,W_m)]$

(ii) $Y \subseteq X^m \leftrightarrow_{Def} Y \subseteq \underbrace{X \times \dots \times X}_{m}$

(iii) $(Y: X_1 \times \dots \times X_m \rightarrow Z) \leftrightarrow_{Def} Y \subseteq X_1 \times \dots \times X_m \times Z$
$\wedge (\forall W_1\epsilon X_1)\dots(\forall W_m\epsilon X_m)(\exists!U\epsilon Z)[(W_1,\dots,W_m,U) \epsilon Y]$.

iv) $(Y: X^m \rightarrow Z) \leftrightarrow_{Def} (Y: \underbrace{X \times \dots \times X}_{m} \rightarrow Z)$

(v) $X = \{U\} \leftrightarrow_{Def} U \epsilon X \wedge (\forall Y\epsilon X)(Y=U)$

These should not be considered for the moment as giving any independent meaning to '$X_1 \times \dots \times X_m$', '$X^m$' or '$\{U\}$'.

Now it is obvious how to associate with each n and σ of length n a formula $Str_\sigma(X_1,\dots,X_n)$ of L_1^* which expresses that $(X_1,\dots,X_n)$ is a structure of signature σ. Following this, with each sentence θ in the <u>first-order</u> language $L_\sigma^{(1)}$ of structures of signature σ is associated a formula $\overline{\theta}(X_1,\dots,X_n)$ which expresses that $(X_1,\dots,X_n)$ <u>satisfies</u> θ. Let $\overline{\overline{\theta}}(X)$ be the formula of L_1^*,

(7) $$(\exists X_1,\ldots,X_n)\{X = (X_1,\ldots,X_n) \wedge Str_\sigma(X_1,\ldots,X_n) \wedge \overline{\theta}(X_1,\ldots,X_n)\}$$

Then $\overline{\overline{\theta}}(X)$ expresses that X is a structure of signature σ which satisfies θ.

For example, corresponding to the usual sentence θ_{Po} which holds in $\mathfrak{A} = (A,R)$ (where A is non-empty, $R \subseteq A^2$) just in case A is partially-ordered by R we have:

(8) $\overline{\overline{\theta}}_{Po}(X)$ is $(\exists X_1,X_2)\{X = (X_1,X_2) \wedge \exists U(U \in X_1) \wedge X_2 \subseteq X_1^2 \wedge \overline{\theta}_{Po}(X_1,X_2)\}$ where

$\theta_{Po}(X_1,X_2)$ is $(\forall U \in X_1)[(U,U) \in X_2] \wedge (\forall U,W \in X_1)[(U,W) \in X_2 \wedge (W,U) \in X_2 \rightarrow U = W]$

$\wedge (\forall U_1,U_2,U_3 \in X_1)[(U_1,U_2) \in X_2 \wedge (U_2,U_3) \in X_2 \rightarrow (U_1,U_3) \in X_2]$

Similarly we obtain formulas $\overline{\overline{\theta}}_{Grp}(X)$, $\overline{\overline{\theta}}_{Cat}(X)$ expressing that X has the structure of a group, resp. of a category, etc.

This kind of association of formulas $\overline{\theta}, \overline{\overline{\theta}}$ of L_1^* with sentences θ in languages of varying σ can be carried out just as well for θ in any second-order or even higher finite order language of structures of signature σ. For example, associated with the sentence θ_{WF} which holds in $\mathfrak{A} = (A,R)$ just in case $\mathfrak{A}$ is well-founded and partially ordered is the formula $\overline{\theta}_{WF}(X_1,X_2)$ which is the conjunction of $\overline{\theta}_{PO}(X_1,X_2)$ and the minimal element principle:

(9) $$(\forall Z)\{Z \subseteq X_1 \rightarrow (\forall Y \in Z)(\exists Y_1 \in Z)(\forall Y_2 \in Z)[(Y_2,Y_1) \in X_2 \rightarrow Y_1 = Y_2]\}.$$

Similarly, with any sentence ψ in first or higher-order languages expressing a relation between structures $X_i = (X_{i1},\ldots,X_{in_i})$, $1 \leq i \leq m$, we associate first a formula $\overline{\psi}(X_{11},\ldots,X_{mn_m})$ of L_1^* expressing that the relation holds between $(X_{11},\ldots,X_{1n_1}),\ldots,(X_{m1},\ldots,X_{mn_m})$ and then $\overline{\overline{\psi}}(X)$ of the form

(10) $$(\exists X_{11},\ldots,X_{mn_m})\{X = ((X_{11},\ldots,X_{1n_1}),\ldots,(X_{m1},\ldots,X_{mn_m})) \wedge \overline{\psi}(X_{11},\ldots,X_{mn_m})\}$$

which expresses that X is a member of the relation given by ψ. For example, the following formula $\overline{\overline{\psi}}_{Sub}(X)$ expresses that X is a member of the substructure relation between structures of signature $\sigma_{\mathfrak{A}}$ where $\mathfrak{A} = (A,R)$ with $R \subseteq A^2$:

(11) $\{(\exists X_{11}, X_{12}, X_{21}, X_{22})\{X = ((X_{11}, X_{12}), (X_{21}, X_{22})) \wedge X_{12} \subseteq X_{11}^2 \wedge X_{22} \subseteq X_{21}^2$

$\wedge X. \subseteq X_{21} \wedge X_{12} \subseteq X_{22} \wedge (\forall U, W \in X_1)[(U,W) \in X_{22} \rightarrow (U,W) \in X_{12}]\}.$

§2.2. Type theory with pairing; stratification. We now introduce a typed language with pairing L_{TP} in terms of which we can describe precisely finite type properties of structures of arbitrary signature. It suffices to use the simple types $0, 1, \ldots, n, \ldots$ if we take it that each type is closed under pairing. The basic symbols of L_{TP} are as follows:

(1) (i) variables of type n, $A^n, B^n, C^n, \ldots, X^n, Y^n, Z^n$ for each $n \in \omega$,

(ii) relations $=, \in$

(iii) binary operation ( ).

The sets of terms of type n of L_{TP} are defined inductively by:

(2) (i) each variable of type n is a term of type n;

(ii) if t_1, t_2 are terms of type n then so also is (t_1, t_2).

The atomic formulas of L_{TP} are just those of the form

(3) (i) $t_1 = t_2$ for type (t_1) = type (t_2) and

(ii) $t_1 \in t_2$ for type $(t_1) + 1$ = type (t_2)

The formulas of L_{TP} are generated from these using $\sim, \vee, \exists$. By L_T we mean the sublanguage of L_{TP} without the binary pairing symbol and by $L_{T(\in)}$ the further sublanguage of L_T without the equality symbol.

A formula ϕ of L_1^* is said to be L_{TP}-stratified if we can assign type superscripts to the variables of ϕ in such a way that the resulting ϕ^+ is a formula of L_{TP}; in this assignment, each variable is to receive the same type at all its occurrences. ϕ^+ is not uniquely determined by ϕ; any such ϕ^+ is called a stratification of ϕ. ϕ is said to be $L_{T(\in)}$ (resp. L_T)-stratified if ϕ^+ is an $L_{T(\in)}$ (resp. L_T) formula; this is the same as being stratified in the sense of Quine [Q1] (resp. Jensen [J]). Such notions of stratification can be extended in an obvious way to any many-sorted language for which one has specified the admitted terms and atomic formulas. We call each instance of such a stratification set-up.

The following are simple examples of L_{TP}-stratified formulas:

(4) i) $X \in Y$, (ii) $(X,X) \in Y$

The following formulas are not L_{TP}-stratified

5) (i) $X \in X$, (ii) $(X,Y) \in Z \wedge X$

Obviously the set of L_{TP}-stratified formulas is (primitive) recursive.

The main point for our purpose is that

(6) each of the formulas $\bar{\theta}, \bar{\bar{\theta}}, \tilde{\psi}, \tilde{\tilde{\psi}}$ of the kind considered in §2. is L_{TP}-stratified.

In other words, the kinds of properties of and relations between structures $X = (X_1,\ldots,X_n)$ dealt with there are all expressed by L_{TP}-stratified formulas. $(X_1,\ldots,X_n$ and X may be assigned type 1 and their elements type 0 in these formulas.)

It is natural to consider the following axiom schemes in L_{TP}, for all n:

(7) i) L_{TP}-Comprehension

$$(\exists A^{n+1})(\forall X^n)[X^n \in A^{n+1} \leftrightarrow \phi] ,$$

for each formula ϕ of L_{TP} which does not contain 'A^{n+1}'

(ii) Extensionality $(\forall X^n)[X^n \in A^{n+1} \leftrightarrow X^n \in B^{n+1}] \rightarrow A^{n+1} = B^{n+1}$.

(iii) Pairing $(X_1^n,X_2^n) = (Y_1^n,Y_2^n) \rightarrow X_1^n = Y_1^n \wedge X_2^n = Y_2^n$.

Standard models of these axioms can be formed as follows from any set M_0 on which is defined a pairing operation $(\cdot\ ,\ )_0$ and which contains at least two distinct elements a_0, b_0. One obtains for each n, a set M_n containing distinct a_n, b_n and closed under a pairing operation $(\ ,\ )_n$. To step to $n + 1$, take $M_{n+1} = PM_n = \{X | X \subseteq M_n\}$, $a_{n+1} = \{a_n\}$, $b_{n+1} = \{b_n\}$ and $(X,Y)_{n+1} = (X \times a_{n+1}) \cup (Y \times b_{n+1})$ where $X \times Y = \{(x,y)_n | x \in X$ and $y \in Y\}$. $\in$ is interpreted as the union of the membership relations between M_n and M_{n+1}.

§2.3. The system S_1^*. With each set T of axiom in L_{TP} is associated a set T_{Strat} of formulas, where

(1) $\theta \quad T_{Strat} \Leftrightarrow_{Def}$ for some stratification θ^+ of θ we have $\theta^+ \in T$.

In particular, the following scheme I and axioms II, III form T_{Strat} for the set T of axioms in §2.2(7)(i)-(iii):

(2) I. L_{TP}-Stratified Comprehension

$$(\exists V)(\forall X)[X \in A \leftrightarrow \phi]$$

for each L_{TP}-stratified formula ϕ of L_1^* in which 'A' does not occur.

Extensionality $(\forall X)[X \in A \leftrightarrow X \in B] \rightarrow A = B$.

III Pairing $(X_1,X_2) = (Y_1,Y_2) \rightarrow X_1 = X_2 \wedge Y_1 = Y_2$.

The system consisting of Extensionality and L_T-Stratified Comprehension (i.e. where I is restricted to L_T-stratified ϕ) is a version of Quine's system NF.(5) It is not known whether NF is consistent. The related system NFU shown consistent by Jensen [J] has in place of II the statement

(2) II'. Weak Extensionality

$$(\exists X)(X \in A) \wedge \forall X[X \in A \leftrightarrow X \in B] \rightarrow A = B$$

This is the form suitable for systems of set theory in which one admits "Urelementen."

We define S_1^* to be the set of axioms (2) I, II' and III. S_1^* can be shown to be consistent by a direct adaptation of Jensen's methods. Moreover, just as remarked in [J] for NFU, the system S_1^* is quite weak; its consistency can be proved in elementary arithmetic. We shall not give a separate consistency proof for S_1^*, but only for the much stronger theory S^* to be considered in §3. But S_1^* is already of interest in connection with the aim A(i), as will be shown in the next subsection §2.4. Before this, there are a number of points worth noting.

Remarks.

(i) Pairing. The existence of $\{Y_1,Y_2\}$, i.e. the statement

$$(\exists!Y)(\forall X)[X \in A \leftrightarrow X = Y_1 \vee X = Y_2]$$

is provable in NFU. Then we can define $\{Y\} = \{Y,Y\}$ and $\langle Y_1, Y_2 = \{\{Y_1\},\{Y_1,Y_2\}\}$ and prove as usual the Pairing Axiom for this definition in NFU,

$$\langle X_1, X_2 \rangle = \langle Y_1, Y_2 \rangle \rightarrow X_1 = Y_1 \wedge X_2 = Y_2$$

But there is no obvious way to use this to interpret S_1^* in NFU, for when stratifying in L_{TP} we need to assign to (Y_1, Y_2) the same type n as assigned to both Y_1 and Y_2 while by this definition, $\langle Y_1, Y_2 \rangle$ must be assigned type $n + 2$. Then, for example, in a structure $\langle X_1, X_2 \rangle$ with $X_2 \subseteq X_1^2$, using this definition of pair, X_2 must be assigned type 2 higher than that of X_1. This blocks us from realizing L_{TP}-stratified properties of structures.

On the other hand, Quine [Q_2] gave a quite different and more complicated definition, in NF, of ordered pair - denoted for the moment here as $\langle\langle Y_1, Y_2 \rangle\rangle$ - which does have the property that $\langle\langle Y_1, Y_2 \rangle\rangle$ can always be assigned the same type as Y_1, Y_2 when stratifying. However, to prove the Pairing Axiom

$$\langle\langle X_1, X_2 \rangle\rangle = \langle\langle Y_1, Y_2 \rangle\rangle \rightarrow X_1 = Y_1 \wedge X_2 = Y_2$$

one makes use of full Extensionality (2) II in an essential way. Lacking the known consistency of NF, we have taken the simple alternative of adding pairing to NFU as a primitive and extending stratification to preserve types. Of course, it must then be checked that this causes no problems for the consistency proof.

(ii) <u>An inconsistent stratified system</u>. We may ask whether the consistency of S_1^* may be established on general grounds, i.e., if T <u>is a consistent set of statements in</u> L_{TP}, <u>is</u> T_{Strat} <u>necessarily consistent in</u> L_1^*? We shall now show that the corresponding is false for a simple, at first sight, reasonable looking extension of the stratification set up. Describe the typed language with pairing L'_{TP} just as for L_{TP} in §2.2 (1)-(3), except that in the inductive definition of the set of terms of type n we permit also:

> <u>if</u> t_1 <u>is of type</u> n_1 <u>and</u> t_2 <u>is of type</u> n_2 <u>then</u> (t_1, t_2) <u>is of type</u> $\max(n_1, n_2)$.

The atomic formulas of L'_{TP} are built from these terms just by the same conditions as §2.2(3). Then it is clear what is meant by an L'_{TP}-stratified formula. The system S'_1 consisting of L'_{TP}-Stratified Comprehension and Pairing is inconsistent. For, the following are instances of the schema in S'_1:

$$(*) \qquad \begin{array}{c} (\exists A)(\forall X)\{X \in A \leftrightarrow (\exists Y,Z)[X = (Y,Z) \wedge Y \in Z]\} \\ (\forall A(\exists B)(\forall Y)[Y \in B \leftrightarrow (Y,Y) \notin A] \end{array}$$

Using Pairing we then get a form of Russell's Paradox. Now $S'_1 = T_{Strat}$ where consists of the scheme of Comprehension

$$(\exists A^{n+1})(\forall X^n)[X^n \in A^{n+1} \leftrightarrow \phi]$$

for each ϕ (without 'A^{n+1}') an L'_{TP}-formula, and the schema of Pairing

$$(X_1^{n_1}, X_2^{n_2}) = (Y_1^{n_1}, Y_2^{n_2}) \rightarrow X_1^{n_1} = Y_1^{n_1} \wedge X_2^{n_2} = Y_2^{n_2}$$

for all n_1, n_2. We obtain a model for T (even with Extensionality) on the standard domains $\langle M_n \rangle_{n\in\omega}$ with the pairing operations given at the end of §2.2 extended as follows. For each n and $m \geq n$ and $X \in M_n$, define $X^{m/n} \in M_m$ inductively by:

$$X^{n/n} = X \ , \quad X^{m+1/n} = \{X^{m/n}\}$$

This has the property that if $X^{m/n} = Y^{m/n}$ then $X = Y$. Then define

$$(X,Y) = (X^{n/n_1}, Y^{n/n_2})_n \quad \text{for } X \in M_{n_1},\ Y \in M_{n_2},\quad n = \max(n_1,n_2)\ .$$

T is consistent.

(iii) Extensionality in stratified theories. The preceding shows that even if the association of T_{Strat} with consistent T for a given stratification scheme looks plausible, the justification for using T_{Strat} must be given on other grounds. Since we do not have a clear intuitive interpretation of such theories, the only grounds we have are by means of interpretations in terms of concepts that we already understand. In particular, there is no reason to insist on full Extensionality if we cannot see how to interpret it, unless it is essential for the aim A(i); we return to the latter question below. It happens that in every known model for stratified theories, i.e. of the kind found by Jensen and developed further here, it is evident from the proof that the most we can hope to arrange for is Weak Extensionality. Other candidates for models may be imagined, for example those for the theories of properties considered in Appendix II below. But there also, Extensionality fails. Indeed, it seems whatever intuitions we have about stratified comprehension principles has to do with ideas about properties, which are non-extensional objects. This is not to deny the possibility of consistency of theories such as NF. But our interest in such shifts when we think of the range of the variables as being not classes in the usual extensional sense, but some kind of classes given by definitions. For lack of better intuition, I propose to call these meta-classes here.

(iv) Extensionality and structure theory. Is full Extensionality essential for the informal theory of structures? Weak Extensionality is certainly used at various places, for example when it is shown that distinct equivalence classes of an equivalence relation are disjoint. I do not know of an example where it would be required that two empty objects be identical. On the other hand, I do not have a convincing argument why it is unnecessary. It is convenient at times to admit empty structures. But we could get around possible non-uniqueness of such simply by designating a specific empty object 0, and considering only empty domains coinciding with 0.

§2.4. Development of the unlimited theory in S_1^*. It is convenient here and for §3 to make the following definitions. These apply to any language which contains L_1^* and some designated constant symbol 0. We assume throughout this section that

$$(1) \qquad (\forall X)(X \notin 0)$$

Definition 2.1. For any formula ϕ (in which 'A' is not free) let

$$[X|\phi] = \begin{cases} \text{the unique } A \text{ s.t. } (\forall X)[X \in A \leftrightarrow \phi] & \text{if } (\exists ! A)(\forall X)[X \in A \leftrightarrow \phi] \\ 0 & \text{otherwise.} \end{cases}$$

The formula ϕ may contain parameters, say $Y_1,\ldots,Y_m$, in which case we write this more loosely as

(2) $$[X|\phi(X,Y_1,\ldots,Y_m)] \quad \text{or} \quad [X|\phi(X)]\ .$$

This determines an operation from meta-classes to meta-classes. Given a term $t(X_1,\ldots,X_n$ which may also contain parameters, we put

(3) $$[t(X_1,\ldots,X_n)|\phi(X_1,\ldots,X_n)] =_{Def} [X|(\exists X_1,\ldots,X_n)(X = t(X_1,\ldots,X_n) \wedge \phi(X_1,\ldots,X_n))].$$

is called the <u>defining condition</u> of $[X|\phi]$

<u>Lemma 2.2.</u> <u>If ϕ is an L_{TP}-stratified formula, then (assuming (1)) it is provable in S_1^* that $[X|\phi]$ satisfies its defining condition, i.e.</u>

$$A = [X|\phi] \to (\forall X)[X \in A \leftrightarrow \phi]\ .$$

We obviously get the same for any $[t(X_1,\ldots,X_n)|\ \phi(X_1,\ldots,X_n)]$ where t is built by pairing. Each of the following has an L_{TP}-stratified defining condition.

<u>Definition 2.3.</u>

(i) $\{Y_1,Y_2\} = [X|X = Y_1 \vee X = Y_2]$

(ii) $\{Y\} = \{Y,Y\}$

(iii) $\bigcup Y = [X|(\exists Z)(X \in Z \wedge Z \in Y)]$

(iv) $Y_1 \cup Y_2 = \bigcup\{Y_1,Y_2\}$

(v) $\bigcap Y = [X|(\forall Z)(Z \in Y \to X \in Z)]$

(vi) $Y_1 \cap Y_2 = \bigcap\{Y_1,Y_2\}$

(vii) $-Y = [X|X \notin Y]$

(viii) $V = [X|X = X]$

(ix) $PY = [X|X \subseteq Y]$

(x) $Y_2{}^{Y_1} = [X|X\colon Y_1 \to Y_2]$

(xi) $Y_1 \times Y_2 = [(X_1,X_2)|X_1 \in Y_1 \wedge X_2 \in Y_2]$

(xii) $Y_1 \times \ldots \times Y_{n+1} = (Y_1 \times \ldots \times Y_n) \times Y_{n+1}$

(xiii) $Y^n = \underbrace{Y \times \ldots \times Y}_{n}$

(xiv) $DY = [X|(\exists Z)(X,Z) \in Y]$

(xv) $\breve{Y} = [(X_1,X_2)|(X_2,X_1) \in Y]$

(xvi) $D^{\smile}Y = \breve{DY}$.

(Note that the symbols in §2.1(6) are now given independent meaning.)

<u>Definition 2.4.</u> i) $\mathrm{Fun}(F) \leftrightarrow (F\colon DF \to V)$.

(ii) $F(X) = \begin{cases} \text{\underline{the unique} } Y \text{ \underline{s.t.} } (X,Y) \in F \text{ \underline{if} } \mathrm{Fun}(F) \wedge X \in DF \\ 0 \text{ \underline{otherwise}.} \end{cases}$

Now consider the formulation of statements §1(1)-(4) in S_1^*. For the first, we use the formulas $\overline{\theta}_{PO}(X)$ and $\widetilde{\psi}_{Sub}(X)$ of §2.1(8) and (11) resp., which express, resp., that X is a partially ordered structure and that X is a pair (X_1,X_2) of structures in the substructure relation. Let

(4)(1) $PO = [X|\overline{\theta}_{PO}(X)]$

(ii) $S = [Z|\widetilde{\psi}_{Sub}(Z) \wedge (\exists X_1,X_2)(Z = (X_1,X_2) \wedge X_1 \in PO \wedge X_2 \in PO)]$

5) $$S_1^* \vdash (PO,S) \in PO \; .$$

This simply proceeds by formalizing the informal proof that the substructure relation satisfies the properties to be a partial ordering. Similarly, using the formula $\overline{\theta}_{WF}(X_1,X_2)$ of 2.1(9) we define

(6) $$WF = [(X_1,X_2)|(X_1,X_2) \in PO \wedge \overline{\theta}_{WF}(X_1,X_2)]$$

It is then clear how to define WO, the metaclass of well-ordered structures, and E, the end-extension relation, in S, and to establish

(7) $$S^{\cdot} \vdash (WO,E)$$

To formulate the statements §1(3)-(4) in $S^{\cdot}$ we would have to spel out in a little more detail the basic notions of category theory. There are no obstacles to doing this for the given statements. But since we shall also want to consider the formulation of significant theorems of category theory which require stronger principles than in S_1^*, we delay all details until §3.4. It should however be remarked that the category of all functors between two given categories, mentioned in §1(4), is analogous to the (meta)class $Y_2^{Y_1}$ of all functions between two given (meta)classes Y_1, Y_2. (In addition, this is the only one in the given list of statements that requires formation of $[X|\phi]$ for ϕ with parameters.)

Thus S_1^* provides us with the means to formulate directly and verify statements of the unlimited theory of structures like those of §1. Definitions 2.1, 2.3 and 2.4

extend familiar mathematical operations and notions for classes to metaclasses, for which they continue to satisfy usual properties. It would thus seem that S_1^* can fulfill aim A(i) to a considerable extent. Nevertheless, S_1^* has a serious limitation in this respect, which we may consider to be the principal defect of L_{TP}-stratification.(6) There are many situations in which we want to consider a relation between pairs (X,Y) where X,Y are intuitively at distinct type levels. There is no problem about forming such a pair for any particular X,Y; for example, whenever A is defined in S_1^* so also is (A,PA). The problem arises when we want to consider such pairs for variable X,Y, and want to form the corresponding metaclass. For example, we frequently consider the operation Q from an object X to its equivalence class $[X]_E$ in a given equivalence relation:

(8) (i) $[X]_E = [Y|(X,Y) \in E]$

(ii) $Q = [Z|(\exists X)(X \in DE \wedge Z = (X,[X]_E))]$.

The second part of this definition is given sense by Definition 2.1, but since the defining condition is not L_{TP}-stratified we don't see how to prove in S_1^*:

(9) $$(\forall Z)\{Z \in Q \leftrightarrow (\exists X)[X \in DE \wedge Z = (X,[X]_E)]\} .$$

For, when this is written out without abbreviations, $U = [X]_E$ must be assigned a type one higher than X in any attempt to stratify.

A similar problem arises when one tries to define the direct product

$$P = \Pi_{X \in A} F(X) \quad \text{for} \quad F: A \to V$$

We may try

$$P = [G|Fun(G) \wedge (\forall X \in A)(G(X) \in F(X)]$$
$$= [G|Fun(G) \wedge (\forall X \in A)(\exists Y,Z)\{(X,Y) \in G \wedge (X,Z) \in F \wedge Y \in Z\}] ,$$

but again the defining condition is not L_{TP}-stratified (Y,Z must be assigned the same type as X for pairing but different types for membership); so we don't see how to prove that P satisfies its defining condition in S_1^*.

One way that might be thought to overcome these difficulties is to define new ordered pair operations to adjust type levels, essentially as done for the model considered in Remark (ii) of the preceding section. For example, take

$$(11)\qquad (X,Y)' =_{\text{Def}} \{X\}.Y)$$

when dealing with pairs satisfying a stratified condition in which Y must be assigned type one higher than X. We can prove in S_1^* that $(\ ,\ )^1$ satisfies the pairing axiom and, in place of (9);

$$(12)\qquad (\exists Q)(\forall Z)\{Z \in Q \leftrightarrow (\exists X)(X \in DE \wedge Z = (X,[X]_E)')\}$$

Q satisfies Fun', the property of being a function in the new sense of ordered pair, i.e. we have

$$(\forall X \in DQ)(\exists ! Y)[(X,Y)^1 \in Q] .$$

But this sort of device requires a <u>duplication of all the basic mathematical notions</u> such as relation, function, structure, etc., <u>for each new ordered pair operation that is introduced.</u>[(7)] That is just contrary to the kind of theory which is sought here, which makes no <u>a priori</u> distinctions between kinds of objects.[(8)]

One should seek instead a formal theory in which ordered pairs (X,Y) can be accomodated in a comprehension schema without any joint type restrictions on X,Y. From §2.3 Remark (ii), one obvious extension of the stratification scheme to accomplish this leads to a contradiction. Another candidate is proposed in Appendix III; we do not know if it is consistent. The Burali-Forti Paradox provides a good test for any such proposed extension. Recall that one proceeds there by defining ordinals as equivalence classes $\alpha = [X]_{Is}$ of well-ordered structures X under the equivalence relation

$$(X,Y) \in Is \leftrightarrow_{\text{Def}} X \cong Y .$$

Then for $\alpha = [X]_{Is}$, $\beta = [Y]_{Is}$ one puts $\beta \leq \alpha$ if Y is isomorphic to an initial segment of X. A crucial point in the argument is that each well-ordered X is isomorphic to the ordinals preceding $[X]_{IS}$; this requires mixing types in an essential way.

The theory S^* considered next incorporates a constant V_0 which we can interpret as any reasonably closed universe of sets. While S^* uses the same stratification scheme as S_1^* for meta-classes, it is essentially unrestricted on sets. In S^* we can form the operation $x \mapsto [x]_E$ for an equivalence relation E between sets, and we can form $\Pi_{X \in I} F(X)$ for $I \subseteq V_0$ and $F: I \to PV_0$. Thus while S^* still has the same defect as S_1^* for work with meta-classes in general, we can get a partial accomodation of the basic operations on structures at least for the ones that interest us in practice. At the same time we have in S^* some of the freedom expected of an unlimited theory.

§3. The Theory S^*.

Languages and axioms. The languages L_1^* and L_{TP} are now both extended by adjoining

(1) (i) a constant V_0, and

(ii) variables $a,b,c,\ldots,x,y,z$

called set-variables. The extended languages are denoted L^* and L_{TPS} respectively. The inductive definition in §2.2(2) of the set of terms of type n is extended further as follows:

(2) The constant V_0 and each set-variable is a term of type n, for all n.

The language L of set theory, using only set-variables and the symbols $=$ and is a sublanguage of L^*.

A formula ϕ of L^* is said to be (L_{TPS}-)stratified if we can assign type superscripts to the variables of ϕ in such a way that the resulting formula ϕ^* is a formula of L_{TPS}; in this assignment, each metaclass variable is to receive the same type at all occurrences in ϕ, but no such restriction is made on the assignments to V_0 and set variables. In particular, every formula of L is stratified in this sense. The following are some simple examples of L_{TPS}-stratified formulas

(3) (i) $V_n \in X \wedge X \notin V_0$

(ii) $X \in x$

(iii) $(x,X) \in x$

But the formula $(x,X) \in X$ is not stratified.

<u>Axioms of</u> S'

I. L_{TPS}-<u>Stratified Comprehension</u>

$$(\exists A)(\forall X)[X \in A \leftrightarrow \phi]$$

for each stratified ϕ of L' in which 'A' does not occur.

II. <u>Weak Extensionality</u>

$$(\exists X)(X \in A) \wedge (\forall X)[X \in A \leftrightarrow X \in B] \rightarrow A = B$$

III. <u>Pairing</u>

$$(X_1,X_2) = (Y_1,Y_2) \rightarrow X_1 = Y_1 \wedge X_2 = Y_2$$

IV. <u>Sets and metaclasses</u>

a) $(\exists X)(x = X)$

b) $(\exists x)(X=x) \leftrightarrow X \in V_0$

c) $X \in x \rightarrow X \in V_0$

V. <u>Empty Set</u>

$$(\exists ! a)(\forall x)(x \notin a)$$

For the remaining axioms, $0 =_{Def}$ [the unique a s.t. $(\forall x)(x \notin a)$]. Then by IV c), $(\forall X)(X \notin 0)$. The further symbols are as introduced by Definitions 2.1, 2.3, 2.4.

VI. Operations on sets

a) $\{x_1, x_2\} \in V_0$

b) $\bigcup x \in V_0$

c) $Px \in V_0$

d) $(x_1, x_2) = \{\{x_1\}, \{x_1, x_2\}\}$

VII. Infinite set

$$(\exists a)\ [0 \in a \wedge (\forall x)(x \in a \rightarrow x \cup \{x\} \in a)]$$

VIII. Replacement on V_0.

$$(\forall x \in a)(\forall y_1)(\forall y_2)[\psi(x,y_1) \wedge \psi(x,y_2) \rightarrow y_1 = y_2] \rightarrow (\exists b)(\forall y)[y \in b \leftrightarrow (\exists x \in a)\psi(x,y)]$$

for each formula ψ of $L^{\frown}$ ('b' not in ψ).

IX. $\in$-Induction on V_0.

$$(\forall x)[(\forall y \in x)\psi(y) \rightarrow \psi(x)] \rightarrow (\forall x)\psi(x)$$

for each $L^{\frown}$-formula ψ.

X. Universal Choice.

$$(\exists C)\{Fun(C) \wedge (\forall X)[(\exists U)(U \in X) \rightarrow (\exists U)((X,\{U\}) \in C \wedge U \in X)]\}$$

Theorem 3.1

(i) S' is consistent,

(ii) S' is an extension of ZFC

Proof: Part (i) wil be proved in the next subsections §§3.2, 3.3. Part (ii) quite simple, but a few points are worth noting. IV expresses that every set is a meta-class and that V_0, which is transitive, is exactly the range of the set variables.(9) It is not excluded that there are distinct empty meta-classes, but there is only one empty set. Hence Extensionality holds for sets. If $\phi(x)$ is any formula of set-theory L we take

$$[x|\phi(x)] =_{Def} [X|(\exists x)(X=x \wedge \phi(x))] .$$

Then by Def. 2.1 and L_{TPS}-Stratified Comprehension

$$A = [x|\phi(x)] \rightarrow (\forall x)[x \in A \leftrightarrow \phi(x)] .$$

It is easily checked that the definitions of { , } and U are in agreement with the usual ones. For the operation P we need also to observe

$$X \subseteq x \rightarrow X \in V_0 ;$$

this is a form of the Separation axiom, which follows from Replacement. As to VI d), this merely says (,) agrees with the usual pairing operation on V_0; all we really need, though, is closure of V_0 under pairing. Replacement and ϵ-Induction in ZF are special cases of the schemata VIII and IX, using only L-formulas ψ. Note that no stratification restriction is made on ψ in the full schemata. To obtain Universal Choice for sets, we can take

$$C_0 = [z|(\exists x,u)(z = (x,u) \wedge (x,\{u\}) \in C)]$$

where C exists by X. Then we have

$$\mathrm{Fun}(C_0) \wedge (\forall x)\ [x \neq 0 \rightarrow C_0(x) \in x]$$

Remark. S^* is not a conservative extension of ZFC. For we can interpret the theory MKC of Morse-Kelley (with choice) in S^*, taking as the range of the 2nd order variables all $X \subseteq V_0$ such that $(\exists y)(y \in X) \vee X = 0$. As is known, MKC is stronger than ZFC because we can define truth for L in MKC and use it to prove Cons(ZFC). MKC is essentially the 2nd order system with full comprehension extending ZFC. Even more, we can interpret the simple theory of types §2.2(7) in the obvious way in S^* with the variables of type 0 ranging over V_0. On the other hand, if we simply replace the axioms V - X of S^* by those of ZFC and keep VI d) (for ordered pair on sets), the resulting system can be shown to be a conservative extension of ZFC.

§3.2. The consistency proof: construction of the model. Part (i) of Theorem 3.1 will be established here by Jensen's methods [J]. The proof assembles the following three elements:

(i) Specker's [Sp] reduction of consistency of NF to existence of models $\mathfrak{M}_T = (\langle U_i \rangle, \langle \epsilon_i \rangle)_{i \epsilon Z}$ of extensional type theory with arbitrary integer types $i \epsilon Z$ and having a shifting automorphism

$$\sigma\colon U_i \to U_{i+1}$$

for al Namely, $\mathfrak{M}^* = (U_0, \epsilon^*)$ with

$$a \epsilon^* b \Longleftrightarrow a \epsilon_0 \sigma(b)$$

gives a model of NF. The same construction works just as well for NFU and type theory with Weak Extensionality.

(ii) The use by Ehrenfeucht and Mostowski [E-M] of Ramsey's theorem to get models of first-order theories with indiscernibles $\{c_i\}_{i \epsilon I}$ in given orderings $(I,<)$. When these models are generated by Skolem functions from the indiscernibles we get elementary substructures having automorphisms induced by those of $(I,<)$.

Jensen applied (ii) to get models $\mathfrak{M}$ of (Zermelo set theory) + (Skolem axioms) with indiscernibles $\{c_i\}_{i \epsilon Z}$ and shifting automorphism induced by $\sigma(c_i) = c_{i+1}$. A Z-typed model is formed from $U_i = \{x | x \epsilon c_i\}$, and thence a model $\mathfrak{M}^*$ of NFU by (i). The same methods were used to show the axioms of Infinity and Choice consistent with NFU.

In Part II of [J], Jensen showed how, given any ordinal α, to get such $\mathfrak{M}'$ which are end extensions of α. This employs:

(iii) The generalization of Ramsey's theorem to certain infinite partitions by Erdős-Rado [E-R].

The same methods are used here to get such $\mathfrak{M}^*$ which are end-extensions of any transitive set. This is best put as an application of a theorem about existence of models of $L_{\infty\omega}$ with indiscernibles satisfying certain prescribed properties, presented below as the theorem of Appendix I. Apparently this has not been formulated previously in the literature, though little needs to be added to Jensen's or other well-known work on indiscernibles to obtain it. Appendix I also contains basic definitions concerning indiscernibles and Skolem fragments for $L_{\infty\omega}$ and the statements of known results from which the Theorem of App. I is derived.

We now proceed with the proof of 3.1(i). Let κ be an inaccessible number, R_κ the set of all sets of rank $< \kappa$ and $R'_\kappa = R_\kappa \cup \{R_\kappa\}$. Enlarge the language L of set theory by adding constant symbols $\bar{a}$ for each $a \in R'_\kappa$; this language is denoted $L(R'_\kappa)$. Consider the following set of sentences E_κ of sentences in $L_{\infty\omega}(R'_\kappa)$:

(1) $$\forall x[x \in \bar{a} \leftrightarrow \bigvee_{b \in a} x = \bar{b}] \quad \text{for each } a \in R'_\kappa .$$

A structure satisfies E_κ just in case it is isomorphic to an end extension of R'_κ. Let L_A be the least Skolem fragment of $L_{\infty\omega}(R'_\kappa)$ which contains the sentences of E_κ. L_A contains for each formula $(\exists x)\phi(x,y_1,\ldots,y_n)$ $(n \geq 0)$ an n-ary function (or constant) symbol $f_{\exists x\phi}$. The cardinality of the set Fm_A of L_A-formulas is equal to κ:

(2) $$|Fm_A| = \kappa .$$

Now let δ be the first inaccessible cardinal greater than κ. (R_δ,ϵ) is a model of ZFC. There exist functions $F_{\exists x\phi}$ on R_δ so that

(3) $$\mathfrak{M}_0 = (R_\delta, \epsilon, \langle F_{\exists x\phi} \rangle_{\phi \in Fm_A}$$

is also a model for Sk_A, the Skolem axioms in L_A $(\phi \in Fm_A)$:

(4) $$(\exists x)\phi(x,y_1,\ldots,y_n \rightarrow \phi(f_{\exists x\phi}(y_1,\ldots,y_n),y_1,\ldots,y_n) .$$

Note that if ZFC $\vdash (\exists! x)\phi(x,y_1,\ldots,y_n)$ then $F_{\exists x\phi}$ is the set-theoretical operation on R_δ which gives x as a function of $y_1,\ldots,y_n$. In particular, the operations { , } of unordered pair, and (,) of ordered pair (defined from { , } as usual) are Skolem functions, as are the operations $\bigcup$ of union and P of power set. This is a very important point in the proof.

Consider the set

(5) $I_0 = \{R_\alpha \;\; \kappa < \alpha < \delta$ and R_α is closed under all the functions $F_{\exists x\phi}$ for $\phi\epsilon Fm_A\}$.

I_0 is linearly ordered by the ϵ-relation; we write $<$ for the restriction of ϵ to I_0.

(6) $$|I_0| = \delta.$$

For it is seen by the method of Montague and Vaught [M-V] that if $R_\alpha \epsilon I_0$ then we can find $R_\beta \epsilon I_0$ with $\alpha < \beta$; also I_0 is closed under increasing unions.

It follows by the Theorem of Appendix I that there exists a model

(7) $$\mathfrak{M}_1 = (M_1, \epsilon_{\mathfrak{M}_1}, \langle G_{\exists x\phi}\rangle_{\phi\epsilon Fm_A})$$

containing a set $\{c_i\}_{i\epsilon Z}$ of indiscernibles ordered in the type of the integers, i.e. satisfying (i) of the following, and satisfying two further conditions:

(8) i) $\mathfrak{M}_1 \models \phi(c_{i_1},\ldots,c_{i_n}) \Leftrightarrow \mathfrak{M}_1 \models \phi(c_{j_1},\ldots,c_{j_n})$ whenever $i_1 < \ldots < i_n$ and $j_1 < \ldots < j_n$,

(ii) $\mathfrak{M}_0 \equiv_{L_A} \mathfrak{M}_1$, and

(iii) $\mathfrak{M}_1 \models \phi(c_{i_1},\ldots,c_{i_n})$ for all $i_1 < \ldots < i_n$ whenever $\mathfrak{M}_0 \models \phi(R_{\alpha_1},\ldots,R_{\alpha_n})$ for all $R_{\alpha_1} < \ldots < R_{\alpha_n}$ in I_0.

Now take

(9) $\mathfrak{M} = (M, \epsilon_{\mathfrak{M}}, \langle G_{\exists x\phi}\rangle_{\phi\epsilon Fm_A})$ where

(i) M is the subset of M_1 generated from $\{c_i\}_{i\epsilon Z}$ by all $G_{\exists x\phi}(\phi\epsilon Fm_A)$, and

(ii) $\epsilon_{\mathfrak{M}}$ is $\epsilon_{\mathfrak{M}_1}$ restricted to M, as is each $G_{\exists x\phi}$.

Since $\mathfrak{M}_1 \models Sk_A$,

(10) $\mathfrak{M}$ is an L_A-elementary substructure of $\mathfrak{M}_1$, so that conditions 8(i)-(iii) hold for $\mathfrak{M}$ in place of $\mathfrak{M}_1$.

In particular,

(i) $\mathfrak{M} \models ZFC$

(ii) $\mathfrak{M} \models (\forall x)[x \in \overline{a} \leftrightarrow \bigvee\!\!\!\bigvee_{b\in a} x = \overline{b}]$ for each $a \in R'_\kappa$,

(iii) $\mathfrak{M} \models (\forall y_1,\ldots y_n)[y_1,\ldots,y_n \in c_i \to f_{\exists x\phi}(y_1,\ldots,y_n) \in c_i]$

for each i and $\phi \in Fm_A$,

(iv) $\mathfrak{M} \models (c_i \in c_j)$ whenever $i < j$.

(v) $\mathfrak{M} \models \forall x,y\ [y \in x \wedge x \in c_i \to y \in c_i]$ for each i

Parts (iii)-(v) of this follow from (10) and (8)(iii) since first, each R_α in I_o is closed under all the functions $F_{\exists x\phi}$; and second, $R_\alpha \in R_\beta$ whenever $R_\alpha < R_\beta$ in I_o, and finally each R_α is transitive.

By (11)(ii), $\mathfrak{M}$ is isomorphic to an end-extension of R'_κ. We may thus assume that it is actually an end-extension, i.e.

(i) $R_\kappa \cup \{R_\kappa\} \subseteq M$

(ii) for each $a \in R'_\kappa$, $(\overline{a})_{\mathfrak{M}} = a$, and

(iii) for each $a \in R'_\kappa$ and $b \in M$, $b \in_{\mathfrak{M}} a \Longleftrightarrow b \in a$.

Note that each $a \in R'_\kappa$ is the 0-ary Skolem function corresponding to the sentence $(\exists x)(x=\overline{a})$. It follows by (11)(iii) that

(13) $a \in_{\mathfrak{M}} c_i$ for each $i \in Z$ and $a \in R'_\kappa$

We can establish:

(14) $\mathfrak{M}$ has an automorphism $\sigma: M \to M$ satisfying $\sigma(c_i) = c_{i+1}$ for all $i \in Z$.

To be such an automorphism, σ must also satisfy

(15) $\sigma(G(b_1,\ldots,b_n)) = G(\sigma(b_1),\ldots,\sigma(b_n))$ for each $G = G_{\exists x\phi}$ of $\mathfrak{M}$.

σ is defined as follows. Each element b of M has some representation in the

$$b = G(c_{i_1},\ldots,c_{i_n}$$

for some $G = G_{\exists x\phi}$ of $\mathfrak{M}$, since a composition of Skolem functions is again a Skolem function We take

$$\sigma(b) = G(c_{i_1+1},\ldots,c_{i_n+1}) .$$

To see that σ is well-defined, suppose given another representation

$$b = G'(c_{j_1},\ldots,c_{j_m})$$

$$G(c_{\ +1},\ldots,c_{\ n+1} = G'(c_{j_1+1},\ldots,c_{j_m+1})$$

follows from

$$G(c_{i_1},\ldots,c_{i_n}) = G'(c_{j_i},\ldots,c_{j_m}$$

by indiscernibility. σ is seen to be one-one and onto and, finally, an automorphism, by the same facts.

We next use $\mathfrak{M}$ and the $\{c_i\}_{i\epsilon Z}$ to form a Z-typed structure:

$\mathfrak{M}_T = (\langle U_i \rangle_{i\epsilon Z}, \langle \epsilon_i \rangle_{i\epsilon Z}, \langle G_{\exists x\phi} \rangle_{\phi\epsilon Fm_A}$ <u>where</u>

(i) $U_i = \{a | \ a \in M$ <u>and</u> $a \in_{\mathfrak{M}} c_i\}$ <u>and</u>

(ii) $a \in_i b \Longleftrightarrow a \in U_i,\ b \in U_{i+1},\ a \in_{\mathfrak{M}} b$ <u>and</u> $b \subseteq_{\mathfrak{M}} c_i$.

Here $b \subseteq_{\mathfrak{M}} c_i$ abbreviates $(\forall x)[x \in_{\mathfrak{M}} b \Rightarrow x \in_{\mathfrak{M}} c_i]$; this is added to insure Weak Extensionality. For if we have $b_1, b_2 \in U_{i+1}$ and know simply that $\forall a \in U_i\ [a \in_{\mathfrak{M}} b_1 \Longleftrightarrow a \in_{\mathfrak{M}} b_2]$ it does not follow in general that $b_1 = b_2$. But $\forall a \in U_i\ [a \in_i b_1 \Longleftrightarrow a \in_i b_2]$ does imply $b_1 = b_2$ when $(\exists a \in U_i)(a \in_i b_1)$. Note that we may consider each $G = G_{\exists x\phi}$ as acting on each U_i separately since by (11)(iii)

$$G \upharpoonright U_i^n : U_i^n \to U_i$$

Finally, we form a Specker model $\mathfrak{M}$ from $\mathfrak{M}_T$:

$\mathfrak{M}^* = (U_0, \epsilon^*, \langle G_{\exists x\phi} \rangle_{\phi \in Fm_A})$ <u>where</u> $a \epsilon^* b \iff a \epsilon_0 \sigma(b)$

We give the following interpretation of L^* in U_0:

(i) <u>the variables</u> A,B,C,...,X,Y,Z <u>range over</u> U_0,

(ii) <u>the set-variables</u> a,b,c,...,x,y,z <u>range over</u> R_κ,

(iii) V_0 <u>is interpreted</u> as R_κ,

(iv) = <u>is interpreted as</u> =, and ϵ as ϵ^*.

(v) <u>the pairing symbol is interpreted by the function</u> $G_{\exists x\phi}$ <u>where in</u> $\mathfrak{M}_0$, $F_{\exists x\phi}(a,b) = (a,b) = \{\{a\},\{a,b\}\}$.

We shall show that $\mathfrak{M}^*$ <u>is a model of</u> S^* under this interpretation. Note that we can also consider more generally statements in L_A^*, interpreting $\bar{a}$ as a for each $a \in R_\kappa'$ and $f_{\exists x\phi}$ as $G_{\exists x\phi}$ for each $\phi \in Fm_A$.

3.3. <u>The consistency proof (cont.): verification of the axioms.</u>

To spell out how the properties of $\mathfrak{M}^*$ reduce to those of $\mathfrak{M}$ via $\mathfrak{M}_T$ we consider a <u>language</u> $L_{\mathfrak{M}_T}$ <u>appropriate to</u> $\mathfrak{M}_T$, <u>thus extending the language</u> L_{TPS}. Its symbols and their interpretations are specified as follows.

(1) (i) <u>variables</u> $A^i, B^i, C^i, \ldots, X^i, Y^i, Z^i$ <u>of each type</u> $i \in Z$, <u>interpreted as ranging over</u> U_i,

(ii) <u>set-variables</u> a,b,c,...,x,y,z, <u>the constant</u> V_0, <u>and function symbols</u> $f_{\exists x\phi}$, <u>for</u> ϕ <u>in</u> Fm_A, <u>each interpreted as in</u> L_A^*.

The sets of terms of type i in $L_{\mathfrak{M}_T}$ are generated inductively as follows:

(2) (i) <u>each variable of type</u> i <u>is a term of type</u> i;

(ii) <u>each set-variable and the constant</u> V_0 <u>are terms of type</u> i <u>for all</u> i;

(iii) <u>if</u> $f_{\exists x\phi}$ <u>is n-ary</u> $(\phi \in Fm_A)$ <u>and</u> $t_1,\ldots,t_n$ <u>are of type</u> i <u>then</u> $f_{\exists x\phi}(t_1,\ldots,t_n)$ <u>is of type</u> i.

(2)(ii) is coherent since $R'_\kappa \subseteq U_i$ for all i by 3.2(13),(16). Under any assignment to variables of type j in U_j, each term t of type i has a value $(t)_{\mathfrak{M}_T}$ in U_i. The <u>atomic formulas</u> of $L_{\mathfrak{M}_T}$ are described just as for L_{TP}, but now using the preceding sets of terms; they are just those of the form

(3) (i) $(t_1 = t_2)$ <u>for type</u> (t_1) = type (t_2) <u>and</u>

(ii) $(t_1 \epsilon t_2)$ <u>for</u> type $(t_1) + 1$ = type (t_2); <u>where</u> t_1, t_2 <u>are terms of</u> $L_{\mathfrak{M}_T}$.

Given any appropriate assignment to the variables, the formula $t_1 = t_2$ is to be satisfied in $\mathfrak{M}_T$ just in case $(t_1)_{\mathfrak{M}_T} = (t_2)_{\mathfrak{M}_T}$ and the formula $(t_1 \epsilon t_2)$ is to be satisfied just in case $(t_1)_{\mathfrak{M}_T} \epsilon_i (t_2)_{\mathfrak{M}_T}$ for i = type (t_1). Now there is a possible ambiguity, due to (2)(ii) and the fact that we are only using one symbol ϵ in $L_{\mathfrak{M}_T}$. Each term t of type i is built up from V_0, set variables and variables of type i by the function symbols. We may write it in the form

(4) $$t(x_1,\ldots,x_n,\ Y_1^i,\ldots,Y_m^i)\ .$$

It may be that $n = 0$ or $m = 0$. If $m > 0$ then i is uniquely determined by t. Otherwise we have a term of the form $t(x_1,\ldots,x_n)$; all such are called the <u>terms of indefinite type</u>. It must be checked for atomic formulas $(t_1 \epsilon t_2)$ in which either t_1 or t_2 is of indefinite type that the interpretation (3) is well-determined. This is true because $\mathfrak{M}$ and each U_i is an end extension of R'_κ. In particular if $(t_1)_{\mathfrak{M}} = a$, $(t_2)_{\mathfrak{M}} = b$ under an assignment and t_2 is of indefinite type, then $b \subseteq_{\mathfrak{M}} c_i$ for each i; thus for $a \in U_i$ we have $a \epsilon_i b \Longleftrightarrow a \epsilon_{\mathfrak{M}} b \Longleftrightarrow a \epsilon b$ (and $a \epsilon R'_\kappa$).

With each formula $\psi(x_1,\ldots,x_n,\ Y_1^{i_1},\ldots,Y_m^{i_m})$ of $L_{\mathfrak{M}_T}$ is associated a formula $\psi^-(x_1,\ldots,x_n,\ y_1,\ldots,y_m)$ which may be considered the <u>translation of</u> ψ <u>into the language of</u> $(\mathfrak{M}, \langle c_i \rangle_{i \in Z})$.

(5) (i) <u>For</u> t <u>of the form</u> (4), t^- <u>is</u> $t(x_1,\ldots,x_n,\ y_1,\ldots,y_m)$;

(ii) $(t_1 = t_2)$ <u>is</u> $(t_1^- = t_2^-)$;

(iii) $(t_1 \epsilon t_2)^-$ <u>is</u> $[t_1^- \epsilon t_2^- \wedge (\forall x)(x \epsilon t_2^- \rightarrow x \epsilon c_i)]$ <u>for</u> t_2 <u>of type</u> $i+1$;

(iv) $(\ )^-$ <u>preserves</u> $\sim$ <u>and</u> $\vee$;

$((\exists x)\psi)^-$ <u>is</u> $(\exists x \in \overline{R}_\kappa)\psi^-$; <u>and</u>

(vi) $((\exists Y^i)\psi^-)$ <u>is</u> $(\exists y \in c_i)\psi^-$.

In (iii) the choice of i is arbitrary for t_2 of indefinite type. The following is then easily proved by induction.

<u>Lemma</u> 1. <u>If</u> $\psi(x_1,\dots,x_n, Y_1^{i_1},\dots,Y_m^{i_m})$ <u>is a formula of</u> $L_{\mathfrak{M}_T}$ <u>and</u> $a_1,\dots,a_n \in R_\kappa$, $b_1 \in U_{i_1},\dots,b_m \in U_{i_m}$ <u>then</u>

$$\mathfrak{M}_T \models \psi(a_1,\dots,a_n, b_1,\dots,b_m) \iff (\mathfrak{M}, c_i\ _{i\in Z}) \models \psi^-(a_1,\dots,a_n, b_1,\dots,b_m).$$

We also have associated with each such ψ a formula ψ^* of L_A^* <u>got by suppressing types</u>, i.e. each variable Y^i of ψ is replaced by a distinct variable y.

<u>Lemma</u> 2. <u>If</u> $\psi(x_1,\dots,x_n, Y_1^{i_1},\dots,Y_m^{i_m})$ <u>is a formula of</u> $L_{\mathfrak{M}_T}$ <u>and</u> $a_1,\dots,a_n \in R_\kappa$, $b_1,\dots,b_m \in U_0$ <u>then</u>

$$\mathfrak{M}^* \models \psi^*(a_1,\dots,a_n,b_1,\dots,b_m) \iff \mathfrak{M}_T \models \psi(a_1,\dots,a_n,\sigma^{i_1}(b_1),\dots,\sigma^{i_m}(b_m))$$

Proof. By induction on ψ. There are a few points to consider here. We use repeatedly the fact that σ preserves all Skolem functions $G_{\exists x\phi}$ and in particular that $\sigma(a) = a$ for each $a \in R_\kappa'$ (o-ary Skolem functions).

(i) If ψ has the form $t_1 = t_2$ we can write it in the form $t_1(x_1,\dots,x_n,Y_1^i,\dots,Y_m^i) = t_2(x_1,\dots,x_n,Y_1^i,\dots,Y_m^i)$ (not all variables need actually occur on each side); i is uniquely determined except in case both t_1, t_2 are indefinite terms, in which case the statement is trivial. Then by applying σ^i to both sides we get

$$,a_n,b_1,\dots,b_m) = t_2(a_1,\dots,a_n,b_1,\dots,b_m)$$

$$\iff t_1(a_1,\dots,a_n,\sigma^i(b_1),\dots,\sigma^i(b_m)) = t_2(a_1,\dots,a_n, \sigma^i(b_1),\dots,\sigma^i(b_m).)$$

(each term evaluated in the appropriate structure)

(ii) If ψ has the form $t_1 \in t_2$ we can write it in the form

$$t_1(x_1,\dots,x_n,Y_1^i,\dots,Y_p^i) \in t_2(x_1,\dots,x_n,Y_{p+1}^{i+1},\dots,Y_m^{i+1})$$

for some $0 \leq p \leq m$, where again i is uniquely determined except when both t_1,t_2 are of indefinite type; in that case the statement is again trivial. Let $d_1 = t_1(a_1,\dots,a_n,b_1,\dots,b_p)$ and $d_2 = t_2(a_1,\dots,a_n,b_{p+1},\dots,b_m)$. We have

$$d_1 \in^* d_2 \Leftrightarrow d_1 \in_0 \sigma(d_2) \qquad \text{(by defn.)}$$
$$\Leftrightarrow \sigma^i(d_1) \in_i \sigma^{i+1}(d_2) \ .$$

The result follows from $\sigma^i(d_1) = t_1(a_1,\dots,a_n,\sigma^i(b_1),\dots,\sigma^i(b_p))$, $\sigma^{i+1}(d_2) = t_2(a_1,\dots,a_n,\sigma^{i+1}(b_{p+1}),\dots,\sigma^{i+1}(b_m))$.

(iii) The induction step is trivial for $\sim$, $\vee$, and existential quantification with a set variable $\exists z$. We suppose then that it holds for θ and prove it for ψ when

$$\psi(x_1,\dots,x_n,Y_1^{i_1},\dots,Y_m^{i_m}) \text{ is } (\exists Z^j)\theta(x_1,\dots,x_n,Y_1^{i_1},\dots,Y_m^{i_m},Z^j)$$

$$\mathfrak{M}^* \models \psi(a_1,\dots,a_n,b_1,\dots,b_m \Leftrightarrow (\exists c\in U_0)\ \mathfrak{M}^* \models \theta(a_1,\dots,a_n,b_1,\dots,b_m,c)$$
$$\Leftrightarrow (\exists d\in U_j)\ \mathfrak{M}^* \models \theta(a_1,\dots,a_n,b_1,\dots,b_m,\sigma^{-j}(d))$$
$$\Leftrightarrow (\exists d\in U_j)\ \mathfrak{M}_T \models \theta(a_1,\dots,a_n,\sigma^{i_1}(b_1)\dots,\sigma^{i_n}(b_m),d)$$
$$\Leftrightarrow \mathfrak{M}_T \models \psi(a_1,\dots,a_n,\sigma^{i_1}(b_1),\dots,\sigma^{i_n}(b_m)) \ .$$

This completes the proof of Lemma 2.

While by the definitions of §3.2, the properties ϕ of $\mathfrak{M}^*$ are completely determined by those of $\mathfrak{M}$, we get a usable reduction only in certain cases; in particular this is provided by Lemmas 1 and 2 when ϕ is stratified, so $\phi = \psi^*$ for some ψ in $L_{\mathfrak{M}_T}$.(10) We now proceed with the verification of the axioms in $\mathfrak{M}^*$.

I. L_{TPS}-<u>Stratified Comprehension is true in</u> $\mathfrak{M}^*$. Consider an instance $(\exists A)(\forall X)[X \in A \leftrightarrow \phi]$ of this scheme. Let the parameters of ϕ be $x_1,\dots,x_n$, $Y_1,\dots,Y_m$; we write ϕ as $\phi(x_1,\dots,x_n, Y_1,\dots,Y_m,X)$. Since ϕ is

L_{TPS}-stratified we can find ψ in $L_{\mathfrak{M}_T}$ such that $\phi = \psi^*$; without loss of generality we can assign X the type 0 in this stratification. Write ψ as $\psi(x_1,\dots,x_n, Y_1^{i_1},\dots,Y_m^{i_m},X^0)$. By Lemmas 1 and 2, for any $a_1,\dots,a_n \in R_\kappa$, $b_1,\dots,b_m \in U_0$ and $x \in U_0$

$$\mathfrak{M}^* \models \phi(a_1,\dots,a_n, b_1,\dots,b_m,x) \Leftrightarrow \mathfrak{M}_T \models \psi(a_1,\dots,a_n,\sigma^{i_1}(b_1),\dots,\sigma^{i_m}(b_m),x)$$

$$\Leftrightarrow (\mathfrak{M},\langle c_i\rangle_{i\in Z} \models \psi^-(a_1,\dots,a_n,\sigma^{i_1}(b_1),\dots,\sigma^{i_m}(b_m),x)$$

$$\Leftrightarrow \mathfrak{M} \models \theta(x)$$

$\theta(x)$ is a formula of L together with a finite number of constant symbols from R'_κ and from $\{c_i\}_{i\in Z}$. By Separation in $\mathfrak{M}$ we can find $b^+ \in M$ with

$$x \in_{\mathfrak{M}} b^+ \Leftrightarrow x \in_{\mathfrak{M}} c_0 \text{ and } \mathfrak{M} \models \theta(x) .$$

Since $b^+ \subseteq_{\mathfrak{M}} c_0$ we have $b^+ \in_{\mathfrak{M}} P(c_0$ But $P(c_0) \in_{\mathfrak{M}} c_1$ and c_1 is transitive in $\mathfrak{M}$ from §3.2(11). Hence $b^+ \in_{\mathfrak{M}} c$ and $\sigma^{-1}(b^+) \in_{\mathfrak{M}} c_0$. Let $b = \sigma^{-1}(b^+)$; we have for al $x \in_{\mathfrak{M}} c_0$:

$$x \in^* b \Leftrightarrow x \in_{\mathfrak{M}} \sigma(b) \text{ and } \sigma(b) \subseteq_{\mathfrak{M}} c_0$$

$$\Leftrightarrow x \in_{\mathfrak{M}} b^+$$

$$\mathfrak{M}^* \models \phi(a_1,\dots,a_n, b_1,\dots,b_n,x)$$

$(\forall x_1,\dots,x_n, Y_1,\dots,Y_m)(\exists B)(\forall X)[X \in B \leftrightarrow \phi(x_1,\dots,x_n, Y_1,\dots,Y_m,X)]$ is true in $\mathfrak{M}^*$.

II. <u>Weak Extensionality is true in</u> $\mathfrak{M}^*$. For, if $a,b \in U_0$ and $(\exists x \in U_0)(x \in^* a)$ and $(\forall x \in U_0)[x \in^* a \Leftrightarrow x \in^* b]$ then $(\exists x \in U_0)(x \in_{\mathfrak{M}} \sigma(a) \;\&\; \sigma(a) \subseteq_{\mathfrak{M}} c_0)$ and $(\forall x \in U_0)[x \in_{\mathfrak{M}} \sigma(a) \;\&\; \sigma(a) \subseteq c_0 \Leftrightarrow x \in_{\mathfrak{M}} \sigma(b) \;\&\; \sigma(b) \subseteq_{\mathfrak{M}} c_0]$. Hence both $\sigma(a), \sigma(b) \subseteq_{\mathfrak{M}} c_0$ and $(\forall x \in_{\mathfrak{M}} c_0)[x \in_{\mathfrak{M}} \sigma(a) \Leftrightarrow x \in_{\mathfrak{M}} \sigma(b)]$, so $\sigma(a) = \sigma(b)$ by Extensionality in $\mathfrak{M}$, and then $a = b$.

III. The Pairing Axiom is true in $\mathfrak{M}^*$. This is a special case of the following:

(6) If ϕ is any sentence of L_A^* which does not contain the ϵ symbol and $\mathfrak{M}_0 \models \phi$ then $\mathfrak{M}^* \models \phi$.

To begin with $\mathfrak{M}_0$ (of §3.2(3)) satisfies the relativization $\phi^{(R_\alpha)}$ for each $R_\alpha \in I_0$, since R_α gives an elementary substructure of $\mathfrak{M}_0$. By §3.2(8)(iii) and (10), $\mathfrak{M} \models \phi^{(c_0)}$. If we ignore ϵ, $\mathfrak{M}^*$ is the same as $\mathfrak{M}$ restricted to the $\mathfrak{M}$ members of c_0 - so $\mathfrak{M}' \models \phi$.

Remark. (6) is the essential point which permits extension of Jensen's proof to systems with additional axioms which may contain Skolem function symbols, but are free of 'ϵ'. By Schema I, we cannot extend (6) to all ϕ in L_A^*; for example, $\sim(\exists A)(\forall X)(X \epsilon A)$ is true in $\mathfrak{M}$ but false in $\mathfrak{M}^*$. It is instructive to see what goes wrong with another example, using the symbol for the singleton function. $\mathfrak{M}$ satisfies $(\forall X,Y)[Y \in \{X\} \leftrightarrow Y = X]$, but this is false in $\mathfrak{M}^*$ since for $a \in_{\mathfrak{M}} c_0$ we have $a \in^* \{c_{-1}\} \Leftrightarrow a \in_{\mathfrak{M}} \sigma(\{c_{-1}\})$ & $\sigma(c_{-1}) \subseteq_{\mathfrak{M}} c_0$. But $\sigma(\{c_{-1}\}) = \{\sigma(c_{-1})\}$ $= \{c_0\} \not\subseteq_{\mathfrak{M}} c_0$, so $\{c_{-1}\}$ has no ϵ^*-members. This does not contradict $\mathfrak{M}^* \models (\forall X)(\exists A)(\forall Y)[Y \in A \leftrightarrow Y = X]$ by Schema I; we return to this in connection with Axiom X below.

Axioms IV, V, VI and VII of S^* are true in $\mathfrak{M}^*$. These are the axioms relating sets and meta classes and stating closure of V_0 under the usual operations. They are easily checked because V_0 is interpreted as R_κ and σ is the identity on R'_κ, so ϵ^* agrees with $\epsilon_{\mathfrak{M}}$ on this set. By this we may further strengthen our consistency result as follows:

(7) $\mathfrak{M}^* \models \phi^{(V_0)}$ for any sentence ϕ of L which is true in R_κ.

VIII. Replacement on V_0 is true in $\mathfrak{M}^*$. For suppose $\psi(x,y)$ is any formula of L^* with additional parameters fixed in U_0. Let $a \in R_\kappa$, and suppose $\mathfrak{M}^* \models (\forall x \epsilon a)(\forall y_1,y_2)[\psi(x,y_1) \wedge \psi(x,y_2) \rightarrow y_1 = y_2]$. Define $a_1 = \{x | x \in a$ and $\mathfrak{M}^* \models \exists y\, \psi(x,y)\}$, $F: a_1 \rightarrow R_\kappa$, $F(x) =$ the unique y s.t. $\mathfrak{M}^* \models \psi(x,y)$. Since κ is inaccessible, $b = D\breve{}F$ belongs to R_κ and $\mathfrak{M}^* \models (\forall y)[y \in b \leftrightarrow (\exists x \epsilon a)\psi(x,y)]$.

This depends only on the interpretation of the set variables as ranging over R_κ and that $\mathfrak{M}^*$ is an end-extension of (R_κ, ϵ). For the same reason we have:

IX. ϵ-<u>Induction on</u> V_0 <u>is true in</u> $\mathfrak{M}^\sim$.

X. <u>Universal Choice is true in</u> $\mathfrak{M}^\sim$. If we write this axiom out without the abbreviations of §2.4, it takes the form

(8) $(\exists B)((\forall X,Y_1,Y_2)[(X,Y_1) \in B \wedge (X,Y_2) \in B \rightarrow Y_1 = Y_2]$

$$\wedge (\forall X)\{(\exists U)(U \epsilon X) \rightarrow (\exists U,Y)[(X,Y) \in B \wedge (\forall Z)(Z \epsilon Y \leftrightarrow Z = U)\}$$

This is a stratified sentence; we may write it as ψ^* for ψ in $L_{\mathfrak{M}_T}$, assigning type 2 to the variable B, type 1 to each of X,Y_1,Y_2,Y, and type 0 to U and Z. By Lemma 2, (8) is true in $\mathfrak{M}^*$ just in case $\mathfrak{M}_T \models \psi$, equivalently by Lemma 1, if

$$(\mathfrak{M}, \langle c_i \rangle_{i \epsilon Z}) \models \psi'$$

In this case ψ' is directly equivalent to:

(9) $(\exists b \epsilon c_2)((\forall x,y_1,y_2 \epsilon c_1)[(x,y_1) \in b \wedge (x,y_2) \in b \wedge b \subseteq c_1 \rightarrow y_1 = y_2]$

$\wedge (\forall x \in c_1)\{(\exists u \epsilon c_0)(u \epsilon x \wedge x \subseteq c_0$

$\rightarrow (\exists u \epsilon c_0)(\exists y \epsilon c_1)[(x,y) \in b \wedge b \subseteq c_1 \wedge (\forall z \epsilon c_0)(z \in y \wedge y \subseteq c_0 \leftrightarrow z = y]\})$.

In $\mathfrak{M}$ we have the existence of

$$r = \{(x,y) | y \subseteq x \subseteq c_0 \wedge (\exists ! z)(z \epsilon y)\} .$$

$Dr = \{x | x \subseteq c_0 \wedge x \neq 0\}$. Since $Pc_i \subseteq c_{i+1}$ and each c_i is closed under pairing, by 3.2(11), we have $r \subseteq c_1$ hence $r \in c_2$. By the Axiom of Choice in $\mathfrak{M}$, there exists $b \subseteq r$ with $Fun(b) \wedge Db = Dr$. Hence also $b \subseteq c_1$ and finally $b \in c_2$ (all these statements in the sense of $\mathfrak{M}$). It is direct to show that b satisfies (9) in $\mathfrak{M}$.

This completes the proof that $\mathfrak{M}^*$ is a model of S' and hence of Theorem 3.1.

Remarks. (i) The existence of infinitely many inaccessibles is accepted with no hesitation (these days) by those who accept set-theoretical foundations as expressed by ZFC at all. In this sense we have given a consistency proof of S^* by currently accepted means. The role of the two inaccessibles κ,δ used in the proof is as follows: κ is used to insure that V_0 satisfies replacement with respect to arbitrary formulas. δ is used to get an end extension of V_0 with indiscernibles using the Erdős-Rado Theorem. (The consistency of the weaker system indicated at the end of §3.1 can be established without assuming any inaccessibles; Ramsey's Theorem suffices for the combinatorial part.) For further discussion relevant to this point cf. also the end of the next subsection.

(ii) The language $L_{\mathfrak{M}_T}$ used as a tool for the consistency proof may actually be thought of as providing a much more liberal stratification set-up, with ϕ stratified if it is of the form ψ^* for some finite ψ in $L_{\mathfrak{M}_T}$ without any constant symbols from $\mathfrak{M}$. For example, not only sets and V_0 but also $V_1 = PV_0$, $V_2 = PV_1$, etc. can be assigned any type when stratifying by this set-up. We could further permit the use of variables ranging over V_1, V_2, etc. which would again be assigned any type. The problem would be to give this a simple organization.

§3.4. Category theory in S^*. The main purpose here is to examine some of the advantages provided by S^* for the freer formulation of the unlimited theory of structures. These are illustrated in connection with one rather general theorem from category theory, known as Yoneda's Lemma. At the same time we shall see that the defect (discussed in §2.4) of the present stratification set-up for pairing blocks us from establishing a completely unrestricted formulation of this theorem.

We must first indicate in more detail how the basic notions and examples of category theory[11] are to be represented in S^*. For this purpose it is convenient to use Greek letters $\alpha,\beta,\gamma,\ldots,\xi,\eta,\zeta$ as additional variables ranging over meta-classes; these will only be used, though, for morphisms in a category.

A category is here taken to be a structure of the form

(1) $A = (O,M,C,D_0,D_1)$

where O consists of the <u>objects</u> of A, M of its <u>morphisms</u>, C is the <u>composition relation</u> and D_0, D_1 are the <u>domain</u> and <u>codomain</u> operations, resp. Thus $C \subseteq M^3$ and $(\alpha,\beta,\gamma) \in C$ if $\alpha \circ \beta = \gamma$ in A. Each $D_i\colon M \to O$. When $X = D_0(\alpha)$, $Y = D_1(\alpha)$ then $\alpha\colon X \to Y$ in the sense of A.(12) We write

(2) $\quad \mathrm{Hom}\,(X,Y) = [\alpha | \alpha \in M \wedge D_0(\alpha) = X \wedge D_1(\alpha) = Y]$

for the meta-class of all A-morphisms $\alpha\colon X \to Y$. When indicating dependence of these various collections and notions on A, we use subscripts 'A', e.g. O_A, M_A, Hom_A, etc.

The conditions to be a category are first-order so, following §§2.1, 2.3, we have the existence in S^* of a meta-class Cat such that

(3) $\quad A \in \mathrm{Cat} \Longleftrightarrow A$ <u>has the structure of a category</u>.

Suppose ϕ is any finite-order property of structures of a certain signature for which we have associated a finite-order definition of <u>homomorphic map</u> between structures satisfying ϕ. We then form

(4) $\quad$ <u>the category</u> ϕ_V <u>of all</u> X <u>such that</u> $\bar{\phi}(X)$

This is the structure A of the form (1) whose objects are given by $O = [X | \bar{\phi}(X)]$ and whose morphisms are triples $\alpha = (F,X,Y)$ where $X, Y \in O$ and F is a homomorphic map of X into Y. The members of C are the triples (α,β,γ) of the form $\alpha = (F,X,Y)$ $\beta = (G,Y,Z)$ and $\gamma = (F \circ G,X,Z)$ where $F \circ G$ is the usual composition of maps. Just as for O, each of M,C,D_0,D_1 is defined by an L_{TP}-stratified condition. Then the structure $A = \phi_V$ so formed can be shown in S^* to satisfy all the conditions to be category, i.e.

(5) $\quad S' \vdash (\phi_V \in \mathrm{Cat})$

In particular we may speak in S' of

(6) $\quad$ <u>the category</u> Class_V <u>of all meta-classes</u>,

i.e. the category of all X in V without any additional structure

$(\overline{\Phi}(X) \leftrightarrow (\forall U \epsilon X)(U=U))$. Similarly we may deal with Grp_V = the category of all groups, Top_V = the category of all topological spaces, etc. Finally, we write Cat_V for the category of all categories. This takes the form

(7) $\quad Cat_V = (Cat, Funct, \circ, D_0, D_1)$

where Funct consists of all functors between categories. The statement $(Cat_V \epsilon Cat)$ is established in S^* as a special case of (5); this represents the informal statement §1(3).

Given categories $A = (O_A, M_A, \ldots)$, $B = (O_B, M_B, \ldots)$, each functor $F: A \rightarrow B$ in Cat_V is determined by a pair of maps $F_{Ob}: O_A \rightarrow O_B$ and $F_{Mor}: M_A \rightarrow M_B$. As usual we write $F(X)$ for $F_{Ob}(X)$ and $F(\alpha)$ for $F_{Mor}(\alpha)$. Finally we put

(8) $\quad Funct(A,B) = [F|F$ is a functor from A to $B]$

$\qquad\qquad = Hom_{Cat_V}(A,B)$.

The natural transformations $\eta: F \rightarrow G$ between $F,G \epsilon Funct(A,B)$ may be considered to be maps $\eta: O_A \rightarrow M_B$ satisfying $\eta(X): F(X) \rightarrow G(X)$ (in B) for each $X \epsilon O_A$, and the usual commutativity conditions. We can then form

(9) $Nat(A,B) = [\eta|\ \eta: F \rightarrow G$ for some $F,G \epsilon Funct(A,B)]$.

With the usual composition, domain and codomain of natural transformations we obtain the structure

(10) $\quad B^A = (Funct(A,B), Nat(A,B), \circ, D_0, D_1)$

which can be shown in S^{*} to belong to Cat. That is now the formulation of the statement §1(4) in S^*. As already noted, A, B are considered here to be variable categories, and this is the only one of the statements in §1(1)-(4) which makes essential use of parameters in the comprehension schema for S^*. We have also used parameters implicitly at various points, e.g. in the definition (2) above.

Now let U be any meta-class and $A = (O,M,\ldots)$ any category. We denote by $A \upharpoonright U$ the substructure of A obtained by restricting both O, M to U,

$A \upharpoonright U = (O \cap U, M \cap U, \ldots)$.

When A is Φ_V we write Φ_U for $A \upharpoonright U$. Instead of $(\mathrm{Class}_V)_U$ we write Class_U, similarly Grp_U, Top_U, etc. A is said to be a U-<u>category</u> if $A \upharpoonright U = A$.

We wish particularly to consider the cases $U = V_0$ = <u>the (meta)-class of all sets</u> and $U = V_1 = PV_0$ = <u>the meta-class of all (meta) classes of sets.</u>(13) Note that this is the first place where we begin to make use of the special class V_0 in S^*; all of the development sketched up to this point can actually already be carried out in S_1^*. We write

(i) $\mathrm{Set} = \mathrm{Class}_{V_0}$, (ii) $\mathrm{Class} = \mathrm{Class}_{V_1}$

for the <u>category of all sets</u> and the <u>category of all classes of sets</u>, resp.

If $A = (O_A, M_A, \ldots)$ is a V_0-category then

(13) $a,b \in O_A \Rightarrow \mathrm{Hom}_A(a,b) \in V_1$

This permits us to establish a functor

(14) $H^a: A \to \mathrm{Class}$ <u>for each</u> $a \in O_A$,

given by (i) $H^a(b) = \mathrm{Hom}_A(a,b)$ <u>for each</u> $b \in O_A$ and

(ii) $H^a(\beta): \mathrm{Hom}_A(a,b) \to \mathrm{Hom}_A(a,c)$ <u>for each</u> $\beta: b \to c$ <u>in</u> A, <u>where</u>

$(H^a(\beta))(\alpha) = \beta \circ \alpha$ <u>for each</u> $\alpha: a \to b$ <u>in</u>

Written out more explicitly, (ii) takes the form

(ii)' $H^a(\beta) = [(\alpha, \beta \circ \alpha) \mid \alpha: a \to D_0(\beta)]$

or

(ii)" $H^a = [(\beta, H^a(\beta)) \mid \beta \in M_A]$

$= [(\beta, Z) \mid \beta \in M_A \wedge Z = [(\alpha, \beta \circ \alpha) \mid \alpha: a \to b]]$

Since sets can be paired with classes without restriction in L_{TPS}, the defining condition for H^a is L_{TPS}-stratified. But we are blocked from defining an analogous

$$H^X: A \rightarrow Class_V$$

for arbitrary A because $H^X(\beta)$ is essentially of one type higher than β.

For any functor category B^A and $F,G: A \rightarrow B$, let

(15) $\quad Nat(F,G) = [\eta |\ \eta$ is a natural transformation of F into G]

$$= Hom_{B^A}(F,G)$$

The following is a form of Yoneda's Lemma which can be established in A^*.

<u>YL</u>*. <u>Suppose</u> A <u>is a</u> V_0<u>-category and</u> F <u>is a functor from</u> A <u>to</u> Class. <u>Then for each</u> $a \in O_A$ <u>there is a bijection</u>

$$\nu: F(a) \rightarrow Nat(H^a,F)$$

The usual proof defines $\nu(x)$ as η_x for $x \in F(a)$, where η_x is determined by

(i) $\eta_x(b): H^a(b) \rightarrow F(b)$ for $b \in O_A$, and

(ii) $(\eta_x(b))(\alpha) = (F(\alpha))(x)$ for each $\alpha \in H^a(b)$.

Again writing (ii) out more explicitly gives us

(ii)' $\eta_x(b) = [(\alpha,y) |\ \alpha \in H^a(b) \wedge y = (F(\alpha))(x)]$

$$= [(\alpha,y) |\ \alpha \in Hom(a,b) \wedge (\alpha,(x,y)) \in F] ,$$

and

(ii)" $\eta_x = [(b,Z) |\ b \in O_A$ and $Z = \eta_x(b)]$.

Once more the defining conditions is L_{TPS}-stratified. Finally, $\nu = [(x,\eta_x) |\ x \in F(a)]$ is defined by one further abstraction involving pairs of sets and classes. Thus the existence of ν with its expected properties can be proved in S^*.

Note that once more we are blocked from making a more general statement (for A an arbitrary category) because η_x makes essential use of (intuitively) mixed types

On the other hand, a typical restricted formulation YL of Yoneda's Lemma (such as given in [M2] p. 61, but re-expressed in present terms, reading "member of V_0" for "small") requires that A be a V_0-category for which $Hom_A(a,b) \in V_0$ for all $a,b \in O_A$. YL^* is prima-facie stronger than such YL.

It is also usual to express YL in the form of a natural equivalence between two functors E, N: $A \times (Set^A) \to Set$, with $E(a,F) = F(a)$ for $F \in Funct(A,Set)$. We meet a problem when attempting a corresponding formulation in S^* for functors

$$E,N: A \times (Class^A) \to Class$$

because of the mixed type levels in the definition of E. This problem would not arise if one used a more liberal kind of stratification scheme of the kind suggested in Remark ii) of the preceding subsection.

Remark. To avoid possible confusion about aims, something should be said comparing the kind of foundation for category theory sought here and the set-theoretical foundations pursued in [F]. There the aim was to eliminate unusual hypotheses, e.g. about the existence of inaccessible cardinals, from the foundations of category theory by examining the kinds of closure conditions on "universes" that are actually needed in practice. Here the aim is to set up a formal theory to encompass unrestricted statements from the theory of structures and categories more or less as they are presented to us. As discussed in the preceding section, we have not hesitated to assume the existence of inaccessible cardinals to establish the existence of such a theory. These aims are evidently not in conflict. Nevertheless, one might expect that an improved solution for the present aims would converge with the other ones.

Appendix I. Models of $L_{\infty\omega}$ sentences with indiscernibles.

The following definitions and facts are taken from Keisler [K] (cf. espec. Lectures 13, 14) and Barwise and Kunen [B-K]. The formulas of $L_{\infty\omega}$ are built up using arbitrary conjunctions and disjunctions in addition to the usual finitary operations of the 1st-order predicate calculus. By a fragment L_A of $L_{\infty\omega}$ we mean one with a set Fm_A of formulas closed under subformulas and the finitary operations. L_A is a Skolem fragment if it contains for each formula $\exists x\, \phi(x,y_1,\ldots,y_n)$ $(n \geq 0)$ an n-ary function (or constant) symbol $f_{\exists x\phi}$; every fragment can be enlarged to a Skolem fragment. If L_A is such, Sk_A denotes the set of sentences in L_A of the form

1) $\forall y_1,\ldots,y_n[\exists x\, \phi(x,y_1,\ldots,y_n) \rightarrow \phi(f_{\exists x\phi}(y_1,\ldots,y_n),y_1,\ldots,y_n)]$

In the following, L_A is assumed to be any Skolem fragment.

$\mathfrak{M}$ ranges over one-sorted L_A-structures; the domain of $\mathfrak{M}$ is denoted by M. $(\mathfrak{M},a_1,\ldots,a_n)$ is the structure $\mathfrak{M}$ with $a_1,\ldots,a_n$ adjoined as distinguished elements. Given a linearly ordered structure $(I,<)$ with $I \subseteq M$, I is called a set of n-variable indiscernibles for $\mathfrak{M}$ (relative to L_A) if whenever $a_1,\ldots,a_n, b_1,\ldots,b_n \in I$ and $a_1 < \ldots < a_n$ and $b_1 < \ldots < b_n$ and $\phi(x_1,\ldots,x_n)$ is in L_A then

(2) $\mathfrak{M} \models \phi(a_1,\ldots,a_n \Leftrightarrow \mathfrak{M} \models \phi(b_1,\ldots,b_n)$

is called a set of indiscernibles if it is a set of n-variable indiscernibles for each n.

Lemma 1 [B-K] (p. 314). Suppose that for each n, $\mathfrak{M}_n$ is a model of Sk_A and $(I_n, <_n)$ is an infinite linearly ordered set of n-variable indiscernibles in $\mathfrak{M}_n$. Suppose further that for all $a_1 < \ldots < a_n$ in $(I_n, <_n)$ and $b_1 < \ldots < b_n$ in $(I_{n+1}, <_{n+1})$ we have

(4) $(\mathfrak{M}_n, a_1,\ldots,a_n) \equiv_{L_A} (\mathfrak{M}_{n+1},b_1,\ldots,b_n)$.

Then given any infinite linearly ordered set $(I,<)$ there is a model $\mathfrak{N}$ for which

$(I,<)$ is a set of indiscernibles and

(5) $(\mathfrak{M}_n, a_1, \ldots, a_n) \equiv_{L_A} (\mathfrak{N}, b_1, \ldots, b_n$

whenever $a_1 < \ldots < a_n$ in I_n and $b_1 < \ldots < b_n$ in N. In particular, $\mathfrak{M}_n \equiv_{L_A} N$ for all n.

The following definitions and results are taken from Jensen [J]; some of the hypotheses are weakened for simplicity. $[C]^n$ is the set of all subsets X of (with cardinality $(X) = n$. A partition of $[D]^n$ in C is a map

(6) $f: [D]^n \to C.$

$I \subseteq D$ is a set of indiscernibles for f if $f([I]^n) = \{c\}$ for some $c \in C$. $\langle f_n \rangle_{n \geq 1}$ is a family of partitions f_n of $[D]^n$ in C then $\langle c_n \rangle_{n \geq 1}$ is called realizable if for every n and every $\beta < \delta = |D|$ there exists $I_n \subseteq D$ with $|I_n| = \beta$ and $f_k([I_n]^k) = \{c_k\}$ for every $k \leq n$.

Lemma 2 [J] (p. 288). If $|C| = \gamma$, $|D| = \delta$, $\gamma < \delta$ and δ is inaccessible then any family $\langle f_n \rangle_{n \geq 1}$ of partitions f_n of $[D]^n$ in C has some realizable sequence $\langle c_n \rangle_{n \geq 1}$.

(This is proved by making use of the Erdös-Rado generalization of Ramsey's Theorem.

Lemma 3. Suppose δ is inaccessible and $|Fm_A| < \delta$. Suppose $\mathfrak{M}$ is a structure $\mathfrak{M} = (M, \ldots)$ and that $I_o \subseteq M$ with $|I_o| = \delta$. Let $<$ be any linear ordering of I_o. Then there is for each n an infinite set $I_n \subseteq I_o$ of n-variable indiscernibles such that for $a_1 < \ldots < a_n$ in I_n and $b_1 < \ldots < b_n$ in I_{n+1} we have

(7) $(\mathfrak{M}, a_1, \ldots, a_n) \equiv_{L_A} (\mathfrak{M}, b_1, \ldots, b_n)$.

Proof. Let $\gamma = 2^{|Fm_A|}$, so $\gamma < \delta$. Now, as in [J], consider the following family of partitions f_n of $[I_o]^n$ in $P(Fm_A)$:

(8) <u>when</u> $a_1 \ldots < a_n$, $f_n(\{a_1,\ldots,a_n\}) =$

$$= \{\phi(x_1,\ldots,x_n) \mid \phi \text{ is in } Fm_A \text{ with at most } x_1,\ldots,x_n \text{ free}$$

$$\text{and } \mathfrak{M} \models \phi(a_1,\ldots,a_n$$

By Lemma 2 there exists a realizing sequence $\langle T_n \rangle_{n \geq 1}$ for $\langle f_n \rangle_{n \geq 1}$; each $T_n \subseteq Fm_A$. By the definition above, there is for each n an infinite set $I_n \subseteq I_o$ for which

(9) $f_k([I_n]^k) = \{T_k\}$ <u>for each</u> $k \leq n$.

In particular,

(10) $f_n([I_n]^n) = T_n$ <u>and</u> $f_n([I_{n+1}]^n \quad T_n\}$

Hence if $a_1 < \ldots < a$ in I_n and $b_1 < \ldots < b_n$ in I_{n+1} we have

$$f_n(\{a_1,\ldots,a_n\}) = T \quad \text{and} \quad f_n(\{b_1,\ldots,b_n\}) = T_n$$

so that for any $\phi(x_1,\ldots,x_n)$ in L_A

$$\mathfrak{M} \models \phi(a_1,\ldots,a_n) \Leftrightarrow \mathfrak{M} \models \phi(b_1,\ldots,b_n) ;$$

.e. (7) holds.

<u>Theorem.</u> <u>Suppose</u> δ <u>is inaccessible and</u> $|Fm_A| < \delta$. <u>If</u> $\mathfrak{M}$ <u>is a model of</u> Sk_A <u>and</u> $I_o \subseteq M$ <u>with</u> $|I_o| = \delta$ <u>and</u> $(I_o,<)$ <u>is linearly ordered then for any other linearly ordered</u> $(I,<)$ <u>we can find a model</u> $\mathfrak{N}$ <u>for which</u>

(i) I <u>is a set of indiscernibles for</u> $\mathfrak{N}$,

(ii) $\mathfrak{M} \equiv_{L_A} \mathfrak{N}$, <u>and</u>

(iii) <u>if</u> $\mathfrak{M} \models \phi(a_1,\ldots,a_n)$ <u>for all</u> $a_1 < \ldots < a_n$ <u>in</u> I_o <u>(where</u> ϕ <u>is in</u> L_A) <u>then</u> $\mathfrak{N} \models \phi(b_1,\ldots,b_n)$ <u>for all</u> $b_1 < \ldots < b_n$ <u>in</u> I.

Proof. This is now a corollary of Lemmas 1 and 3.

Remark. The main new point is that $\mathfrak{N}$, I can be chosen to satisfy prescribed conditions in the sense of (iii). But even without (iii), the formulation seems to be new.

Appendix II. Some formal systems for structural properties relative to any universe.

The systems considered here develop a suggestion due to Kreisel giving a different means for treating statements of the sort considered in §1.(14) This is obtained by analyzing a common explanation of such, which runs something as follows for the first statement of §1 (to take a specific example),

(1) $(PO,S) \in PO.$

Namely, consider any reasonably closed universe of sets V_0. Let $C_{V_0}^{PO}$, $C_{V_0}^{S}$ be respectively the classes of all structures $x = (x_1,x_2)$ in V_0 which are partially ordered and of all pairs $(x,y) = ((x_1,x_2), (y_1,y_2))$ of such structures which are in the substructure relation $x \subseteq y$. Then the pair $(C_{V_0}^{PO}, C_{V_0}^{S})$ itself satisfies the property of being partially ordered. In this sense (1) is true.

More generally, suppose given a 1st-order property A expressed by a 1st order formula θ_A. Thus we know exactly what is meant by saying that a structure $x = (x_1,\ldots,x_n)$, consisting of definite sets $x_1,\ldots,x_n$, satisfies A; namely, $x \models \theta_A$. In particular, this is determined if x lies in a specified universe V_0 (each $x_i \in V$ of V_0 (which we would call cla Hence for each such A and V_0 we can

o of A on V_0,

(2) $C_{V_0}^{A} = \{x \mid x = (x_1,\ldots,x_n$ with each $x_i \in V_0$, and $x \models \theta_A\}$.

Now suppose all the properties $A,B,\ldots,X,Y,Z$ we have to deal with determine extensions in this way. Then we can explain

(3) $(X_1,\ldots,X_n \in A$,

i.e. that the property A applies to the n-tuple of properties $(X_1,\ldots,X_n)$ (15) as follows:

(4) $(C_{V_o}^{X_1},\ldots,C_{V_o}^{X_n})$ satisfies the property A, for any reasonable closed universe V_o.

This does not pretend to be a general explanation of the relation of application between properties; it depends on dealing with properties for which we have a prior set-theoretical definition of satisfaction. Thus nothing is said of properties expressed by formulas involving quantified variables which are themselves interpreted as ranging over properties.

The next step is to set up a formal system in which statements like (1) can be established directly, following this approach. This can be done in the language L^*. Moreover, it is not necessary to restrict attention to 1st-order properties θ_A of structures in the usual sense; we can deal with properties expressed by formulas $\theta(X_1,\ldots,X_n)$ of L^* in which all the quantified variables range over V_o, and where the determination of whether θ holds depends only on the extensions of the X_i on V_o. (In L^* we read: x is in the extension of X, for $x \in X$.) Finally, θ may contain meta-class parameters under the same conditions. (We continue to call $A,B,C,\ldots,X,Y,Z$ variables for meta-classes, though we have more definite properties in mind here.) Then a formula θ of L^* is said to be V_o-determined if:

(5) (i) each bound variable of θ is a set-variable, and

(ii) if X is a meta-class variable which occurs in θ then each such occurrence is in an atomic formula of the form $(s \in X)$, where s is a set-term, and

(iii) the same for the constant 'V_o' in place of 'X'

Set-terms are built up from set-variables only by means of pairing.

Let TP_1 be the theory obtained from the theory S^* of §3.1 in the following way. The stratified comprehension scheme I is replaced by:

I_{TP_1} $(\exists A)(\forall X)\{X \epsilon A \leftrightarrow (\exists X_1,\ldots,X_n)[X = t(X_1,\ldots,X_n \wedge \theta]\}$

where θ is V_0-determined, t is a term with variables $X_1,\ldots,X_n$ each of which occurs in θ (but 'A', 'X' do not).

Axioms II-IX are the same as for S^*, but the axiom X of Universal Choice is replaced by one of the usual forms of the Axiom of Choice for sets[16]. Thus TP_1 is a subtheory of S^*. However, the consistency of TP_1 can be proved by much more elementary methods than for S^*, more or less following the introductory ideas. Briefly, this proceeds as follows:

Let $M = R_\kappa$ where κ is inaccessible. We interpret the set variables as ranging over M and V_0 as M. The language $L^*(M)$ has set-constant symbols for each element of M. A formula θ of $L^*(M)$ is said to be M-determined just as in (5), though now set-terms may contain set-constants. Let $\mathscr{T}_0$ be the class of pairs $\langle t,\theta\rangle$ where θ is M-determined and t,θ have the same free meta-class variables $X_1,\ldots,X_n$, and only these free variables Put

(6) $$C_M^{\langle t,\theta\rangle} = \{t(a_1,\ldots,a_n) \mid a_1,\ldots,a_n \in M \text{ and } (M,\epsilon) \models \theta(a_1,\ldots,a_n)\}$$

Let $\mathscr{T}$ be the closure of $\mathscr{T}_0$ under pairing; $\mathscr{A},\mathscr{B},\ldots$ range over $\mathscr{T}$. Then $C_M^{\mathscr{A}}$ is extended to arbitrary $\mathscr{A}$ in M in such a way that each $C_M^{\mathscr{A}} \subseteq M$; here we use the pairing operation $(C_0,C_1) = (C_0 \times \{0\} \cup C_1 \times \{1\})$ for subclasses C_i of M. Define

(7) $$\mathscr{B}\,\eta\,\mathscr{A} \Leftrightarrow \mathscr{A}\epsilon\mathscr{T}_0,\ \mathscr{B}\epsilon\,\mathscr{T},\ \mathscr{A} \text{ is of the form } \langle t,\theta\rangle \text{ with free vars. } X_1,\ldots,X_n,$$
$$\mathscr{B} \text{ is of the form } t(\mathscr{B}_1,\ldots,\mathscr{B}_n) \text{ and } M \models \theta(C_M^{\mathscr{B}_1},\ldots,C_M^{\mathscr{B}_n})$$

Note that only $\mathscr{T}_0$ objects have η-members. We can identify two such objects if they agree on $\mathscr{T}$. We can also identify certain $\mathscr{A}$ with elements $\underline{a}$ of M. Then η gets mapped into a relation ϵ which satisfies the axioms of TP_1. For verification of the schema I_{TP_1}, one first checks the case without parameters and then gets the general case by a substitution argument.

Question. Is this model already a model of S^*? It seems unlikely but it is not easy to test impredicative formulas in it. Perhaps a model of S^* could be built by a certain transfinite iteration of the process used in forming this model

Now we return to the formulation of the unlimited theory of structures in TP_1. For any V_0-determined $\theta(X_1,\ldots,X_n,A_1,\ldots,A_n)$, the meta-class $[t(X_1,\ldots,X_n)|\theta(X_1,\ldots,X_n,A_1,\ldots,A_m)]$ provably satisfies its defining condition. In particular we can define $PO = [(X_1,X_2)|\theta_1(X_1,X_2)]$ and $S = [((X_{11},X_{12}),(X_{21},X_{22})|\ \theta_2(X_{11},X_{12},X_{21},X_{22})]$ taking θ_1 to be the conjunction of formulas $Str_\sigma(X_1,X_2)$ and $\overline{\theta}_{PO}(X_1,X_2)$ of 2.1 with all quantifiers relativized to V_0, and similarly θ_2 using $\overline{\Psi}_{Sub}(X_{11},X_{12},X_{21},X_{22})$; in other words we simply write out the properties of (X_1,X_2) being partially ordered or of the substructure relation, using only set-variables with the quantifiers[17] (restricted quantifiers $(\exists x \in X(\ldots)$ are all right since these are really of the form $(\exists x)(x \in X \wedge \ldots))$. Then $TP_1 \vdash (PO,S) \in PO$. In the proof of this we consider only the sets (x_1,x_2) which belong to PO. This is in accord with the idea at the beginning of this section. We can also treat the statement §1(3) in the same way. But §1(2) and §1(4) present special problems for TP_1. We first consider the latter.

Given two categories A,B we can form the structure $B^A = (Funct(A,B), Nat(A,B),\ldots)$; we use here V_0-determined formulas with parameters A,B to define $Funct(A,B)$, etc. (This essentially reduces to the definition of $X\colon A_1 \to A_2$ by the V_0-determined property $(\forall x \in A_1)(\exists! y \in A_2)[(x,y) \in X] \wedge (\forall z \in X)(\exists x \in A_1)(\exists y \in A_2)(z=(x,y))$.) The problem for §1(4) is not the definition of B^A, but the proof that $(B^A \in Cat)$. Here we are to use the definition of Cat in TP_1 which requires us to verify that a given $(O,M,\ldots)$ belongs to Cat by quantifying over all sets in O, sets in M, etc. In the case of B^A, this means quantifying over all sets which are functors from A to B. If A,B are "large" categories such as Grp, Class, etc., there simply are no such sets.

As to the statement of §1(2) in TP_2, the problem is that this uses the 2nd order notion of well-ordering which is not realized by an object WO of TP_1, in the sense that it is not given by a V_0-determined property in the above sense. However, we can expand TP_1 to a theory TP_2 for a theory of properties of sets and classes, i.e. properties of elements of V_0 and PV_0, where we may also allow quantifiers over PV_0. TP_2 can be proved consistent in much the same way as TP_1. In TP_2 we can now give sense to application of certain 2nd order properties to other 2nd order properties. The problem about functor categories disappears partly, since we can prove $(\forall A, B \in PV_0)[A, B \in Cat \rightarrow B^A \in Cat]$. But we are still blocked from establishing this for arbitrary B, A. One has further problems with a general formulation of Yoneda's lemma. Finally, the formalism of TP_2 is more cumbersome then would be hoped for.

Appendix III. Unrestricted pairing in stratified systems; a problem.

To overcome the main defect of L_{TP} or L_{TPS} stratification, but still using some sort of stratification set-up, we would want to start with a typed language in which objects from arbitrary types can always be paired. One way of doing this was presented in §2.3 Remark (ii), denoted L'_{TP}. But a more natural way is to use the following language L''_{TP}. First the type symbols (t.s.) are generated inductively as follows:

(1) (i) 0 is a t.s.;

(ii) if σ, τ are t.s. then so also are (σ, τ) and $[\tau]$.

The idea is that (σ, τ) is the type of a pair of objects of respective types σ and τ and that $[\tau]$ is the type of a collection of objects of type τ. The basic symbols of L''_{TP} are then specified as follows:

(2) (i) variables $A^\tau, B^\tau, C^\tau, \ldots, X^\tau, Y^\tau, Z^\tau$ for each t.s

(ii) relations $=, \in$

(iii) binary operation (,).

The sets of terms of type τ are defined inductively by:

(3) (i) each variable of type τ is a term of type τ;

(ii) if s is a term of type σ and t is a term of type τ then is of type (σ, τ).

The atomic formulas of L''_{TP} are just those of the form

(4) (i) $t_1 = t_2$ for type (t_1) = type (t_2) and

(ii) $t_1 \in t_2$ for type $(t_1) = \tau$, type $(t_2) = [\tau]$.

L''_{TP}-formulas are built up using $\sim$, $\vee$, $(\exists X^{\tau})$ for any t.s. τ. It is then explained in the usual way when a formula ϕ of L_1^* is said to be L''_{TP}-stratified.

This leads us to consider a system S''_1 based on L''_{TP}-Stratified Comprehension and Pairing, as basic Axioms. L''_{TP}-stratification is finer than L'_{TP}-stratification since the typing doesn't confuse pairs with collections. Nevertheless, the system S''_1 is also inconsistent and for the same reason as S'_1. The axioms (*) of S'_1 which lead to Russell's Paradox are instances as well of L''_{TP}-Stratified Comprehension. Slightly rewritten, they are:

(5) (i) $(\exists A_1)(\forall X)\{X \in A_1 \leftrightarrow (\exists Y,Z)(X=(Y,Z) \wedge Y \in Z)\}$

(ii) $(\exists A)(\forall X)\{X \in A \leftrightarrow (X,X) \notin A_1\}$.

Note that the only way to stratify (i) in the present set-up is to assign Y some type τ, then Z type $[\tau]$, X type $(\tau,[\tau])$ and A_1 type $[(\tau,[\tau])]$. The only way to stratify (ii) is to assign X some type σ and then A_1 type $[(\sigma,\sigma)]$ (and finally A type $[\sigma]$). In other words, there is something incoherent in this set-up, since we may substitute a stratified definition for a parameter in a stratified formula, where the result cannot be stratified.

There are certain formulas ϕ for which this situation does not arise. Write ϕ as $\phi(X,A_1,\dots,A_n)$ with distinguished variable X and parameters $A_1,\dots,A_n$. Call ϕ uniformly stratified if whenever $\theta_1(X),\dots,\theta_n(X)$ are stratified and without parameters then

(6) $$\bigwedge_{1\le i\le n} (\forall X)\ [X \in A_i \leftrightarrow \theta_i(X)] \rightarrow \phi(X,A_1,\dots,A_n)$$

is also stratified. We may think of ϕ as determining a stratified operation $F(A_1,\dots,A_n)$ which takes stratified definable A_i to the result of substituting their definitions in $[X|\phi(X,A_1,\dots,A_n)]$.

Now the formulas

(7) i) $(X,X) \in A_1$, (ii) $X \in A_1 \wedge X \in A_2$

are <u>not</u> uniformly stratified though they are stratified. For example, for (i), take θ_1 so that $(\forall X)[X \in A_1 \leftrightarrow \theta_1(X)]$ can be stratified only if A_1 receives a type of the form $(\tau,[\tau])$. There are non-trivial uniformly stratified formulas, for example

(8) (i) $X \in A_1 \times A_2$ <u>and</u> (ii) $X: A_1 \to A_2$

where $((X \in A_1 \times A_2)$ is $(\exists Y,Z)\{X = (Y,Z) \wedge Y \in A_1 \wedge Z \in A_2\}$, and $(X: A_1 \to A_2)$ is as defined in §2.1(6).)

The notion of uniformly stratified formula applies to any stratification set-up It turns out that <u>every</u> L_{TP} <u>or</u> L'_{TP}-<u>stratified formula is uniformly stratified</u>. This follows simply from the fact that if we can stratify ϕ by ϕ^+ in one of those languages, then the result ϕ^{+m} of raising al' types by m in ϕ^+ also gives a stratification of ϕ. Thus restricting attention to uniformly stratified formulas is not proof against contradiction.

Nevertheless it seems to me that the following system S_2^* may well be consistent: S_2^* has the schema of comprehension restricted to L''_{TP}-uniformly stratified formulas and the Pairing Axiom.

<u>Question</u>. <u>Is</u> S_2^* <u>consistent</u>?

By (7)(i) the immediate inconsistency (5) is blocked. In any case, even though we may overcome the defect of pairing in this way, the restriction in S_2^* is quite severe. It is not implausible to reject the formation of $A_1 \cap A_2$ in general, but DA_1 and $\breve{A}_1$ are not forthcoming for <u>relations</u> A_1, and these as well as other operations should be available.

<u>Problem</u>. <u>Assuming the answer to the question is positive, find a simple consistent extension of</u> S_2^* <u>in which the unlimited theory of structures can be developed without obstacles</u>.

FOOTNOTES

(1) Guggenheim Fellow 1972-73; visiting the Mathematical Institute, Oxford, Fall 1972 and U.E.R. de Mathématiques, Université Paris VII, Spring 1973. I am grateful to the Guggenheim Foundation and to these institutions for their kind and generous assistance during which most of the work for the present paper was carried out. This had its source in work (done at Stanford University) summarized in Appendix II. The material there was presented in extended form in a talk to the Congrès de Logique d'Orleans Sept. 3-14, 1972 under the title 'The "category of all categories" etc., in a theory of classes and properties.' The new theories here provide an improvement in a particular but important respect, necessitating a considerable change in presentation.

(2) The principal means (in category theory) are in terms of Grothendieck's idea of "universes" satisfying certain set-theoretical closure conditions, developed further by Gabriel and Sonner, and by MacLane's distinction between "small" and "large" categories in terms of a theory of sets and classes; c.f. further [M1] and [F].

(3) I have in mind here particularly Quine's [Q1], type-free systems of Ackermann developed further by Schütte [S] Chp. VIII, and Gilmore [G]. The work on type-free combinatory or λ-calculi seems so far to be only marginally related.

(4) The choice of language for the kind of systems we are seeking is not uniquely determined. In particular, the familiar simplifications made here which are perfectly well justified in the usual set-theoretical context could be taking us off on the wrong path at this point. It may be that operations and relations should play quite different roles, as they do in constructive mathematics. Martin-Löf [ML] gives an attractive formulation of the latter which brings this out.

(5) To be more precise, is treated as defined in terms of ϵ in NF and Extensionality is rewritten accordingly.

(6) I am indebted to Dana Scott for bringing aspects of this defect to my attention and for related discussions.

(7) It is even worse in the case of $G \in \Pi_{X \in A} F(X)$ where we would have to use the notion of function in two different senses for F and G, in order to stratify the definition

(8) Note that the replacement in §2.1 of designated individuals c in structures by their singletons $\{c\}$ already implicitly used the device of a separate pairing operation.

(9) This should not be thought of as expressing that all conceivable sets belong to V_0, only that V_0 is the universe of sets of initial interest. We reserve the word 'set' for members of V_0 when working in S^*. But in the models obtained of S^* not only the members of V_0, but also V_0 itself, PV_0, etc. will be standard sets.

(10) Technically it would be preferable to define ϕ to be stratified in a given stratification set-up) just in case $\phi = \psi^*$ for some typed ψ. But in practice we want to see whether a given untyped ϕ can be written as such ψ^*; we do this by assigning type indices as possible.

(11) We assume familiarity with these; cf. e.g. [M2].

(12) An ambiguity of notation is introduced here; this will be compounded below. The context will determine whether we are dealing with maps $Z: X \to Y$ in the sense of §2.1, or the relation $\alpha: X \to Y$ in a category A as just defined, or $F: A \to B$ when F is a functor from A to B, as explained below. It should also be noted that for simplicity we have omitted explicit use of I in A where $I: O \to M$, $I(X) = 1_X$ for each $X \in O$. However, in detailed explanations of the conditions to be a category, a functor, etc., simplicity would require displaying I.

(13) We define a <u>class</u> in S^* to be a member of PV_0. Weak extensionality does not exclude the possibility of more than one empty class. To consider only such one would define the classes to be the members of $\{0\} \cup [X | X \subseteq V_0 \wedge \exists y(y \in X)]$. Note that morphisms between classes are triples (F,X,Y) where X,Y and hence F are classes. We can without harm identify such triples again with classes

Otherwise we should consider the closure of V_1 under pairing instead of V_1.

This suggestion was presented in unpublished notes. Some points of difference of the present development are noted below.

Bernays had used the symbol 'η' for the relation of application to distinguish it from the relation of membership in his theory of sets and classes, thinking of classes as properties. Kreisel follows Bernays in this notation, especially since one is now considering axioms which are distinctive for properties. I have used the same symbol 'ϵ' to simplify the relationship with S^*.

16) The main difference of TP_1 from the theory proposed by Kreisel is in permitting meta-class parameters in the comprehension scheme. This seems necessary if one is to try to account for statements like §(1(4); cf. below.

To be sure there is no confusion, note that this is different from the definition of PO and S in the system S^* (§2.4(4)) where the variables in the quantifiers range over arbitrary meta-classes.

BIBLIOGRAPHY

[B-K] J. Barwise and K. Kunen, Hanf numbers for fragments of $L_{\infty\omega}$, Israel Journ. Math. 10 (1971), 306-320.

E-M] A. Ehrenfeucht and A. Mostowski, Models of axiomatic theories admitting automorphisms, Fund. Math. 43 (1956), 50-68.

E-R] P. Erdös and R. Rado, A partition calculus in set theory, Bull. A.M.S. 62 (1956), 427-488.

[F] S. Feferman, Set-theoretical foundations of category theory (with an Appendix by G. Kreisel) Springer Lecture Notes 106 (1969), 201-247.

[G] P.C. Gilmore, Attributes, sets, partial sets and identity, Compositio Math. 20 (1970). 53-79.

R.B. Jensen, On the consistency of a slight (?) modification of Quine's New Foundations, in Words and Objections: Essays on the work of W.V. Quine, eds. D. Davidson and J. Hintikka, Dordrecht 1969, 278-291.

[K] H.J. Keisler, Model theory of infinitary logic, Amsterdam 1971.

[M1] S. MacLane, Categorical Algebra and set-theoretic foundations, in Axiomatic Set Theory (Proc. A.M.S. Sympos. Pure Math. Vol. XIII, Part I) Providence 1971, pp. 231-240.

S. MacLane, Categories for the working mathematician, New York 1971.

[M-L] P. Martin-Lof, An intuitionistic theory of types, to appear.

[M-V] R. Montague and R.L. Vaught, Natural models of set theories, Fund. Math. 47 (1959), 219-242.

[Q1] W.V. Quine, New foundations for mathematical logic, Amer. Math. Monthly 44 (1937), 70-80.

[Q2] W.V. Quine, On ordered pairs, Journ. Symbolic Logic 10 (1945), 95-96.

[S] K. Schütte, Beweistheorie, Berlin, 1960.

[Sp] E. Specker, Typical ambiguity in logic, in Methodology and Philosophy of Science; Proc. of the 1960 International Congress (eds. Nagel, Suppes, Tarski), Stanford 1962, pp. 116-123.