Zeitschr. f. math. Logik und Grundlagen d. Math.
Bd. 33, S. 461–480 (1987)

A POSSIBLE MODAL FORMULATION OF COMPREHENSION SCHEME

by JAN KRAJÍČEK in Prague (Czechoslovakia)[1])

§ 0. Introduction

We will propose a set theory MST formalized in modal logic and we will try to show that its consequences are relatively powerful in relation to its simple axiomatization. The result was announced in [10].

The naive comprehension scheme

(CA) $\exists y\, \forall t(t \in y \equiv \varphi(t))$,

where φ is any property, is a very elegant principle. This principle, or better its substance, describes the naive set-universe. Unfortunately, in the most customary formalization, where any formula of the set-theoretical language is accepted as a property, (CA) is inconsistent. As it is well-known the contradiction can be proved e.g. by following RUSSELL's argument. Thus various approaches try to formulate a set-theory which would be mathematically powerful and consistent. They replace (CA) by a list of weaker axioms guaranteeing the existence of some sets (e.g. ZF) or restrict it (e.g. NF). But, at the same time, they lose important features of (CA): homogeneity, simplicity and elegance or a certain intuitive picture.

In ZF, for example, we have axioms which suggest how to make new sets from already formed sets, and we have enough of them that the described set-universe is sufficiently complex to be mathematically rich. An underlying universal idea of the ZF-axioms is the so called "size doctrine". This principle together with the axiom of foundation gives us a sufficiently clear intuitive picture about the ZF-universe. But the homogeneity of assumptions is lost and nothing guides us in choosing of new axioms from a variety of mutually incomparable possibilities.

A different approach is formalized by QUINE's NF. Here (CA) is applied only to the "stratified" formulas. The source of this modification lies in the type theory. It is syntactical rather than semantical; thus there is not a clear picture about the described universe. Serious difficulties arise when one tries to develop mathematics in NF. The theory must be supported by additional axioms to be sufficiently strong (see [7]).

We hope that the features of (CA) justify the program of studying modifications of (CA).

We will describe a set theory MST which is a formalization of our modification of (CA) using a modal logic. We also show that MST covers some nontrivial mathematics, in particular it contains PEANO's arithmetic. A partial consistency result is proved in Chapter 8. Some new partial consistency results together with some results concerning mathematics in MST were obtained after finishing this paper. They are partly reported in [11] and partly in preparation.

[1]) We thank P. PUDLÁK for many discussions and valuable remarks. His assistance was essential to our work.

In the literature we found a few systems closely connected to MST (see [1], [2], [3], [5], [6] and [8]). These connections will be discussed in detail in Chapter 8. FEFERMAN's paper [4] contains a general discussion of various approaches to some foundational problems. It also offers a wide relevant bibliography.

§ 1. The formal theory MST

We begin this Chapter with a motivation in terms of knowledge. However, it should be understand only like a heuristic. We think that a serious epistemic interpretation of the following theory is possible but is not needed for the present paper. Our goal here is to propose a set-theory based on some modal reformulation of (CA) to which we were lead by the following considerations.

Let us imagine the following situation: There exists a set-universe which is the object of our consideration. The only atomic predicates are "to be equal" and "to be an element of". Each atomic sentence and hence each sentence is true or false in the set-universe. Our wish is to recognize the truth, i.e. the sentences true in the set-universe. So some true sentences are known to us. Still others could become known, i.e. are in principle knowable,

For formalizing the modal operator "to be knowable" we extend the usual classical set-theoretical language by adopting a new unary logical connective $\Box$ which should be an epistemic modality. Thus our language (the modal set-theoretical language) is the modal predicate calculus with identity (see [9]) with a binary predicate $\in$ as the only non-logical symbol.

1.1. When we decide to try to understand the set-universe we can already take the fact of looking for the knowledge as a part of the knowledge. Put otherwise, we may accept assumptions which manifest the principles and the correctness of our knowledge. Hence the following two axiom schemes and one deduction rule should be accepted:

(i) $\Box\varphi \to \varphi$

(ii) $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$

(i), (ii): T-axioms,

(iii) $\dfrac{\varphi}{\Box\varphi}$ Necessitation rule (N-rule).

This (it will be specified later) extension of the classical predicate calculus is called T in [9].

1.2. The main idea of MST is that (CA) does not refer to the whole set-universe but only to its knownable part, to our "universe of discourse". That means: it seems to us from the point of view of our knowledge that the set-universe behaves as if (CA) were sound.

In the chosen language this modification of (CA) (Modal (CA) or shortly (MCA)) can be described as follows:

(MCA) For any formula $\varphi(t, a_1, \ldots, a_k)$ of the modal set-theoretical language with free variables among $t, a_1, \ldots, a_k$ the universal closure of the following formula holds:

$$\exists y\, \forall t((\Box\varphi(t, a_1, \ldots, a_k) \equiv \Box t \in y) \,\&\, (\Box\neg\varphi(t, a_1, \ldots, a_k) \equiv \Box t \notin y)).$$

The a_i's are called *parameters* and will be omitted further.

1.3. So far we have not accepted any concrete "theory of knowledge", any non-logical epistemic assumptions. The second principle of MST is of this kind. As usual $\Diamond\varphi$ abbreviates $\neg\Box\neg\varphi$. The principle is

(LP) $\Diamond x = y \rightarrow \Box x = y$.

There are other (in T) equivalent formulations of (LP)

(i) $\Box x = y \vee \Box x \neq y$

or

(ii) $(x = y \rightarrow \Box x = y) \,\&\, (x \neq y \rightarrow \Box x \neq y)$.

We leave aside the question whether (LP) has logical or empirical character.

(LP) is a kind of a finitistic assumption. There are also close connections to LEIBNIZ's principle "No two monads are exactly alike" from his theory of monads (see [13]). This may be formalized as $\Diamond x = y \rightarrow x = y$ (where $\Diamond x = y$ simulates indistinguishability) or equivalently $x \neq y \rightarrow \Box x \neq y$. Thus LEIBNIZ's principle coincides with the second conjunct of (ii). Surely, the first is a trivial consequence of the substitution properties of identity. Hence the name (LP) seems to be justified for this axiom.

We remark that (LP) is also discussed in [9]; the first conjunct of (ii) is called (LI) and the second (LNI) there.

1.4. Let us summarize the definitions. The language of the formal theory MST, the modal set-theoretical language, contains the classical propositional connectives $\neg$, &, $\vee$, $\rightarrow$, $\equiv$, the quantifiers $\exists$, $\forall$, and the new unary (modal) connective $\Box$. Variables $x, y, \ldots, x_0, x_1, \ldots, a, b, \ldots$ ranges over individuals. There are only two binary predicates: $=$ (identity) and $\in$ (membership). Formulas of the modal set-theoretical language are all formulas built up in this language. In particular, formulas may contain free variables. Axioms and rules of the modal predicate calculus T are:

(i) the substitution-instances of all classical propositional tautologies for all formulas,

(ii) the substitution-instances of the schemes (and their variants replacing x, y by other variables)

  a) $\varphi \equiv \forall x\varphi$, provided x is not free in φ,

  b) $\forall x\varphi(x) \rightarrow \varphi(y)$,
  provided x does not occur in φ in the scope of a quantifier $\exists y$ or $\forall y$,

  c) $\forall x(\varphi \rightarrow \psi) \rightarrow (\forall x\varphi \rightarrow \forall x\psi)$, for all formulas φ, ψ,

(iii) the substitution-instances of the schemes

$$\varphi \rightarrow \varphi \quad \text{and} \quad \Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi$$

the axioms (and their variants using other variables)

$$x = x, \quad (x = y) \equiv (y = x), \qquad (x = y \,\&\, y = z \rightarrow x =$$

$$(x_0 = x_1 \,\&\, y_0 = y_1) \rightarrow (x_0 \in y_0 \equiv x_1 \in y_1$$

(v) the inference rules

modus ponens $\dfrac{\varphi, \varphi \to \psi}{\psi}$

generalization $\dfrac{\varphi}{\forall x \varphi}$ (and its variants using other variables than x),

necessitation (N-rule) $\dfrac{\varphi}{\Box \varphi}$

for all formulas φ, ψ.

The only non-logical assumptions of MST are (MCA) and (LP).

Let us explicitly stress that the N-rule is generally applicable (in contrast to theories of [2] and [3]). In particular, all instances of (MCA) are "knowable". This gives to the whole system features of a "logical calculus".

Instead of writing MST $\vdash \varphi$ we will shortly write $\vdash \varphi$. In the next Chapters various results will be proved in extended underlying logical systems over T. These extensions will be explicitly stressed by A $\vdash \varphi$, which denotes the deducibility of φ in MST + A.

We will use freely various results about modal logics which are proved in [9]. When we talk about equivalent formulas we always suppose that this equivalence can be established in (possibly extended) logic T.

§ 2. Russell's paradox

2.1. Let us discuss Russell's paradox formally. Applying (MCA) to Russell's formula $t \notin t$ we obtain

$$\vdash \exists y \, \forall t((\Box t \notin t \equiv \Box t \in y) \; \& \; (\Box t \in t \equiv \Box t \notin y))$$

and hence

$$\vdash \exists y(\Box y \in y \equiv \Box y \notin y).$$

Now surely

$$\vdash((\Box y \in y \to y \in y) \; \& \; (\Box y \notin y \to y \notin y))$$

and the only escape from the contradiction gives

$$\vdash \exists y((\Box y \in y \equiv \Box y \notin y) \; \& \; \Diamond y \in y \; \& \; \Diamond y \notin y).$$

Fortunately, this situation does not lead to inconsistency because $\Box y \in y \vee \Box y \notin y$ is not a theorem of T.

We even profit by this trivial but important corollary:

2.2. Corollary. $\vdash \exists y(\Diamond y \in y \; \& \; \Diamond y \notin y)$.

2.3. In the Russian translation of [7] Esenin-Volpin derived Russell's paradox in intuitionistic logic. If we try his derivation for MST it fails, because a scheme analogical to $(\varphi \to \neg \varphi) \to \neg \varphi$, on which it is based, i.e. $(\Box \varphi \to \Box \neg \varphi) \to \Box \neg \varphi$, is not provable in T. (It is, in fact, equivalent to $\Box \varphi \vee \Box \neg \varphi$.)

Another dangerous modification of RUSSELL's paradox is that of CURRY (see [7]). It goes in this way: Let ψ be any sentence. Take the set $y = \{t \mid t \in t \to \psi\}$. In particular for y

$$y \in y \equiv (y \in y \to \psi)$$

and ψ follows. The reader could calculate himself that this argument also fails for MST.

§ 3. A possible strengthening of the underlying logic

There is surely a wide spectrum of modal logics which may come to our attention if we consider possible extensions of the underlying logic of MST. Consult, for example, the book [9].

Let us list a few of the most obvious axiom schemes:

(i) S4: $\Box\varphi \to \Box\Box\varphi$,

(ii) S5: $\Diamond\Box\varphi \to \Box\varphi$,

(iii) B: $\Diamond\Box\varphi \to \varphi$,

(iv) BF: $\forall x\Box\varphi(x) \to \Box\forall x\varphi(x)$,

(v) $\Box\exists x\varphi(x) \to \exists x\Box\varphi(x)$.

S4 and S5 are known as LEWIS's systems. Axiom B is called BROUWER's axiom because of its relations to intuitionism (see [9], p. 58). BF is known as BARCAN's formula. The axiom (v) is also discussed in [9], p. 144.

The converses of all axioms except B are easily provable in T.

The problem of evidence for these assumptions is difficult. Surely, a necessary mathematical condition for adjoining any of them to MST is the consistency of the extended theory. Now we prove some limitations in this direction.

3.1. Lemma. MST + S5 *is inconsistent.*

Proof. First observe that, since T + S5 $\vdash$ T + S4, T + S5 $\vdash \Diamond\Box\varphi \to \Box\Box\varphi$. Now let us apply (MCA) to the Russell-like formula $\Box t \notin t$:

$$\vdash \exists y\, \forall t((\Box\Box t \notin t \equiv \Box t \in y)\ \&\ (\Box\Diamond t \in t \equiv \Box t \notin y))$$

and thus

(A) $\vdash \exists y((\Box\Box y \notin y \equiv \Box y \in y)\ \&\ (\Box\Diamond y \in y \equiv \Box y \notin y))$.

Since $\Diamond\Box\varphi \to \Box\Box\varphi$ is equivalent to $\Box\Box\varphi \vee \Box\neg\Box\varphi$, we have (by the above observation)

(B) S5 $\vdash \forall y(\Box\Box y \notin y\ \&\ \Box\Diamond y \in y)$.

Trivially we also have

$$\vdash \forall y(\Box\Box y \notin y \equiv \Box y \in y) \to \neg\Box\Box y \notin y$$

and hence, from (A), (B), S5 $\vdash \exists y(\Box\Diamond y \in y\ \&\ \Box y \notin y)$ and so

$$\text{S5} \vdash \exists y(\neg\Box y \notin y\ \&\ \Box y \notin y).$$

We are done.

By [9] T + S4 + B = T + S5 and thus the former system is inconsistent too.

3.2. The assumption (v) looks like a kind of a constructivistic principle. We feel it is inadequate and, in fact,

Lemma. MST + (v) *is inconsistent.*

Proof. We will use only the second conjunct of the statement (A) derived in the previous proof. Hence we begin with

$$\vdash \exists y(\Box\neg\Box y \notin y \equiv \Box y \notin y).$$

It follows, using also N-rule,

$$\vdash \Box\exists y((\Box\neg\Box y \notin y \equiv \Box y \notin y) \,\&\, \neg\Box y \notin y).$$

Now

$$(\mathrm{v}) \vdash \exists y(\Box(\Box\neg\Box y \notin y \equiv \Box y \notin y) \,\&\, \Box\neg\Box y \notin y \,\&\, \neg\Box y \notin y)$$

and thus

$$(\mathrm{v}) \vdash \exists y(\Box y \notin y \,\&\, \neg\Box y \notin y)$$

We are done.

Observe that we proved, in fact, slightly more: both the schemes

$$\exists y\, \Box\forall t(\Box t \in y \equiv \Box\varphi(t)) \quad \text{and} \quad \exists y\, \Box\forall t(\Box t \notin y \equiv \Box\neg\varphi(t))$$

are inconsistent (use $\varphi(t) \equiv \neg\Box t \in t$) to get the first one).

We mention an interesting remark due to P. PUDLÁK:

Remark. (LP) + (v) + $\exists x, y(x \neq y) \vdash \Diamond\varphi \to \Box\varphi$ and thus this system trivializes the operator $\Box$.

Proof. Abbreviate the above system by S. Certainly

$$\mathrm{S} \vdash \exists x, y((x = y \to \varphi) \,\&\, (x \neq y \to \neg\varphi)),$$

thus by N-rule and (v)

$$\mathrm{S} \vdash \exists x, y((\Box x = y \to \Box\varphi) \,\&\, (\Box x \neq y \to \Box\neg\varphi)).$$

Using (LP) we obtain $\mathrm{S} \vdash \exists x, y((x = y \to \Box\varphi) \,\&\, (x \neq y \to \Box\neg\varphi))$ and so

$$\mathrm{S} \vdash \Diamond\varphi \to \Box\varphi.$$

Using this remark and the ordinary RUSSELL's argument we get another proof of the inconsistency of MST + (v).

Let us now discuss some advantages of the other mentioned axioms.

3.3. Axiom S4 seems sound with 1.1 and we do not see a contradiction in adopting it. It has the following interesting consequence: If we define a string of $\neg$, $\Box$ before a formula to be "a modality", then in T we have an infinite number of distinct modalities, e.g. $\Box\varphi$, $\Box\Box\varphi$, $\Box\Box\Box\varphi, \ldots$, but in S4 there are only a finite number of nonequivalent modalities.

3.4. We do not see that B is sound with 1.1 but we also do not see that it is contradictory with MST. Adopting B has the following technical advantage: Imagine that we want to construct (by (MCA)) a set y of t's which satisfy φ. Instead of applying (MCA) to φ apply it to $\Box\varphi$. Hence we obtain:

$$\vdash \exists y\, \forall t((\Box t \in y \equiv \Box\Box\varphi(t)) \,\&\, (\Box t \notin y \equiv \Box\neg\Box\varphi(t))).$$

Since trivially $\Box t \notin y \to t \notin y$ we conclude, by contraposition, from the second conjunct

$$\vdash \exists y \forall t((\Box t \in y \equiv \Box\Box\varphi(t)) \& (\Box t \notin y \equiv \Box\neg\Box\varphi(t)) \& (t \in y \to \Diamond\Box\varphi(t)))$$

ıd hence

$$B \vdash \exists y \forall t((\Box t \in y \equiv \Box\Box\varphi(t)) \& (\Box t \notin y \equiv \Box\neg\Box\varphi(t)) \& (t \in y \to \varphi(t))).$$

Thus in the presence of B each "class" $\{t \mid \varphi(t)\}$ can be approximated in the sense of the formula above by a set.

3.5. BARCAN's formula may be seen as saying that we know the range of the set-universe, or better that the accessible part of the set-universe reflects all possible properties of individuals. This is suggested by an equivalent formulation of BF:

$$\Diamond\exists x\varphi(x) \to \exists x\Diamond\varphi(x).$$

The system T + S4 + BF has the following important metamathematical consequence (we learned an inessentially weaker form of it in [14]): Denote by $\vdash_1$ the derivability in the logic T + S4 + BF and by $\vdash_2$ the derivability from $\{\Box\varphi \mid \varphi$ is an axiom of T + S4 + BF$\}$ in the logic T + S4 + BF without the use of N-rule. Then we have:

Lemma. $\varphi_1, \ldots, \varphi_k \vdash_1 \psi$ *iff* $\Box\varphi_1, \ldots, \Box\varphi_k \vdash_2 \psi$.

Proof. Since (by T) $\Box\varphi_i \vdash_2 \varphi_i$ $(i = 1, \ldots, k)$ it is sufficient to show that the set of $\vdash_2$-consequences of the $\Box\varphi_i$'s is closed under $\Box$, i.e. if $\Box\varphi_1, \ldots, \Box\varphi_k \vdash_2 \alpha$, then $\Box\varphi_1, \ldots, \Box\varphi_k \vdash_2 \Box\alpha$ too. This is done by induction on the lenght of the proof of α.

3.6. Let us finish this Chapter with a short remark. The character of the problems connected with reasonable extensions of the underlying logic of MST is also (and, maybe, mainly) philosophical. The paragraph 1.1 is our only mentioned criterion for these questions. The present short experience with MST indicates that S4 + BF is the "right" extension of MST. A number of important results of the next Chapters are proved under these assumptions. A successful attack on these problems would probably require developing a metatheory for MST, especially some model theory. Because, for the time being, we have only partial results concerning the consistency of MST, there is much that has to be done.

§ 4. $\Box$-decidable formulas and decidable sets

In this Chapter we begin to develop MST.

4.1. Metadefinition. We call a formula φ $\Box$-*decidable* iff $\Box\varphi \vee \Box\neg\varphi$ holds.

We will frequently use other equivalent conditions (cf. 1.4):

(i) $\neg\Box\varphi \to \Box\neg\varphi$,

(ii) $\Diamond\varphi \to \Box\varphi$,

(iii) $(\varphi \to \Box\varphi) \& (\neg\varphi \to \Box\neg\varphi)$.

Note that if φ is decidable (i.e. $\vdash\varphi$ or $\vdash\neg\varphi$) then it is also $\Box$-decidable but the converse is not generally true.

4.2. We abbreviate the formula $\Box\varphi \vee \Box\neg\varphi$ by $\mathfrak{D}(\varphi)$. We begin with a simple result.

Lemma. $\vdash(\mathfrak{D}(\varphi) \,\&\, \mathfrak{D}(\psi)) \rightarrow [\mathfrak{D}(\neg\varphi) \,\&\, \mathfrak{D}(\varphi \,\&\, \psi)]$ *and* (i)

$$\mathrm{BF} \vdash \mathfrak{D}(\varphi) \rightarrow [\mathfrak{D}(\forall x\varphi) \,\&\, \mathfrak{D}(\exists x\varphi)]$$ (ii)

The axiom (LP) implies the following

4.3. Corollary. *Let φ be a formula built up without use of $\Box, \forall, \exists, \in$. Then* $\vdash\mathfrak{D}(\varphi)$

4.4. Definition. We call a set y *decidable* ($D(y)$ in symbols) iff

$$\forall t(\Box t \in y \vee \Box t \notin y)$$

holds.

The rest of this Chapter is devoted to proving that the domain of decidable sets is rich and behaves reasonably.

4.5. Theorem. $\vdash\mathfrak{D}(\varphi) \rightarrow \exists y\, \forall t((t \in y \equiv \varphi(t)) \,\&\, D(y))$.

Proof. We have

$$\vdash\mathfrak{D}(\varphi) \rightarrow \forall t(\varphi(t) \equiv \Box\varphi(t)) \,\&\, (\neg\varphi(t) \equiv \Box\neg\varphi(t))$$

and using (MCA) we are done.

4.6. Corollary.

(i) $\vdash\exists y\, \forall t(t \notin y)$,

(ii) $\vdash\exists y\, \forall t(t \in y)$,

(iii) $\vdash\exists y\, \forall t(t \in y \equiv (t = a_1 \vee \ldots \vee t = a_k))$,

(iv) $\vdash\exists y\, \forall t(t \in y \equiv (t \neq a_1 \,\&\, \ldots \,\&\, t \neq a_k))$.

Moreover $D(y)$ in each case.

Proof. Consider the provably $\Box$-decidable (see 4.3) formulas $t \neq t$, $t = t$, $t = a_1 \vee \ldots \vee t = a_k$ and $t \neq a_1 \,\&\, \ldots \,\&\, t \neq a_k$.

4.7. Theorem.

(i) $\vdash(D(a) \,\&\, D(b)) \rightarrow \exists y\, \forall t((t \in y \equiv (t \in a \,\&\, t \in b)) \,\&\, D(y))$,

(ii) $\vdash(D(a) \,\&\, D(b)) \rightarrow \exists y\, \forall t((t \in y \equiv (t \in a \vee t \in b)) \,\&\, D(y))$,

(iii) $\vdash D(a) \rightarrow \exists y\, \forall t((t \in y \equiv t \notin a) \,\&\, D(y))$,

(iv) $\vdash D(a) \rightarrow \exists y\, \forall t((t \in y \equiv (t \in a \vee t = c)) \,\&\, D(y))$.

Proof. Observe that decidability of a implies the $\Box$-decidability of $t \in a$. Using 4.5 we are done.

Note that, already now, we can conclude that MST interprets some nontrivial mathematics: From 4.6 and 4.7 it follows that MST interprets a weak fragment of set theory based on axioms

(i) $\exists y\, \forall t(t \notin y)$,

(ii) $\exists y\, \forall t(t \in y \equiv (t \in x \vee t = z))$.

It is sufficient to consider the domain of decidable sets. On the other side it is known (see [12]) that this theory globally interprets arithmetic with bounded induction $\mathrm{I}\Delta_0$.

4.8. Theorem.

(i) $\quad BF \vdash [D(a) \& \forall b(b \in a \to D(b))] \to \exists c\, \forall t(t \in c \equiv \exists b(b \in a \& t \in b))$,

(ii) $\quad BF \vdash D(a) \to \exists c\, \forall t(D(t) \to (t \in c \equiv t \subseteq a))$,

where $t \subseteq a$ *abbreviates* $\forall s(s \in t \to s \in a)$,

(iii) $\quad BF \vdash (\mathfrak{D}(\varphi(x, t)) \& D(a)) \to \exists c\, \forall t(t \in c \equiv \exists x(x \in a \& \varphi(x, t)))$.

Moreover, c is decidable in (i) *and* (iii).

Proof. (i) Write ψ for $D(a) \& \forall b(b \in a \to D(b))$ and $\varphi(t)$ for $\exists b(b \in a \& t \in b)$. Observe

$$\vdash \psi \to \varphi(t) \equiv \Box\varphi(t),$$

$$\vdash \psi \to (\neg\varphi(t) \equiv \forall b(D(b) \to (b \notin a \vee t \notin b)))$$

and

$$\vdash D(a) \& D(b) \to ((b \notin a \vee t \notin b) \equiv (\Box b \notin a \vee \Box t \notin b))$$

Thus

$$\vdash \psi \to (\neg\varphi(t) \equiv \forall b(b \in a \to (\Box b \notin a \vee \Box t \notin b)))$$

and so

$$\vdash \psi \to \neg\varphi(t) \equiv \forall b(\Box b \notin a \vee \Box t \notin b)$$

This gives $BF \vdash \psi \to \neg\varphi(t) \equiv \Box\neg\varphi(t)$ and by 4.5 we are done.

(ii) It is enough to calculate: $BF \vdash (D(x) \& D(a)) \to \mathfrak{D}(x \subseteq a)$ by which we are done.

(iii) Again it is sufficient to prove that the formula $\exists x(x \in a \& \varphi(x, t))$ is $\Box$-decidable. But this is immediate from the assumptions.

4.9. We write $Suc(x, y)$ as an abbreviation for $\forall t(t \in y \equiv (t \in x \vee t = x))$. Concerning axioms of infinity we observe, that the universal set v (which exists) satisfies $(x \in v \& Suc(x, y)) \to y \in v$. However, more plausible versions of axiom of infinity will be proved in the next Chapters.

In this Chapter we proved the existence of union ($\bigcup a$), intersection ($a \cap b$), etc. on some appropriate domain. Thus it might be easier to write e.g. $\varphi(x \cup \{x\})$ instead of $\exists y(Suc(x, y) \& \varphi(y))$, i.e. to introduce definable terms. Unfortunately, in view of result 3.2, this is not generally possible and one must be very carefull when using terms in arguments.

This also shows that extensionality loses some of its importance in MST.

§ 5. Extensionality

MST implies some restrictions to introducing definable terms (see 3.2 and 4.9). Hence extensionality loses one of its usual application.

Many concepts and results using extensionality can be interpreted without it (see also [3] and [4]).

We did not adopt it and, in fact, MST disproves it.

5.1. Theorem. $\vdash \exists x, y(\forall t(t \in x \equiv t \in y) \& x \neq y)$

Proof. Define the formula $\varphi(t)$ with parameters a, b, c as follows:

$$\varphi(t) \equiv (t = a \,\&\, c \in c) \vee (t = b \,\&\, c \notin c).$$

Clearly $\vdash a \neq b \to (\varphi(a) \equiv c \in c) \,\&\, (\varphi(b) \equiv c \notin c)$, so by N-rule and (LP)

(A) $\qquad \vdash a \neq b \to (\Box\varphi(a) \equiv \Box c \in c) \,\&\, (\Box\varphi(b) \equiv \Box c \notin c)$.

Also $\vdash \varphi(t) \to (t = a \vee t = b)$, and again by N-rule and (LP)

(B) $\qquad \vdash \Diamond\varphi(t) \to (t = a \vee t = b)$.

Now, by (MCA)

$$\vdash \exists y\, \forall t((\Box\varphi(t) \equiv \Box t \in y) \,\&\, (\Box\neg\varphi(t) \equiv \Box t \notin y) \,\&\, (t \in y \to \Diamond\varphi(t)))$$

(the last conjunct follows from the second one). By (B) we have

$$\vdash \exists y\, \forall t((\Box\varphi(t) \equiv \Box t \in y) \,\&\, (\Box\neg\varphi(t) \equiv \Box t \notin y) \,\&\, (t \in y \to (t = a \vee t = b)).$$

Thus "this y" is an at most two-element set and so, by 4.6 (iii) and extensionality, it is decidable. Hence we have

$$\text{ext} \vdash \exists y\, \forall t((\Box\varphi(t) \equiv t \in y) \,\&\, (\Box\neg\varphi(t) \equiv t \notin y),$$

from which follows

$$\text{ext} \vdash \Box\varphi(t) \vee \Box\neg\varphi(t).$$

This, together with (A), gives

$$\text{ext} \vdash a \neq b \to (\Box c \in c \vee \Box c \notin c).$$

Since $\vdash \exists a, b(a \neq b)$ (e.g. the empty set and the universal set are different), we have

$$\text{ext} \vdash \Box c \in c \vee \Box c \notin c.$$

This contradicts 2.1.

Let us remark that in this proof we applied (MCA) only to nonmodal formulas, i.e. those without $\Box$ (see § 8).

5.2. Various other modifications of extensionality may be considered. For example:

(i) $\qquad \forall t(\Box t \in x \equiv \Box t \in y) \to x = y$,

(ii) $\qquad \forall t((\Box t \in x \equiv \Box t \in y) \,\&\, (\Box t \notin x \equiv \Box t \notin y)) \to x = y$,

(iii) $\qquad \Box\forall t(t \in x \equiv t \in y) \to x = y$

(this formulation was suggested by P. Pudlák),

(iv) $\qquad \forall t(t \in x \equiv t \in y) \,\&\, D(x) \,\&\, D(y) \to x = y$.

It is simple to observe that (i) $\to$ (ii) $\to$ (iii), and that BF implies (iii) $\to$ (iv).

In § 8 we sketch a proof of partial consistency result concerning a slight variant of (ii), (iii) and (iv).

Concerning (i) we have the following result:

Lemma. MST *disproves* (i).

Proof. By (MCA)

$$\vdash \exists y\, \forall t((\Box t \in y \equiv \Box r \in r) \,\&\, (\Box t \notin y \equiv \Box r \notin r)).$$

If we choose for the parameter r RUSSELL's set from § 2, we obtain

(A) $\qquad \vdash \exists y \, \forall t(\neg \Box t \in y \;\&\; \neg \Box t \notin y).$

Hence if we choose for x and y from (i) a decidable empty set (see 4.6 (i)) and "the y from (A)", respectively, we conclude

$$\text{(i)} \vdash \exists y \, \forall t(\Diamond t \in y \;\&\; \Diamond t \notin y \;\&\; D(y)).$$

This is a contradiction.

Although some of the above modifications of extensionality are interesting (and useful) we shall not assume extensionality at all.

§ 6. Small sets

Imagine that we are looking on the set-universum from outside of it. The result 4.6 (iii) implies that for all finite collections of sets from the set-universe (i.e. finite from outside of the set-universe) there exists a decidable set in the set-universe with the same extension (i.e. with the same members). Hence the following should characterize "finite sets":

$$(\forall y \subseteq x) \, D(y).$$

Unfortunately, the proof of 5.1 implies that there cannot be a set x satisfying the formula above and having more than one element.

However, this suggests to characterize finite sets as those whose all subcollections (e.g. those represented by some formula) are "simple" in the sense of $\Box$. Since (MCA) is our principle which enables us to represent (in a particular way) a collection of sets defined by some formula by a set, we are lead to the following definition which should substitute a notion of finiteness:

6.1. Definition. x is *small* ($S(x)$ in symbols) iff

(i) $\qquad \forall y \, \exists z(D(z) \;\&\; \forall t(t \in z \equiv (t \in x \;\&\; \Box t \in y)))$

and

(ii) $\qquad D(x)$

We require small sets to be decidable since decidable sets are easily handled. We do not consider this definition as a definitive approximation of "finiteness" in MST; the notion of "smallness" is introduced mostly for the purposes of § 7.

6.2. Lemma. *For any formula $\varphi(t)$ we have*

$$\vdash S(a) \to \exists b \, \forall t(D(b) \;\&\; (t \in b \equiv (\Box \varphi(t) \;\&\; t \in a))).$$

Proof. By (MCA), $\vdash \exists c \, \forall t(\Box \varphi(t) \equiv \Box t \in c)$ and by smallness of a we are done.

6.3. Lemma.

(i) $\qquad \vdash b \subseteq a \;\&\; S(a) \;\&\; D(b) \to S(b),$

(ii) $\qquad \vdash S(a) \;\&\; S(b) \to \exists c(S(c) \;\&\; \forall t(t \in c \equiv (t \in a \vee t \in b))),$

(iii) $\qquad \vdash S(a) \;\&\; S(b) \to \exists c(S(c) \;\&\; \forall t(t \in c \equiv (t \in a \;\&\; t \in b))),$

(iv) $\qquad \vdash S(a) \to \exists c(S(c) \;\&\; \forall t(t \in c \equiv (t \in a \vee t = d))).$

Proof. Easily by 4.7.

§ 7. Arithmetic

In this Chapter we develop PEANO's arithmetic PA in MST augmented by S4 and BF.

7.1. We begin with the familiar von NEUMANN's definition: x is an *ordinal number* ($On(x)$ in symbols) iff

(i) x is transitive

and

(ii) x is strictly well-ordered by $\in$.

For ordinal numbers we will also use the letters $\alpha, \beta, \ldots$

7.2. "Natural" candidates for natural numbers are perhaps small ordinal numbers. Since it is not evident why there could not be a limit small ordinal number we shall use this definition: α is a *natural number* ($N(\alpha)$ in symbols) iff

(i) $On(\alpha)$,

(ii) $S(\alpha)$,

(iii) $\forall\beta(\beta \in \alpha \to S(\beta))$,

(iv) α is strictly well-ordered by $\in^{-1}$.

($x \in^{-1} y$ is obviously defined by $y \in x$).

7.3. Lemma.

(i) *Any decidable empty set is a natural number.*

(ii) *For any natural number α there exists a natural number β which is its successor* (see 4.9).

(iii) *Any element of a natural number is a natural number.*

(iv) *If $\alpha \in \beta$ are natural numbers, then there exists a natural number γ which is a successor of α and satisfies $\gamma \in \beta$ or $\gamma = \beta$.*

Precise formulations are obvious. Remark that without extensionality e.g. successors are not uniquely given.

Proof. (i) is trivial, (ii) use 6.3 (iv), (iii) use (iii) of 7.2. For (iv) choose γ to be the minimal element of "$\beta \setminus (\alpha \cup \{\alpha\})$" or the β itself.

7.4. Metadefinition. A formula $\varphi(t)$ is called a *cut* in N iff

(i) $\varphi(t) \to N(t)$,

(ii) "α decidable empty" $\to \varphi(\alpha)$,

(iii) $\varphi(\alpha)$ & $Suc(\alpha, \beta)$ & $N(\beta) \to \varphi(\beta)$.

A cut $\varphi(t)$ is called *nontrivial* if, moreover, the following holds:

(iv) $\exists\beta(N(\beta)$ & $\neg\varphi(\beta))$.

Remark that any cut $\varphi(t)$ in N satisfies (for β a natural number):

(v) $(\forall\alpha \in \beta)\, \varphi(\alpha) \to \varphi(\beta)$

(any nonempty β has an maximal element whose successor is β itself).

7.5. Lemma. *No formula $\Box\varphi(t)$ is a nontrivial cut in N.*

Proof. We shall argue informally. Suppose that $\Box\varphi(t)$ is a nontrivial cut in N. So there exists a natural number α which does not satisfy $\Box\varphi$. Let us fix any such a number α. Since empty natural numbers satisfy $\Box\varphi$, α is not empty. By 6.2 there exists $b \subseteq \alpha$, such that $\forall t(t \in b \equiv (t \in \alpha \ \& \ \Box\varphi(t)))$. This b is not empty (since $\min \alpha$ satisfies $\Box\varphi$) and hence it has a maximal element, say, β. By 7.3 (iv) there are two possibilities: α is a successor of β or there exists $\gamma \in \alpha$ which is a successor of β. In both cases we immediately come into contradiction with the choice of α and β.

7.6. Lemma. S4 + BF $\vdash \Box N(t) \equiv N(t)$.

Proof. In a view of 7.5 it is enough to prove that $\Box N(t)$ is a cut in N. Clearly $\vdash \Box N(t) \to N(t)$. Also

$$\vdash \forall x(\text{“}x \text{ decidable empty”} \to N(x)),$$

hence, by N-rule,

$$\vdash \forall x(\Box \text{ “}x \text{ decidable empty”} \to \Box N(x)).$$

Since S4 + BF $\vdash$ “x decidable empty” $\to \Box$ “x decidable empty”, we have

$$\text{S4} + \text{BF} \vdash \forall x(\text{“}x \text{ decidable empty”} \to \Box N(x)).$$

It remains to prove

$$\Box N(\alpha) \ \& \ Suc(\alpha, \beta) \ \& \ N(\beta) \to \Box N(\beta).$$

Clearly $\vdash N(\alpha) \ \& \ Suc(\alpha, \beta) \ \& \ D(\beta) \to N(\beta)$, hence

$$\vdash \Box N(\alpha) \ \& \ \Box Suc(\alpha, \beta) \ \& \ \Box D(\beta) \to \Box N(\beta).$$

Since $\vdash N(\alpha) \to D(\alpha)$ and

$$\text{S4} + \text{BF} \vdash D(\alpha) \ \& \ D(\beta) \to (\Box D(\beta) \ \& \ (Suc(\alpha, \beta) \to \Box Suc(\alpha, \beta))),$$

we may conclude

$$\text{S4} + \text{BF} \vdash \Box N(\alpha) \ \& \ Suc(\alpha, \beta) \ \& \ N(\beta) \to \Box N(\beta).$$

7.7. Theorem (Induction). S4 + BF *implies that there are no nontrivial cuts in N.*

Proof. Suppose that $\varphi(t)$ is a cut in N. Define the formula $\psi(t)$ by

$$\psi(t) \equiv N(t) \ \& \ (\forall s \in t)\, (((\forall r \in s)\, \varphi(r)) \to \varphi(s)) \to (\forall s \in t)\, \varphi(s).$$

$\psi(t)$ says that induction holds up to t.

Now we claim:

$$\text{S4} + \text{BF} \vdash \text{“}\Box\psi(t) \text{ is a cut in } N\text{”}.$$

For this we need:

(i) $\vdash \Box\psi(t) \to N(t)$,

(ii) S4 + BF $\vdash$ “α decidable empty” $\to \Box\psi(\alpha)$,

(iii) S4 + BF $\vdash \Box\psi(\alpha) \ \& \ Suc(\alpha, \beta) \ \& \ N(\beta) \to \Box\psi(\beta)$.

(i) is trivial, (ii) is easy by 7.6 and (iii) is proved analogically to the proof of the corresponding part in 7.6. From the claim, using 7.5, it follows

$$\text{S4} + \text{BF} \vdash \psi(t) \equiv N(t).$$

By 7.4 (v) we have

$$\vdash \text{“}\varphi \text{ is a cut in } N\text{”} \to (\psi(t) \to (\forall s \in t)\, \varphi(s)),$$

hence we obtained

$$\text{S4} + \text{BF} \vdash \text{“}\varphi \text{ is a cut in } N\text{”} \to \forall t \varphi(t).$$

7.8. Now we sketch how to introduce the arithmetical structure on N. At first we need to define $=_N$ and $\leqq_N$. We define, for example, "$\alpha \leqq_N \beta$" iff "there exists a small $1-1$ mapping of α into β" and "$\alpha =_N \beta$" iff "$(\alpha \leqq_N \beta \,\&\, \beta \leqq_N \alpha)$" ("mapping" will be defined obviously). Now, using induction, one proves the usual properties of $=_N$ and $\leqq_N$. Having this we define in some reasonable way the addition. For example: "$\alpha + \beta = \gamma$" iff "there exist small sequence $s_0, \ldots, s_\beta$ such that $s_0 = \alpha$, $s_{\eta+1} = s_\eta + 1$ ($\eta + 1$ abbreviates the successor of η) and $s_\beta = \gamma$" ("sequence" will be defined as usual). So we easilly prove $\alpha + 0 = \alpha$ and $(\alpha + \beta) + 1 = \alpha + (\beta + 1)$. Then, using induction, we prove that "+" is defined for each two natural numbers. The same can be done with multiplication and induction will guarantee the wanted arithmetical properties of "+" and "·". Observe that induction also implies that this can be done uniquely. Moreover, because 7.7 holds for all (not only "arithmetical") formulas, the domain of natural numbers forms a standard model (i.e. without cuts) of PA in MST. On the other side there remains the problem whether N defines a set.

Finally remark that by using the modification 5.2 (iv) of extensionality we may considerably simplify a structure of N; in particular, $x =_N y$ and $x \leqq_N y$ are $x = y$ and $(x = y \vee x \in y)$, respectively.

§ 8. Partial consistency result

In this Chapter we will construct (in ZF) a Kripke-style model for considerable fragment of MST.

8.1. Let M be a fixed infinite set.

Definition. A *world* over M (shortly *world*) is an ordered pair $w = \langle w^+, w^- \rangle$ such that:

$$\text{(i) } w^+, w^- \subseteqq M^2 \quad \text{and} \quad \text{(ii) } w^+ \cap w^- = \emptyset.$$

The set of worlds is denoted by W. The set W is partially ordered by

$$w_1 \leqq w_2 \quad \text{iff} \quad w_1^+ \subseteqq w_2^+ \,\&\, w_1^- \subseteqq w_2^-.$$

8.2. For each world $w \in W$ we define a Kripke structure $K_w = \langle |K_w|, \leqq_w, \models_w \rangle$ by

(i) $|K_w| = \{w' \in W \mid w \leqq w'\}$,

(ii) $\leqq_w = |K_w|^2$,

(iii) for $w_1 \in |K_w|$ and $a, b \in M$:

a) $w_1 \models_w a \in b$ iff $\langle a, b \rangle \in w_1^+$,

b) $w_1 \models_w a = b$ iff $a = b$,

c) $\models_w$ is in obviously way extended to all sentences.

For a formula φ with some free variables we define: $w_1 \models_w \varphi$ iff $w_1 \models_w \bar{\varphi}$, where $\bar{\varphi}$ is the universal closure of φ.

8.3. Let $\mathscr{F}_1$ be the set of all formulas of the modal set-theoretical language with just one free variable t and with parameters from M. Let $^\wedge\colon \mathscr{F}_1 \to M$ be a fixed mapping for which $\varphi^\wedge = \psi^\wedge$ iff $\varphi = \underbrace{\neg \ldots \neg}_{2n \text{ times}} \psi$ or $\psi = \underbrace{\neg \ldots \neg}_{2n \text{ times}} \varphi$.

8.4. Now we define a sequence of worlds $\{w_\alpha\}$ (α an ordinal) by

(i) $w_0 = \langle \emptyset, \emptyset \rangle$,

(ii) $w_\alpha = \langle \bigcup_{\beta<\alpha} w_\beta^+, \bigcup_{\beta<\alpha} w_\beta^- \rangle$ (if α a limit),

(iii) $w_{\alpha+1}^+ = w_\alpha^+ \cup \{\langle a, \varphi(t)^\wedge \rangle \mid \langle a, \varphi(t)^\wedge \rangle \notin w_\alpha^- \mathbin{\&} w_\alpha \models_{w_\alpha} \Box\varphi(a)\}$,

$w_{\alpha+1}^- = w_\alpha^- \cup \{\langle a, \varphi(t)^\wedge \rangle \mid \langle a, \varphi(t)^\wedge \rangle \notin w_\alpha^+ \mathbin{\&} w_\alpha \models_{w_\alpha} \Box\neg\varphi(a)\}$.

It is simple to realize that this definition is correct, i.e. it defines a sequence of worlds. Also is immediate that the sequence $\{w_\alpha\}$ is monotonous in W ($\alpha \leqq \beta \to w_\alpha \leqq w_\beta$). Thus, surely, there exists ξ such that $w_\xi = w_{\xi+1}$ (in fact $(\forall\eta \geqq \xi)\ w_\eta = w_\xi$). Let us fix any such ξ.

8.5. Definition. We call a formula φ *monotonic* iff

$$(\forall w_1, w_2 \in W)\,(w_1 \leqq w_2 \to (w_1 \models_{w_1} \Box\varphi \to w_2 \models_{w_2} \Box\varphi)).$$

Let $\mathrm{m}\mathscr{F}$ be the set of monotonic formulas, $\mathrm{m}\mathscr{F}_1 = \mathrm{m}\mathscr{F} \cap \mathscr{F}_1$.

Lemma. *Let* $\varphi(t) \in \mathrm{m}\mathscr{F}_1$, $c \in M$ *and* $w \in |K_{w_\xi}|$. *Then*

(i) $w \models_{w_\xi} \Box c \in \varphi^\wedge \to \Box\varphi(c)$,

(ii) $w \models_{w_\xi} \Box c \notin (\neg\varphi)^\wedge \to \Box\varphi(c)$,

(iii) $w \models_{w_\xi} \Box\neg\varphi(c) \to \Box c \in (\neg\varphi)^\wedge$,

(iv) $w \models_{w_\xi} \Box\neg\varphi(c) \to \Box c \notin \varphi^\wedge$.

Proof. We will write "$\Leftrightarrow$" instead of "iff" and "$\Rightarrow$" instead of "if ... then ...".

(i) $w \models_{w_\xi} \Box c \in \varphi^\wedge \Leftrightarrow w_\xi \models_{w_\xi} \Box c \in \varphi^\wedge \Leftrightarrow \langle c, \varphi^\wedge \rangle \in w_\xi^+ \Rightarrow (\exists\alpha \leqq \xi)\ w_\alpha \models_{w_\alpha} \Box\varphi(c)$. By monotonicity of φ, $w_\xi \models_{w_\xi} \Box\varphi(c)$.

(ii) $w \models_{w_\xi} \Box c \notin (\neg\varphi)^\wedge \Leftrightarrow w_\xi \models_{w_\xi} \Box c \notin (\neg\varphi)^\wedge \Leftrightarrow \langle c, (\neg\varphi)^\wedge \rangle \in w_\xi^- \Rightarrow (\exists\alpha \leqq \xi)\ w_\alpha \models_{w_\alpha} \Box\varphi(c)$. By monotonicity of φ w_ξ, $\models_{w_\xi} \Box\varphi(c)$.

(iii) $w \models_{w_\xi} \Box\neg\varphi(c) \Leftrightarrow w_\xi \models_{w_\xi} \Box\neg\varphi(c)$. Observe, that monotonicity of φ implies $\neg(\exists\alpha \leqq \xi)\ w_\alpha \models_{w_\alpha} \Box\varphi(c)$. Thus $\langle c, (\neg\varphi)^\wedge \rangle \notin w_\xi^-$ and, by definition, $\langle c, (\neg\varphi)^\wedge \rangle \in w_{\xi+1}^+ = w_\xi^+$. So $w_\xi \models_{w_\xi} \Box c \in (\neg\varphi)^\wedge$.

(iv) Again (see (iii)) $\neg(\exists\alpha \leqq \xi)\ w_\alpha \models_{w_\alpha} \Box\varphi(c)$, hence $\langle c, \varphi^\wedge \rangle \notin w_\xi^+$ and, by definition, $\langle c, \varphi^\wedge \rangle \in w_{\xi+1}^- = w_\xi^-$. So $w_\xi \models_{w_\xi} \Box c \notin \varphi^\wedge$.

8.6. Lemma. $\mathrm{m}\mathscr{F}$ *contains all nonmodal formulas.*

Proof. Easily from definition of satisfaction.

8.7. We define a theory U with following axioms:

(i) the logical axioms T + S4 + S5 + BF (and N-rule too),

(ii) (LP),

(iii) $\exists y\,\Box\forall t((\Box t \in y \equiv \Box\varphi(t)) \mathbin{\&} (\Box t \notin y \equiv \Box\neg\varphi(t)))$, for nonmodal φ.

Our main consistency result is

Theorem. U *is consistent.*

Proof. We will prove that K_{w_ξ} is a model for U, i.e. for each $w \in |K_{w_\xi}|$ $w \models_{w_\xi} \mathrm{U}$. This is nontrivial only for the axioms of (iii). By 8.5 (using 8.6) we have

$$w \models_{w_\xi} \forall t((\Box t \in \varphi^\wedge \to \Box\varphi(t)) \;\&\; (\Box\neg\neg\varphi(t) \to \Box t \in (\neg\neg\varphi)^\wedge))$$

and

$$w \models_{w_\xi} \forall t((\Box t \notin (\neg\neg\varphi)^\wedge \to \Box\neg\varphi(t)) \;\&\; (\Box\neg\varphi(t) \to \Box t \notin \varphi^\wedge)).$$

Thus, since $\varphi^\wedge = (\neg\neg\varphi)^\wedge$,

$$w_\xi \models_{w_\xi} \Box\forall t((\Box t \in \varphi^\wedge \equiv \Box\varphi(t)) \;\&\; (\Box t \notin \varphi^\wedge \equiv \Box\neg\varphi(t)))$$

and, using S4, we obtain

$$w \models_{w_\xi} \exists y \Box\, \forall t((\Box t \in y \equiv \Box\varphi(t)) \;\&\; (\Box t \notin y \equiv \Box\neg\varphi(t))).$$

8.8. Corollary. MST *with instances of* (MCA) *only for nonmodal formulas is consistent with* S4 + BF + S5.

8.9. Let us sketch how to modify the construction above to obtain a model for the theory from 8.8 together with some modification of extensionality. The variant of 5.2 (ii) we mentioned is the following:

(ii′) $\forall t((\Box t \in x \equiv \Box t \in y) \;\&\; (\Box t \notin x \equiv \Box t \notin y)) \;\&\; \exists t(\Box t \in x \vee \Box t \notin x) \to x = y.$

Observe that BF implies: (ii′) → (iv) ((iv) refers to 5.2).

The suitable modification of the construction to obtain also a model of (ii′) is the following: Let us call a pair of sets $\langle\{t \mid \langle t, a\rangle \in w^+\}, \{t \mid \langle t, a\rangle \in w^-\}\rangle$ a type realized by a in w. Instead of "completing" the types of $\varphi^\wedge$'s in any step of the construction consider the set X of types $\langle\{t \mid w \models_{K_w} \Box\varphi(t)\}, \{t \mid w \models_{K_w} \Box\neg\varphi(t)\}\rangle$, $\varphi \in \mathscr{F}_1$, and construct the next world w' in such a way, that each nonempty (i.e. different from $\langle\emptyset, \emptyset\rangle$) type from X is realized in w' by exactly one element. This implies that (ii′) is true in all K_{w_α}'s. Precise formulation of this construction needs some more definitions and technical results and we omit it.

Let us finish this Chapter with remark that a construction similar to ours (using fixed-point of some sequence of structures) was used already by FEFERMAN.

§ 9. Related systems

In the literature we found a few systems closely connected to MST. We shall discuss them in this Chapter.

9.1. The systems which we found are those of FITCH [5], [6], of GILMORE [8] and of FEFERMAN [1], [2], [3]. Since the systems of FITCH and of GILMORE are covered by those of FEFERMAN (see [2], p. 78 and p. 86) we shall be concerned only with the latter one.

9.2. FEFERMAN's systems differ in formalization in all three references and it is not easy to compare them. We think that the system of [3] reflects sufficiently the ideas of both [1] and [2]. Also its formalization is nearest to that of MST. Thus we shall describe FEFERMAN's ideas and compare them with MST using [3].

9.3. Let us begin with citing FEFERMAN ([3], pp. 88–90): "(i) There is nothing currently on the horizon which gives hope of obtaining anything like the Frege-Russell (pre-PM) or Church-Curry programs for a foundation of mathematics in a type-free theory. These programs are first of all monistic ("everything is a class" or "everything is a function") and secondly attempt to extract all of mathematics from some few relatively simple principles of a very general "logical" character.

(ii) Instead these programs are more profitably reversed. We should start with a part S of mathematics that is already accepted, or at least reasonably well-understood – be it number theory, or analysis, or even stronger theories. S should then be extended to a theory $\hat{S}$ which admits instances of self-application not avaible in S. This should be done as a matter of convenience, thus the extension should be conservative."

FEFERMAN is lead to selfapplicative notions also by his goal to find foundations for category theory which would be unrestricted in the sense, that large categories, categories of categories, etc. would be naturally included.

9.4. FEFERMAN's idea: Let S be any consistent theory in any logic including classical. The language $\mathscr{L}$ of S includes the operations 0 (zero), $'$ (successor), $\langle , \rangle$ (pairing), such that $x' \neq 0$, $x' = y' \to x = y$, $\langle u, v\rangle = \langle a, b\rangle \to u = a \;\&\; v = b$ are theorems of S. The first step is to extend $\mathscr{L}$ by adding a ternary predicate $App(x, y, z)$ with the intended meaning "the operation x is defined at y with value z". This is also expressed by $xy \cong z$ and App is then denoted by $\cong$. The extended language is denoted by $\mathscr{L}(\cong)$. The theory S is extended to $S_{\cong}$ in $\mathscr{L}(\cong)$ by adding some appropriate axioms concerning $\cong$. We shall omit them here because they are irrelevant for our discussion. The next step is to extend $\mathscr{L}(\cong)$ by a pair of new binary predicates $\eta, \bar{\eta}$ and by a new unary logical operator $\Box$. The extended language is denoted by $\mathscr{L}(\eta, \bar{\eta}, \cong)$. The intended meaning of these new symbols is "for any property φ there exists y such that $t\eta y$ iff it can be verified (in some sense) that $\varphi(t)$, and $t\bar{\eta}y$ iff it can be verified that $\neg\varphi(t)$, and $\Box\varphi$ iff it can be verified that φ". The logic is extended over that of S by the modal system T + S4 + BF (+ N-rule) and by the axiom schemes

(i) $\varphi \to \Box\varphi$, φ atomic,

(ii) $\neg\varphi \to \Box\neg\varphi$, φ atomic of $\mathscr{L}(\cong)$,

(iii) $t\bar{\eta}y \equiv \Box\neg t\eta y$.

For any theory Z we write $Z \vdash^* \varphi$ iff for some $\psi_1, \ldots, \psi_k \in Z$, $(\psi_1 \;\&\; \ldots \;\&\; \psi_k) \to \varphi$ is derivable in the described logic.

The final extension $\hat{S}$ of S has the logic described above (with derivability $\vdash^*$) and the axioms of $S_{\cong}$ together with

(D) $\neg(t\eta y \;\&\; t\bar{\eta}y)$ (disjointness)

and

(C) For each formula φ of $\mathscr{L}(\eta, \bar{\eta}, \cong)$ with free variables among t, a_i $(i \leqq n)$:

$$\forall a_1, \ldots, a_n \exists y \forall t((t\eta y \equiv \Box\varphi(t)) \;\&\; (t\bar{\eta}y \equiv \Box\neg\varphi(t))) \text{ (comprehension)}.$$

Observe, that N-rule is not generally applicable in $\hat{S}$, i.e., $\hat{S} \vdash^* \psi$ does not imply $\hat{S} \vdash^* \Box\psi$.

We have to remark that the original version of axiom (C) is stronger than this one, because it also contains a "functional representation" of y in terms of $\cong$ and a_i's. We used this version because we are interested only in the $(\Box, \eta, \bar{\eta})$-part of the original $\hat{S}$. However, this is rather incorrect, since it is not clear why the original $\hat{S}$ of [3] should be conservative (with respect to sentences without $\cong$) over our version of $\hat{S}$ (if we understand [3] correctly, then it is not).

9.5. The assumptions about S (which are needed for a "functional part" of $\hat{S}$), the different languages and the different logics make it impossible to compare $\hat{S}$ with MST directly. We propose to compare the following subtheory S^* of $\hat{S}$: The language $\mathscr{L}^*$ of S^* consists of the predicates $\eta, \bar{\eta}$ and the logical operator $\Box$. The logical assumptions of S^* are those of modal predicate calculus + T + S4 + BF (+ N-rule) together with (LP) and:

(A1) $\quad t\eta y \to \Box t\eta y,$

(A2) $\quad t\bar{\eta} y \to \Box t\bar{\eta} y,$

(A3) $\quad t\bar{\eta} y \equiv \Box \neg t\eta y.$

The axioms of S^* are (D) and (C) (restricted to $\mathscr{L}^*$). The derivability from S^* means the same as in $\hat{S}$, namely $S^* \vdash^* \varphi$ iff for some $\psi_1, \ldots, \psi_k \in S^*$ (i.e. (D) + some instances of (C)) the formula $(\psi_1 \& \ldots \& \psi_k) \to \varphi$ is derivable (in the usual sense) in the above described logic of S^* alone. We remark again that the $\vdash^*$-consequences of S^* are not closed under N-rule.

9.6. We denote by MST* the theory MST + S4 + BF. Using an idea of 3.5 we are able to construct two theories T_1 and T_2 with classical derivability concept and with the same theorems as S^* and MST*. The direct comparing of T_1, T_2 will be possible and thus, in fact, the comparing of S^*, MST* too.

Theory T_1: The language of T_1 is $\mathscr{L}^*$ and the axioms are

(i) $\quad$ T + $\Box$(T + S4 + BF),

(ii) $\quad$ $\Box$(LP),

(iii) $\quad$ $\Box$(A1), $\Box$(A2), $\Box$(A3),

(iv) $\quad$ (D), (C).

Theory T_2: The language of T_2 is the modal set-theoretical language and the axioms are

(i) $\quad$ T + $\Box$(T + S4 + BF),

(ii) $\quad$ $\Box$(LP),

(iii) $\quad$ $\Box$(MCA).

The derivability concept of both T_1 and T_2 is the same, the classical (see 3.5). We state a result analogical to 3.5 without any proof:

Lemma.

(i) $\quad S^* \vdash^* \varphi$ *iff* $T_1 \vdash_2 \varphi$, $\qquad$ (ii) $\quad MST^* \vdash \varphi$ *iff* $T_2 \vdash_2 \varphi$,

where $\vdash_2$ *denotes derivability in* T_1, T_2 (see 3.5).

9.7. Using $\Box$(A1) and $\Box$(A3) we can reformulate (C) as follows:

(C′) $\exists y\, \forall t((\Box t\eta y \equiv \Box\varphi(t))\ \&\ (\Box\neg t\eta y \equiv \Box\neg\varphi(t)))$.

Thus (C′) is identical with (MCA). This suggests two things:

a) try to interpret T_2 in T_1 by interpreting $\in$ as η,

b) try to interpret T_1 in T_2 by interpreting $x\eta y$ as $\Box x \in y$ and $x\bar{\eta}y$ as $\Box x \notin y$.

However, both fails.

9.8. a) fails because there is no chance to prove $\Box$(C′) in T_1. A question surely arises whether $\Box$(MCA) is necessary for developing T_2, or equivalently, whether the logic of S* is not sufficient for MST. The question is artificial since the logic of MST arose from the idea of modification of (CA). However, the answer is, as we shall see, no; the lack of N-rule for MST would have lead to essential difficulties.

9.9. b) looks more satisfactorily: the described interpretation is correct for whole T_1 except $\Box$(A3). This suggests the third possibility:

c) try to interpret η as $\in$ and to use $\Box$(A3) as "definition" of $\bar{\eta}$ (i.e., to interpret $x\bar{\eta}y$ as $\Box x \notin y$).

This c) leads to a correct interpretation of T_1 without $\Box$(A1) in T_2.

9.8 and 9.9 show that there is not a simple way to interpret one of T_1, T_2 in another.

9.10. We finish this Chapter by sketching how usual mathematical objects are introduced into $\hat{S}$ (for details see [3]). The natural numbers are not finite ordinals but terms $0, 0', 0'', \ldots$ A function from a to b is not some subset of the cartesian product $a \times b$, but it is a function in the sense of $\cong$. Also $a \times b$ itself is not the set of usual ordered pairs but of terms $\langle u, v\rangle$ such that $u\eta a$ and $v\eta b$. It is an open question whether natural numbers form an object N. However, this is easily shown to be consistent with $\hat{S}$. If we try to define the natural numbers as finite ordinals then $\hat{S}$ will not prove induction in the whole generality-this is caused by the lack of N-rule (see [2], p. 107). This answers the question of 9.8.

References

[1] Aczel, P., and S. Feferman, Consistency of the unrestricted abstraction principle using an intensional equivalence operator. In: To H. B. Curry, Essays on Combinatory Logic, Lambda Calculus and Formalism (J. R. Hindley and J. P. Seldin, eds.), Academic Press, New York 1980, pp. 67–98.

[2] Feferman, S., Non-extensional type-free theories of partial operations and classifications I. In: Proof Theory Symposion, Kiel 1974. Springer Lecture Notes in Mathematics **500** (1975), pp. 73–118.

[3] Feferman, S., Categorial foundations and foundations of category theory. In: Logic, Foundations of Mathematics and Computability Theory (R. E. Butts and J. Hintikka, eds.), D. Reidel Publ. Co., Dordrecht 1977, pp. 149–169.

[4] Feferman, S., Toward useful type-free theories I. J. Symbolic Logic **49** (1984), 75–111.

[5] Fitch, F. B., A consistent modal set theory (abstract). J. Symbolic Logic **31** (1966), 701.

[6] Fitch, F. B., A complete and consistent modal set theory. J. Symbolic Logic **32** (1967), 93–102.

[7] Fraenkel, A. A., Bar-Hillel, Y., and A. Levy, Foundations of Set Theory. North-Holland Publ. Comp., Amsterdam 1973.

[8] GILMORE, P. C., The consistency of partial set theory without extensionality. In: Axiomatic Set Theory, Proc. of Symp. in Pure Math. **13**, Part II (1974), pp. 147–153.
[9] HUGHES, G. E., and M. J. CRESSWELL, An Introduction to Modal Logic. Methuen, London 1968.
[10] KRAJÍČEK, J., A possible modal reformulation of coprehension scheme. Comment. Math. Univ. Carolin **24** (1983), 387–388.
[11] KRAJÍČEK, J., Modal set theory. In: Proc. 2nd Easter Conf. on Model Theory 1984, pp. 87–99.
[12] PUDLÁK, P., Some prime elements in the lattice of interpretability types. Trans. Amer. Math. Soc. **280** (1983), 255–275.
[13] RUSSELL, B., History of Western Philosophy. Unwin Paperbacks, London 1979.
[14] SMORYŃSKI, C., Self-reference and modal logic. Manuscript.

J. Krajíček
Math. Inst. Acad. Sci.
Žitná 25
11567 Praha 1
Czechoslovakia

(Eingegangen am 28. Juli 1986)