STS:
A STRUCTURAL
THEORY OF SETS

Alexandru Baltag

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the Department of Mathematics
Indiana University
June 1998

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.

Jon Barwise, Ph.D.

Michael Dunn, Ph.D.

Doctoral Committee

Lawrence S. Moss, Ph.D.

Slawomir Solecki, Ph.D.

July, 15, 1998

To the memory
of my father Cezar
and my mother Elena,

who by their words and their life
taught me to love the ways of the spirit
and to trust the reality of beautiful ideas

Acknowledgements

First I would like to thank my advisor, Jon Barwise, for his essential contribution to the development of this dissertation. His philosophical views on the foundations of mathematics and his work on non-wellfounded sets and semantical paradoxes have inspired, directed and fertilized the present work. I would also like to thank him for his extraordinary generosity, his insights and feedback, his full, uninterrupted and overwhelming support and help over the years, and most importantly his enthusiastic openness to even the "wildest" ideas that I could think of.

I would like to thank Larry Moss, my other advisor, for all his insightful help and his important contribution to the growing and the unfolding of the ideas presented here. His work on constructing a model for non-wellfounded sets using modal logic is one of the sources of my present approach to set theory. His passion for mathematics and his masterful knowledge of modern logical and set-theoretical techniques played an essential role in my logic training. Moreover, his generous friendship and his moral, practical and spiritual support have given me strength in some of the toughest moments of my life.

I want to express my thanks to Mike Dunn for agreeing to serve on my Dissertation Committee, and also for his lessons in logic and his help. Attending his class on Algebraic Logic was one of the most useful experiences of my graduate years. I also thank Slawek Solecki for kindly agreeing to serve on my Committee, and for his comments and feedback on the present work.

I would like to thank my college Professor Solomon Marcus, for his lessons, help and encouragement, both during my college years and after. My thanks to Serban Basarab and my other colleagues at the Institute of Mathematics in Bucharest for all their help. I am grateful to the people at the Institute of International Education, for giving me the Fulbright grant that made possible my participation in a graduate study in the US. I also want to thank my colleagues and professors at the Department of Mathematics of Indiana University for their support, and to secretaries Tracy Johnson and Misty Cummings for the kindness and efficiency shown in helping me with all graduate issues. I give my kind thanks to my friends Ioana Moise, Jordi Puigsegur and Nik Swoboda, for helping me to tame this monster called LATEX, and in general for being my friends-in-need during the writing of this dissertation.

I am grateful to my family for believing in me and always supporting me: to my deceased parents, my grandparents, my dear brother Adrian, his lovely wife and his extraordinary parents-in-law, who helped me and my father with such miraculous dedication. I give my love and my gratitude to Adriana and Matei Calinescu, whom I am proud to consider not just as my best friends, but as my adoptive parents; their protective presence, manifested through continuous help and advice, overwhelming kindness and generosity, stimulating intellectual friendship and uninterrupted moral support, has been the main outside force that has helped me keep my mental balance and my spiritual drive throughout the years I have spent in the US. And I would also like to thank my ex-wife Dana, for putting up with me for such a long time and for giving me all the love that she was able to give me.

Contents

Chapter 1

Introduction

It is often said that *mathematics is the science of pattern*, that its objects are purely structural in nature and have no proper identity apart from their structure. To the working mathematician who takes the time to reflect on what he is doing, this statement may seem almost self-evident. As usually understood by philosophers, however, the so-called *structuralistic conception* of mathematics is seen as more or less directly opposed to the "substantialist" approach of Cantorian Set Theory and is indeed frequently employed to deflate the later's foundational claims. This version of structuralism has a strong nominalistic, or at least anti-Platonistic flavor: it is seen as a way to "de-ontologize" the mathematical universe by destroying the illusion of the objective existence of individual mathematical objects endowed with unique features and a clear-cut identity.

While accepting the above-mentioned structuralist slogan, I have little sympathy for the latter views. More generally, the whole deflationary program seems to go against deep, old and provably useful common-sense intuitions about mathematics. My mind can easily and consistently hold, at least at an informal level (as I think most mathematicians do), two ideas which the structuralists tend to regard as incompatible:

(1) The objects of mathematics are simply *all the possible structures* (regardless of whether or not they are actually realized as completed, realistic objects) and *there is nothing to them but their structure:* their identity is given by the structure only.

(2) Nevertheless, the act of unifying such a pure structure into a completed whole (as Cantor has done), taking it as an *actual, well-defined* individual object, to which other structures could be applied, is a fully legitimate mental act, without which mathematics would be impossible. For, to quote Jon Barwise, the slogan mentioned in the above paragraph has to be corrected: all science is the study of patterns; mathematics is the science of *patterns of patterns.*

In this paper, I will not try to argue directly for the truth or the compatibility of these two ideas. I shall rather take them for granted, and moreover take them *at face value*, in order to articulate a new, coherent, more liberal conception of what "sets" are. I call this the *structural-analytical* concept of set. Building on the intuitions and the work of Abramski, Aczel, Barwise, Boffa, Devlin, Forti, Honsell, Malitz, Moss, Weydert and others, my analytical theory of sets is based on a view of the mathematical activity as a process of gradual, open-ended (transfinite) unfolding of all possible structures. This is opposed to the classical, "synthetical" and *cumulative* (or *iterative*) conception of set, in which the idealized mathematician is gradually building the universe himself, in an open-ended process of set-formation. The classical quasi-constructivist approach is replaced by a dual, "de-constructivist" one, based on the intuition that *we are not the creators of the mathematical universe, only its explorers*; sets are not all built up in transfinitely many stages (although some of them are), but their structure is unfolded, analyzed, distinguished in successive stages.

To obtain a coherent and definite picture based on the above-mentioned two ideas, some more ingredients must be added:

(3) The traditional distinction between *sets* and *classes* has to be properly understood as a subtle distinction between two aspects of structures, namely the *actual* and the *potential* structures. These correspond to two ways of presenting a structure: as an object or as a mere relation between objects. This dichotomy has been traditionally confused with many others: finite versus infinite, small versus large etc. We need to carefully distinguish these various meanings of the set/class distinction.

(4) The necessity of giving an identity criterion for sets leads, in the structural conception, to the notion of *observational equivalence* between structures. Sets having indistinguishable structures must be identical. This is the basis of axioms like the classical Extensionality, Aczel's Strong Extensionality and my axiom of Super-Strong Extensionality. As explained below in section **4**, I define observational equivalence between two processes of unfolding as their capacity to simulate each other up to any (ordinal) degree of accuracy. This is a generalization of the standard notion of bisimulation; there are several equivalent ways to do this, and the one I shall use in this paper employs *partial structural descriptions*, which can be seen as infinitary modal sentences.

(5) The above-mentioned structural understanding of the set/class distinction removes the traditional limitations imposed on the concept of set, making possible a truly *maximal* conception of the set-theoretical universe. The latter will be based on a literal reading of the expression "all possible structures"; this unrestricted interpretation of the structural metaphor is the source of my Super-Antifoundation Axiom ($SAFA$), a strengthening of Peter Aczel's Antifoundation Axiom (AFA). My axiom says basically that *every transfinite pattern of structural unfolding can be seen as the unfolding of a set.* As a consequence, every definable structure, regardless of size, is in some sense realizable as a set. This is a totally "naive", unrestricted view of sets-as-structures. The only restriction comes from the above-mentioned understanding of set identity, which forces us to identify sets corresponding to observationally equivalent structures. This gives the following precise meaning to the above claim: every definable binary structure is observationally equivalent, but not necessarily isomorphic, to the $\in$-structure of some set.

I will show that the analytical conception is *more comprehensive* than the iterative one: it proves the existence of non-wellfounded sets like Aczel's $\Omega = \{\Omega\}$, of "over-comprehensive" sets like the universal set (of all sets) $\mathbf{U}$ or like the "largest ordinal" $\overline{On}$ (which is just the set of all ordinals, including itself), of fixed points for all monotonic operators, of very large cardinals etc. The resulting universe has interesting closure and fixed-point properties; it also satisfies very strong comprehension principles, namely the so-called Generalized Positive Comprehension Principle, proposed by Malitz, Weydert and Forti.

One can show that, unlike the classical naive theory of sets-as-collections, this "naive" concept of sets-as-structures is provably *consistent*, if we assume some mild large cardinal hypothesis. My model for this theory is just an infinitary generalization of the classical "canonical model" construction, used for proving completeness for modal logic. Under certain assumptions, the model is isomorphic to Forti's topological "hyperuniverse".

Chapter 2

Historical Discussion

2.1 Sets and Classes

The Aristotelian distinction between actual and potential existence was recovered by modern mathematical thinking in the guise of the *set/class* distinction (Cantor's *consistent/inconsistent* multiplicities). I regard this as a distinction concerning structures, namely between two different aspects of a structure: the purely *relational* (or logical) notion of structure and the *substantial* (or mathematical) understanding of structure. The first takes structures as mere relations linking objects, or as predicates, ways of talking about objects; the second deals with structures as objects-in-themselves. We can only *define and explain* relations and classes; but we can actually *observe and explore* objects and sets.

The structuralists seem to assume the first notion as the fundamental one and consider the second notion to be simply inconsistent. On the contrary, I think both aspects are essential to our understanding of structures. The very same structure exists potentially in its objects, as a relation, and can also be considered as an actual existence in its own respect. These two sides of a structure might not match each other perfectly: maybe an actual object is "richer" than its potential counterpart. But, following Cantor (or at least my interpretation of his views), I believe that these two aspects are inseparable: the potential presupposes the possibility of the actual. This means that, in some sense, every possible structure can be "actualized". In set-theoretical terms: every class can somehow be realized as a set.

The central subject of this paper is to give a precise mathematical meaning to this notion of realization of a structure: how can we consistently understand the relation between a potential structure and its actualization?

But first, the distinction between sets and classes needs to be freed from its ties to other pairs of opposite concepts, which are traditionally associated with it: finite/infinite, small/large, predicative/impredicative, circular/non-circular, well-founded/non-wellfounded. These latter distinctions are usually based either on non-structural features (e.g. *size*) or on some quasi-constructivist restrictions on the kind of structures that are permissible in set theory (e.g. *predicativity* or *well-foundedness*).

In this section and the next one, I consider some of the traditional conceptions about infinite totalities: the Aristotelian-medieval view of infinite-as-incompletable, the Cantor-Fraenkel-von Neumann doctrine of limitation of size and Zermelo's iterative conception. I argue that each of these doctrines is based on the confusion between the set/class distinction and one, or some, of the above-mentioned pairs of opposing notions. Nevertheless, these conceptions are useful as ways of isolating certain important kinds of well-behaved sets.

2.2 Actual versus Potential

The opposition between the *potential* and the *actual* played a key role in Aristotelian and medieval thought. Someone may be actually a child, but potentially a full-grown man; something may be a piece of stone, but potentially a sculpture. Aristotle explained movement, time and becoming as the actualization of potentialities. To solve Zeno's paradoxes, he also applied this distinction, and the associated *temporal* metaphor, to the concept of *infinity*. He distinguished between the *potential infinite* and the *actual infinite*, and he observed that the natural numbers are potentially infinite, but not actually so. From this, he arrived at the conclusion that there is in general no actually infinite object, and moreover that such an object cannot be conceived. Hence, in the case of infinity, he made an exception to the general rule that, if something is potentially thus and so, then it may become actually thus and so, or at least it might be imagined to be: the stone may become a sculpture, and this is what makes it potentially so; while it is utterly impossible for anything, including the potentially infinite, to be actually infinite.

To show this, he *redefined* the metaphysical concept of infinite. The Greek expression '*to apeiron*' meant literally "the unlimited, the unbounded", and for Anaximander it also meant perfect, whole, unconditioned, imperishable, inexhaustible, uncountable. This meaning was recovered later by Plotinus and the medieval thinking, and associated with transcendence, absolute, plenitude and divinity. But for Aristotle, infinite meant something completely different, almost the opposite of all the above: he says that 'the infinite turns out to be the contrary of what it is said to be. It is not what has no part outside it that is infinite, but what always has some part outside it' (Aristotle 1983, *Physics*, III, 6). The infinite is the *unfinished, the indefinite, the unrealizable, that which cannot be completed, it cannot be actualized*. It is just an eternally unfulfilled, unending succession which never reaches its limits. The infinite is never given *all at once*, but only *over* time. Observe the temporal metaphor:

> 'In general, the infinite exists through one thing being taken after another, what is taken being always finite, but ever other and other' (Aristotle 1983, *Physics*, III, 6).

For Aristotle, then, *actual implies finite*, and infinity is simply synonymous with "pure potentiality, which can never be actualized". In this way, actual infinity becomes a logical contradiction. The only way I can understand this concept of infinite is to give it a *topological* flavor: it seems to point to processes or totalities which are *not closed* in the modern topological sense (i.e. they do not contain their 'limits' or accumulation points); for how else can we make sense of the above-mentioned definition of the infinite as 'that which has some part outside it'? Compare with the following quote:

> 'Something is infinite if, taking it quantity by quantity, we can

always take something outside' (Aristotle 1983, *Physics*, III, 6).

It is implicitly understood that we can take something outside which *still belongs to it, and should actually be inside*(but it's not). I shall return to this topological reading later.

Recall that Aristotle was applying his analysis of the concept of infinity to the *mathematical* notion of infinity, that which is a possible answer to the question "How many?" (as in "how many prime numbers exist?"). But the mathematical infinite is not obviously identical to the incompletable. As observed by Cantor, this mathematical notion refers to the *size* of "infinitely big" collections, such as the natural numbers, and only asserts their incompletability through *certain means*, namely through counting. *We are* the ones who cannot actualize or realize the completed totality of all natural numbers, in the restricted sense that *we cannot count them.* Hence they are surely "incompletable" (or "infinite" in Aristotle's sense), but *only from a limited point of view*, from the perspective of our counting abilities. There is no a priori reason for which we (or God, for that matter) would not be able to complete them through some other means, or simply grasp them as object-like, actual, undivided wholes. But this is precisely what was assumed to be impossible by the Aristotelian analysis.

2.3 The Actual Infinite

In effect, what Aristotle did was to **(1)** disregard the ancient notion of infinity-as-whole, and **(2)** identify his concept of infinite-as-pure-potentiality with the mathematical notion of infinite-in-size. The Neo-Platonic and medieval Christian thinkers have basically continued to accept **(2)**, while reevaluating **(1)**: the actual infinity, in the form of infinite-as-unconditioned-whole, was established as the central object of religion, theology and philosophy, but was kept out of mathematical science, being regarded as incomprehensible by reason.

The work of Bolzano and Cantor broke with this tradition. The Cantorian concept of set is that of a *definite, object-like, actual, completed totality.* He speaks of a set as "a many which can be thought as a *one*" (Cantor 1883), "that is, every totality of definite elements which *can be united to a whole through a law*". In a 1889 letter, he says:

'When ... the totality of elements of a multiplicity can be thought without contradiction as "being together", so that their collection into "one thing" is possible, I call it a *consistent multiplicity* or a *set.*'

'As an example, I give the collection, the totality of all finite positive, whole numbers. This set is a thing for itself and forms, quite apart from the natural order of the numbers belonging to it, a definite quantum fixed in all parts... '(Cantor 1887-1888).

There are two distinct notions here: the concept of a general "multiplicity" , collection or totality, *possibly "inconsistent"*; and the concept of a "consistent multiplicity", a "definite totality", a collection which can be thought as "one thing", i.e. a *set.* This is Cantor's version of Aristotle's distinction between actual and potential. In the rest of this paper, following the modern, post-von Neumann usage, I shall use the terms *class*, collection and family to denote the first concept, while reserving the term *set* for the second concept. I shall regard classes as having only a *potential*, or second-order, existence, as being *about* the mathematical domain, but not necessarily being *in* it, as definite, completed members. On the contrary, sets are to be seen as actual, complete and definite objects, "first-class citizens" of the mathematical realm. Classes are more process-like (as in the process of successively generating all the natural numbers or all the ordinals), while sets are more object-like. Nevertheless, sets can be seen as coming from classes, as the actualization of some process of generating all their members. Sets can be thus understood as the "unified", completed version of their corresponding classes (as in 'a collection which can be thought as one thing'). The so-called *proper classes* are Cantor's inconsistent multiplicities: the classes which are not sets, i.e. which cannot be considered as being already complete, actual objects. Whether or not they can be somehow completed and actualized is a different question.

Some people would reject the above distinction, because they consider that referring to these purely potential classes as if they were objects is both inelegant and philosophically suspect. They insist that class-talk should be avoided and that it should be used only as a manner of speaking, as in Zermelo-Fraenkel's system ZFC. I actually agree with them! I think that talking about classes is useful and makes good sense; but that, as long as they are just classes, they should be used only as predicates, and not referred to as subjects of predication (except maybe as an eliminable way of speaking). After all, they are supposed to have only a potential existence. Their actual counterparts (if any) are the sets, so the only proper way to talk about a class as an object is to talk about its corresponding set. So I prefer to simply have sets as the only objects, and to refer to classes only indirectly, through their corresponding predicates (=unary formulas), as in ZFC. I think this reflects nicely their evanescent and process-like nature: how else to give a process but through its defining law or rule-of- formation (predicate)?

The Comprehension Problem, which can be seen as originating in Cantor's rejection of Aristotle's identification between the infinite-in-size and the un-actualizable, goes the other way around. It asks whether every class can be actualized into a set. If *not*, then what is the criterion for sethood, and how can it be justified? If *yes*, then in what sense is the resulting object (set) identical with the underlying process (class)? As we shall see, the structural-analytical conception of sets answers paradoxically *yes* and *no*. And the answer is perfectly consistent!

Against Aristotle, Cantor freed the potential/actual distinction from its ties to the size-originated pair finite/infinite. He claims that there are infinite-in-size collections such as the natural numbers, which can be consistently conceived as multiplicities unified into a whole. But how can he claim something like that? How can we, as *finite beings*, make an infinite sequence into an actual and complete object?

Cantor answers by appealing to God's infinite power. Who cares if we can conceive a collection as ‘one’, as long as God can do it? And God can surely do it, as Cantor rejoices:

‘They [the natural numbers] form in their totality a manifold, *unified thing in itself, delimited* from the remaining content of God’s intellect, and this thing is itself again an object of God’s knowledge ’.

He quotes St. Augustin: ‘every infinite is, in a way we cannot express, made finite to God ’. This is what M. Hallett calls “Cantor’s finitism” (Hallet 1984). In Weyl’s words “...for set theory, there is no difference in principle between the finite and the infinite”. *The size-distinction does not matter anymore,* from the moment Cantor decided to see the world “through God’s eyes ”. In other words: mathematics should not depend on trivial human limitations, such as finitude.

There is a catch though: the same argument can be used to show that the collection of all sets is a set, or more generally that every collection should be somehow actualizable. Surely God can conceive the totality of all sets as a unified thing. After all, ‘*every* infinite is made finite to God’. Does this mean that there is no real distinction between potential and actual after all, at least from God’s point of view? This is exactly how (the renowned atheist) Bertrand Russell understood Cantor’s theory (before the Paradox), as asserting the so-called Naive Comprehension Principle: every class is a set. The consequences are known: the set-theoretical paradoxes, and the subsequent restrictions on the notion of set imposed by Russell, Zermelo and others.

As pointed by many, there are no Cantorian texts to back up this naive set theory. By stressing the *unity* (or "wholeness", or "thingness") that collections should posses in order to be sets, Cantor implicitly denies that all classes are sets. But he *does* assert something very close to this. Namely, he says that 'each potential infinite, if it is rigorously applied mathematically, presupposes an actual infinite' (Cantor 1886). Or again:

> The potential infinite is only an auxiliary or relative concept, and always indicates an underlying transfinite without which it can neither be nor be thought (Cantor 1886).

Cantor stresses in particular that the *domains* in which mathematical variables (in a particular theory) take values have to be sets: 'in order for there to be a variable quantity in some mathematical study, the domain of variability must be (...) a definite, actually infinite set '. Apply this to set theory itself and you will obtain the set of all sets.

One wonders how to consistently understand the two apparently contradictory claims: *only* the "completed, definite, unified, whole" collections are "actual objects", *sets*; but *every* potential (infinite) collection "presupposes", "indicates an underlying" actually infinite set. It seems to be only one solution: the presupposed, underlying actual set is *not always identical* to the given potential collection, but can nevertheless be regarded as its *actualization*. To actualize something you might have to change it. But in such a way that you do not lose anything, otherwise there is no sense to say you actualized it.

2.4 Boxes inside Boxes

The two main readings of the Cantorian definition of "sets", that are commonly used to justify the most widely accepted set-theoretical system (ZFC), are the *limitation-of-size* conception and the *iterative* (or cumulative) conception. Both are based on what has been called [Barwise and Moss 1991] the *box metaphor*: "A set is like a box, and forming a set is like putting things in a box". This is a particular intuition about sets, which could be called the "bottom-up" view of sets: one collects things to put in the box.

2.4.1 The Zermelo-Fraenkel Axioms

As mentioned above, the almost universally accepted axiomatic system for set theory is the Zermelo-Fraenkel system ZFC. This is the system proposed by Zermelo and strengthened by Fraenkel and von Neumann (who have added the Replacement Axiom). All classical mathematical theories and concepts can be defined and developed inside ZFC. This system has become the standard foundation for all mathematical studies. It is our opinion that *any reasonable alternative set theory should preserve all the advantages of ZFC, by proving the existence of a transitive class V such that $(V, \in)$ is a model of ZFC*. Moreover, it is highly desirable that the this model is *standard*, i.e. is closed under set-formation.

We present here the Zermelo-Fraenkel axioms. The *language* of ZFC is the first-order language with two binary relations: *membership* $\in$ and *equality* $=$. We use small letters $a, b, x, y, \ldots$ for sets. We use capital letters

$A, B, X, Y, \ldots$ to denote *classes*, i.e. arbitrary formulas having a designated free variable x. We use *class-relations* $R, Q, \ldots$ to denote arbitrary binary formulas, having two designated variables x, y etc. We use *class-functions (or operators)* $F, G, \ldots$ to denote class-relations which are functional, i.e. for which we assume that for every set x there is a unique set y such that $F(x, y)$. For convenience, we extend the membership and equality notations to classes: we write $x \in X$ for $X(x)$, $(x, y) \in R$ for $R(x, y)$, $X = Y$ for $\forall x(x \in X \leftrightarrow x \in Y)$, and $F(x) = y$ for $F(x, y)$.

We also introduce a standard set-theoretical notation for classes: we use $\{x : x \in X\}$ to denote the class X. In ZFC, we say that *the class* $\{x : x \in X\}$ *is a set* if we can prove that

$$\exists y \forall x(x \in y \leftrightarrow x \in X).$$

In this case, the set y with this property is unique, by the axiom of Extensionality (see below). We shall informally identify the set y with the class X in this case, and use the notations X, $\{x : x \in X\}$ for both. In this way, it is clear that every set y can be identified with some class, namely with $\{x : x \in y\}$. (In general, the defining class is not unique.) We define *inclusion* for both sets and classes by:

$$X \subseteq Y \iff \forall z(z \in X \to z \in Y)$$

(the definition for sets is similar and can be considered a particular case of the one for classes, via the above identification of sets with special classes). We also introduce the following standard

Notations:

$$\begin{aligned}
\emptyset &=: \{x : x \neq x\} \\
V &=: \{x : x = x\} \\
\{a, b\} &=: \{x : x = a \text{ or } x = b\} \\
\{a\} &=: \{a, a\} \\
(a, b) &=: \{\{a\}, \{a, b\}\} \\
A \cap B &=: \{x : x \in A \text{ and } x \in B\} \\
A \cup B &=: \{x : x \in A \text{ or } x \in B\} \\
\bigcup A &=: \{x : \exists y (y \in A \text{ and } x \in y)\} \\
\mathcal{P} A &=: \{x : x \subseteq A\} \\
F[A] &=: \{y : \exists x \in A \text{ s.t. } y = F(x)\}
\end{aligned}$$

We assume *all the axioms of first-order logic with equality.* In addition, we postulate the following:

Extensionality. If x and y have the same elements then $x = y$:

$$\forall z (z \in x \leftrightarrow z \in y) \rightarrow x = y.$$

Pairing. If x, y are sets then $\{x, y\}$ is a set.

Separation. If x is a set and X is a class then $x \cap X$ is a set. Equivalently, if x is a set and X is a class such that $X \subseteq x$ then X is a set.

Union. If x is a set then $\bigcup x$ is a set.

Power Set. If x is a set then $\mathcal{P}x$ is a set.

Infinity. There exists an infinite set. More concretely, there exists some set y such that

$$\emptyset \in y \text{ and } \forall x(x \in y \to x \cup \{x\} \in y).$$

The least such set will be denoted by ω (the set of all natural numbers).

Replacement. If x is a set and F is a class-function then the image $F[x]$ is a set.

Foundation. Every set is well-founded, i.e. V is the least class closed under set-formation: if X is a class such that $\mathcal{P} \subseteq X$ then $X = V$. In other words, for every class X we have:

$$\forall x(x \subseteq X \to x \in X) \longrightarrow \forall x(x \in X).$$

Choice. Every set a of non-empty sets has a choice-function: if $\forall x \in a(x \neq \emptyset)$ then there exists some set f, such that $\forall x \in a \exists! y(x, y) \in f$ and $\forall x, y((x, y) \in f \to y \in x)$.

Note that the axiom of Separation is superfluous, since it is implied by the axiom of Replacement. We keep it only for historical reasons. Also, observe that Separation, Replacement and Foundation are actually *axiom-schemes*: they involve a universal quantifier on classes, which means that for each formula (class) X (or F) we have a corresponding instance of each of these axioms. It is well-known though that Foundation can be restated as a single axiom.

The system ZFC^- can be obtained by eliminating the axiom of Foundation from ZFC. As it has been observed by many authors, Foundation is an artificial restriction on the set structures, whose primary justification is purely technical: it makes more definite our concept of set, allowing us to prove things by induction along the membership relation ($\in$-induction). But Foundation is not really necessary for developing any particular branch of mathematics inside set theory. One can simply work in ZFC^-, define the class WF of all well-founded sets (e.g. as the least class closed under set-formation) and prove that this class is a transitive standard model for ZFC (including Foundation).

Note also that all the axioms of ZFC^-, except for Extensionality and Choice, are particular cases of the general Naive Comprehension Principle (stating that every definable class is a set). Indeed, these axioms were obtained by Zermelo by weakening Naive Comprehension, in such a way that no paradoxes arise, but classical mathematics and Cantor's theory of transfinite cardinals can be still developed and Cantor's Well-Ordering Principle can be proved as a theorem.

But, apart from their usefulness, what is the justification for keeping precisely these instances of the Naive Comprehension Principle, and not others? It is natural to look for an intuitive concept of set, more restricted than the "naive" one, but for which the Comprehension-like axioms of ZFC would still be obviously true.

2.4.2 Limitation of Size

The limitation-of-size conception (Fraenkel, Bernays, von Neumann, Levy) understands the "unity" or the "thingness" of a collection as related to its *size*: a class is to be regarded as an object if it is *not too big*. Observe the underlying *spatial metaphor*, which puts a *cardinality* restriction on the box metaphor: "one cannot put too many things in the same box". Fraenkel uses the physical, spatial metaphor of a *encompassing wall*:

'Sets have around them a closed wall which separates them from the outside world', while to form proper classes, 'elements have to be taken from outside every wall, no matter how inclusive'(Fraenkel)

He also says proper classes are "contradictory because of their limitless extent". In a way, this is a rehabilitation of the pre-Cantorian conception that the "actuality" of a collection has something to do with its size: for Aristotle and the medievals, all infinite classes are "proper classes", since only the *finite* collections are actual, while all those with size bigger than every natural number have only a potential existence. Now, this "barrier" is just lifted, so that all classes which are smaller than some big ones (such as the universe U) are considered "sets", all the others (the "proper classes") having just a potential reality.

This identification of the distinction set/class with the small/large distinction is so pervasive today, that I must stress again, for the sake of clarity, that I do *not* adhere to it: for me, classes are just the loose-potential-collections which do not necessarily form a unity, whereas sets are unified-actual-classes.

While accepting that small collections are sets, I agree with Hallett and others that there are *no* good reasons to think that all sets are small (in any sense of the notion of small); such an assumption seems like an ad hoc solution to Russell's and Burali-Forti's paradoxes, solution obtained by throwing out each and every perfectly benign large class, just for the sake of getting rid of the inconsistent Russell class $R = \{x : x \notin x\}$ or the inconsistent class On of all the ordinals. After all, one can easily turn against it Cantor's own arguments, aimed at the Aristotelian sized-based identification of the infinite with the pure potentiality. Maybe not the excessive *size* of R and On is the reason for their "incompleteness", but their sheer "unfinished" character.

Moreover, the limitation-of-size conception does not fulfill its promise: it cannot explain all the axioms of ZFC, and its spirit is actually in contradiction with some of them. It can easily justify axioms such as the existence of the *Empty-Set, Pairing, Replacement* (and *Separation*); for *Replacement* it is indeed the only explanation available at this moment: recall that *Replacement* simply asserts that any collection which is not bigger than a given set is also a set. The conception can be ad-hoc adapted to accommodate *Infinity* (by somehow understanding "smallness" in a sense that doesn't imply finitness) and *Union* (by replacing the smallness by "hereditarily smallness": sets are hereditarily small classes, i.e. small collections made out of small collections etc.). But it offers no justification for either the *Foundation Axiom* or the *Power-Set Axiom.* Why should the power of a small set be itself small? Even if this is true, it is not at all obvious. On the contrary, Cohen's and Easton's results imply that there is no provable (non-trivial) connection between the size of an infinite set a and the size of its power-set $\mathcal{P}a$; the power-set of an infinite set "evades all our attempts to characterize it by size" (Cohen); moreover, because of its non-absoluteness, it is closely related to the size of the whole universe. So, on this conception, the power-set axiom is at least counter-intuitive. One might still be compelled to accept it on the basis of some proof, but not as an axiom. In M. Hallet's words, "the power-set axiom just is a mystery " (Hallett 1984).

In conclusion, I have to stress that the limitation-of-size conception does separate out an important collection of sets: the small ones. Size distinctions are central in set theory, and small sets behave differently from the large ones. In a framework such as the one adopted here, allowing the existence of large sets, the small ones can be neatly characterized as the ones which satisfy the *Replacement* axiom. The hereditarily small sets will play an important role in the axiomatization of our framework.

2.4.3 The Iterative Conception

The iterative conception of set (Zermelo, Schoenfield, Scott, Wang, Gödel) is based on a *temporal* understanding of the box metaphor: boxes are formed in stages and put inside the new boxes. There is a temporal restriction on the set-formation: you cannot put into a box objects which are not yet created. So the "thingness" of a collection is again understood in terms of the actual versus potential distinction: at each moment, only the sets that have already been created are actually existing, and so are available as building blocks for the new sets. The "yet uncreated" sets exist only potentially. Totalities like the set of all sets cannot be formed at any stage, since this would entail an instant actualization of all the sets that might be formed in the future. They contain as members sets that occur at arbitrarily late stages, so that their collection can never be formed. Such totalities cannot be completed and so are condemned to remain in a potential state forever.

More precisely, it is assumed we have a logical sense of Time, consisting of *stages* with a relation of *precedence* between them. These stages are identified with the ordinals, which means the "logical" time is transfinite and well-ordered. The idealized mathematician is building the universe of sets in time. At each stage he forms all the collections consisting of sets formed at earlier stages. He adds these collections to his stock of available sets and then goes to the next stage. The process *has a memory*: it cumulates all the sets that have been formed. Notice the quasi-constructivist flavor of this conception.

The iterative picture is very powerful and intuitive, and has become the dominant conception in modern set theory. It easily justifies the axioms of Pairing, Union, Infinity and Foundation. It can be argued that it also justifies Separation and Power Set. But it clearly *fails* to offer an explanation for Replacement: there is no apparent reason for which the members of a functional image of a previously formed set might not occur at arbitrarily late stages.

The standard answer is that, once we have a function F, we can imagine a new stage beyond any of the old ones and form the range of the function at this new stage. There are many reasons to consider this response as flawed. A definable function is just a formula and each addition of a new stage, each extension of the universe, might change the meaning of the formula. After adding a new stage, we can indeed form the set that corresponds to the old range of F, but this will not be the range of F in the new universe.

Sometimes the answer takes again the form of a size metaphor: one assumes that the collection of stages (the logical "Time") is *large*, compared to each of the sets formed in the process. There are more stages than there are elements in any set. In other words, each of the sets created at any stage is smaller than the collection of all stages. This seems a plausible assumption, and this combination of the temporal iterative metaphor with the spatial metaphor of size can indeed justify Replacement. But how can we consistently understand this mixture of metaphors? How can we ensure, in advance, that the length of available Time will be longer than the enumeration of any of the sets that will be formed during this Time?

Some authors understand this in terms of an "open-ended" concept of time: they think of the temporal stages as being themselves build up "in time", together with the sets. After we have created a set a at some stage α, we don't just add a "next" stage, but we possibly have to add many more stages, enough many to enumerate the sets created at stage α. If this picture would be consistent, it would indeed justify Replacement. But unfortunately, this view seems to me to be obviously inconsistent, despite its apparent intuitive character: what is the meaning of building Time "in time"? Even worse, observe that in this view, we do not just create one "stage" at a time, by building all the sets associated with it, but we can somehow get "ahead of time" to create infinitely many future stages, before actually building the sets associated to them! Such an assumption seems to render meaningless the basic iterative concept of "building sets in successive stages, one after another, in a quasi-constructive manner".

Despite all the talk about open-endedness of time, it seems to me that a reasonable and consistent understanding of the iterative picture would have to take the collection of available temporal stages as *given*. We create sets in time, and some of these sets may reflect the structure of the temporal stages, up to isomorphism; but we cannot jump out of the logical time to create some more time. Any justification of Replacement that assumes such an ad-hoc, magical power of "creation before creation", is logically flawed, and defeats the very purpose the iterative picture is supposed to serve.

But if we think of the collection On of stages as given, then how can we ensure that the sets which will be successively created will be smaller than On itself? The most natural way to do it would be to impose a limitation-of-size condition on the iterative construction: at each stage we only form the *small* collections of previously formed sets, where "small" is understood as "being smaller than the collection On of all stages". In this way, we only form sets which can, in principle, be enumerated at some future stage. This view can successfully explain Replacement. But it unfortunately fails now to justify Power Set, for the same reasons the pure limitation-of-size conception failed to do so: there is no warranty that the powerset of a small set will be itself small.

There are problems with the Power Set anyway: in its full, impredicative acception, this axiom goes against the quasi-constructive character of the iterative conception. The fact that we can eventually build all the subsets of a set does not obviously imply the possibility of having them all collected into a whole. A more realistic understanding of the iterative picture would insist that at each stage we form all the *sets* (i.e. all the collections that *we can actually form* as unified things) consisting of previously formed objects. But whether or not this justifies Power-Set is debatable: there are good arguments (connected to Cohen's work on forcing) to think that we can *never* complete the full power set of an infinite set: as the universe extends, the power-set grows too (non-absoluteness).

Chapter 3

The Structural Conception of Set

The “structure- forgetting metaphor” was proposed by J. Barwise and L.Moss in [Bar 1991], as a motivating intuition for Aczel’s Antifoundation Axiom (AFA) and as an alternative to the usual “box“ metaphor. The idea is that sets are what is left when we take an aggregate (a complex object) and we *abstract everything but its structure*.

By forgetting the nature of the components, the only thing that remains is the aggregation/disaggregation relation between the whole and the components, i.e. the *membership structure*. This structure is *pointed*, in that it has a root: the underlying process of unfolding the structure, by successive decompositions, has a starting point, namely the *very object* under consideration. So we look at sets as *pointed binary structures*.

One can think of this conception as turning the iterative picture on its head: instead of starting at the bottom and building sets in stages, as collections of previously given objects, we are now presented, from the start, with a unified totality (an "object"), which we analyze into its constituents, which in their turn are to be analyzed...and so on. This is a "top-down" approach to the concept of set. In terms of the box metaphor, we do not put things in boxes, they *come in boxes*. Moschovakis (in [Moschovakis 1994]) uses a "gift" metaphor to introduce Aczel's hypersets: we receive each box as a gift from the universe; we just have to "unwrap" the box to see what is inside and to continue doing this for each new box encountered in the process. By forgetting everything but this "pattern of unfolding" we obtain a set.

Observe that under this conception there is no reason to limit the possible structures to wellfounded ones. A "possible world" containing infinitely divisible objects (i.e. not ultimately reducible to unstructured, atomic components) is surely conceivable - and this is enough for mathematics, regardless of whether the real world is "atomic" or not.

So the Foundation axiom has to go. But this immediately poses the question of determining the *identity of sets*: what is the essential feature of a set? As many people have pointed out (cf. Forster in [Forster 1995]), maybe the most fundamental and least controversial principle about sets is the view that *all there is to know about a set is its members.* In the context of Foundation, this principle can be reduced to the Axiom of Extensionality: sets having the same members are identical. Foundation allows us to use this principle to *define the relation of identity* on sets by $\in$-recursion. In the absence of Foundation, we need some stronger notion to determine the identity of sets.

In the context of the structural conception, there is a natural generalization of Extensionality, that comes from the very principle of abstraction we assumed: a set is nothing but a pattern of unfolding of a possible structure; so the *identity of sets should be given by the identity of their analytical patterns.*

We obtain a notion of *structural equivalence*, based on the identity of analytical behavior: two structures are equivalent if their patterns of unfolding are the same. I will refer to this relation as *observational equivalence*, since I think of the above analysis as a series of observations performed on the object in question. As we shall see, there are various ways to rigorously define this notion. All these definitions are equivalent for *small* structures. All of them have the following property, which is obvious in the above informal description of an analytical pattern: an unfolding of a set is given by unfoldings of its elements; in purely structural terms: an unfolding of a pointed structure having some root g is given by unfoldings of pointed structures having (immediate) successors of g as their roots.

So we can state the above criterion for set identity as: *sets having observationally equivalent structures are identical.* Depending on the notion of observational equivalence that is used, we shall call this principle *Strong Extensionality* or *Super-Strong Extensionality.* In its super-strong formulation, this principle can be seen to capture the full structural content of the above-mentioned slogan about sets as being determined only by their members; this is because, as the above discussion shows, the unfolding pattern of a set is uniquely determined by the unfolding patterns of its members.

Observe that this does not reduce set-identity to simple structural isomorphism, because of the *untyped* character of the set concept: the components are themselves regarded as being sets. In other words, sets are structures composed of sets.

For example, take an object, which happens to be a nonempty aggregate of components, each of which is itself a nonempty aggregate, and so on.... Clearly, all the components have the same analytical behavior (pattern of decomposition) as the initial object; if we forget everything but this pattern, then all the components will be identified with the initial object. The resulting set has only one member, namely itself; this is Aczel's $\Omega = \{\Omega\}$, the ultimately frustrating gift (as Moschovakis called it): a box having inside a box having inside a box...all the way down.

Now we can state the main principles of the structural conception:

> **Maximality**(or **Structural Completeness**): *every possible pattern of structural unfolding is the pattern of unfolding of some set.*
>
> **Uniqueness** (or **(Super)strong Extensionality**): *two sets that are observationally equivalent (i.e. have the same patterns of unfolding) are identical.*

Depending on the precise definition of observational equivalence, these principles will be embodied by Aczel's axiom of Antifoundation (AFA) or by our axiom of Super-Antifoundation ($SAFA$). The second of these principles (the uniqueness side) is simply the above-mentioned Principle of (Super)Strong Extensionality. The first principle (maximality) will be referred to as the *existential side* of these axioms. It is a maximality postulate, stating that every possible structural pattern is realized. Depending on the understanding of what a "possible pattern" is, we can obtain (the existential sides of) Aczel's axiom AFA, our axiom $SAFA$ or a weaker version, called *Weak* $SAFA$.

In any of the formulations, the conjunction of the above two principles is easily seen to imply the following claim

> Every (pointed binary) structure is observationally equivalent to a unique set.

This confirms that the underlying conception of set is indeed *purely structural*: there are no restrictions to be imposed to our (pointed binary) structures in order to represent them as set structures, as long as we are satisfied with having a representation only up to observational equivalence. Moreover, there are good reasons to be satisfied with such a representation: (by uniqueness or super-strong extensionality) we cannot hope in general for a better representation, since observationally equivalent sets have to be identical.

Of course, for particular structures, we can have perfect set representations: they may be isomorphic to the structure of a set. But if we are looking for representation up to isomorphism, we can see that the Principle of (Super)Strong Extensionality (in any of its formulations) will still impose some *limitations* on the kind of binary structures that we can get. Namely, these structures will have to be *(super-)strongly extensional*, i.e. the relativization of (Super)Strong Extensionality to these structures has to hold.

The above discussion is obviously informal: we have not made explicit what our background assumptions are; we have not attempted to define the notions of structure and observational equivalence. In the rest of this chapter, we shall give various semi-formal implementations of the above concepts.

The first is just Aczel's theory of hypersets ZFA. We shall later see that, by weakening the background assumptions of this theory, we can afford to strengthen our notions of structure and observational equivalence to match the informal ones of naive set theory. The resulting theory will simply be called "the Structural Theory of Sets" (STS).

3.1 Antifoundation and Strong Extensionality

Working in ZFC^-, i.e. the system obtained by eliminating Foundation from ZFC, Peter Aczel has defined the notion of a *bisimulation* between two structures. The name comes from computer science, but the notion has already been used in set theory under different names.

Definition 3.1.1 A *set-bisimulation* is a relation $\sim$ between sets, having the property that, if two sets are in the relation $\sim$ then every member of one of the sets is in the relation $\sim$ with some member of the other set:

$$a \sim b \ \Rightarrow \ [\forall a' \in a \exists b' \in b \ a' \sim b' \ \& \ \forall b' \in b \exists a' \in a \ a' \sim b'] \, .$$

Two sets are said to be *bisimilar* if there exists some bisimulation $\sim$ between them.

Bisimilarity can be defined more generally for *pointed binary structures*. In the universe of ZFC^{-}, the only binary structures available are the *binary graphs*, which are defined as sets of ordered pairs of objects. The *nodes* of the graph are all the members of these pairs. The underlying binary relation R is called the *edge relation* (or the *accessibility relation*) of the graph. A node g' is an *immediate successor* of a node g if gRg'. A node g' is a *successor* of g if it can be reached from g by a finite chain of immediate successors. A *pointed (binary) graph* $G = (g_0, R)$ is a pair consisting of a binary graph R and a *root* (or a "point") g_0, having the property that every other node is a successor of g_0. One can define bisimilarity between two pointed graphs by replacing in the above definition the membership relation inside each set ($a' \in a$, $b' \in b$) with the converse of the edge relation in the corresponding graph:

Definition 3.1.2 A *bisimulation* between two graphs is a relation $\sim$ between their nodes, having the property that : if two nodes are related by $\sim$ then every immediate successor of one of the nodes is related by $\sim$ with some immediate successor of the other node.

A *bisimulation* between two pointed graphs is just a bisimulation $\sim$ between the two graphs, which relates the two roots. Two pointed graphs are *bisimilar* if they are related by a bisimulation.

Bisimilarity is Aczel's notion of observational equivalence. The same concept is useful in *modal logic*: observe that a pointed graph is just a Kripke structure having a distinguished world ("the actual world"). As we shall see later, bisimilarity can be alternatively defined as *infinitary modal equivalence*, i.e. elementary equivalence with respect to infinitary modal logic. The two definitions are equivalent in ZFC^{-}, but not in a more general setting allowing for "large" sets.

The above considerations on set identity as observational equivalence lead naturally to the following strengthening of the classical axiom of extensionality:

Strong Extensionality(P. Aczel): *Bisimilar sets are identical.*

This axiom basically says that sets are uniquely determined by their membership structure. This gives a clear-cut identity to Aczel's sets. One can *prove* now that every set having the structure described in the above example (i.e. a non-empty set consisting only of non-empty sets, each consisting only of non-empty sets...etc.) has to be identical to the set $\Omega = \{\Omega\}$.

The above discussion on the structural metaphor suggests that strong extensionality should be the only limitation imposed on the set structures. This leads to a maximality principle, stating that, *up to bisimilarity, every pointed binary structure can be seen as a set.* This is the existential half of Aczel's axiom of Antifoundation:

Existential AFA: *Every pointed graph is bisimilar to some set.*

By Strong Extensionality, the set mentioned in this statement is unique. So, putting together the last two axioms, one obtains a version of Aczel's main postulate:

Antifoundation Axiom (AFA): *Every pointed graph is bisimilar to a unique set.*

This statement seems to capture perfectly the structural conception of set: *sets are just pointed graphs modulo observational equivalence (bisimilarity).* There are other equivalent ways of stating this axiom, e.g. in terms of solving systems of equations (Barwise and Moss). Aczel has a different formulation in terms of the notion of decoration.

Definition 3.1.3 A *decoration* of a graph R is a function d_R mapping nodes to sets having the property that

$$d_R(g) = \{d_R(g') : gRg'\} \text{ , for all nodes } g.$$

Aczel states AFA in the form:

Every pointed graph has a unique decoration.

Aczel's system ZFA is obtained by adding the Antifoundation axiom AFA to ZFC^-.

Against ZFA, one can argue that the acceptance of the axioms of ZFC^-, as the background of our set theory, imposes an artificial *limitation-of-size* on the kind of structures sets are. In particular, Replacement and Separation force our structures to be *small*; the sets obtained by collapsing these structures in accordance to AFA will be *hereditarily small.* As it was observed by Barwise and Moss in *Vicious Circles*, it seems that ZFA can be best understood as a general theory of hereditarily small sets. Indeed, if we adopt the structural conception, then the iterative justification for the axioms of ZFC^- (in particular for Separation) is no longer available: sets are not built up from below in a wellfounded manner, but are obtained by collapsing arbitrary structures. The only justification for ZFC^- comes now from the limitation-of-size conception. But, as we know from the above discussion, the power-set axiom is *not* justified by this conception. So we are back to the "power set mystery".

Moreover, one can argue that, from the point of view of the structural conception, the "smallness" condition looks unnatural, a simple artifact of the way we usually represent (pointed binary) structures as *sets in* ZFC^-. It actually goes against the "maximality" principles embodied by AFA, by imposing again an ad-hoc restriction to the possible set-structures. Even worse, this limitation is of a non-structural nature: *size is not a structural characteristic* (in the sense described above), since it is not preserved by bisimilarity (or by any reasonable notion of observational equivalence). Indeed, one can construct graphs of arbitrarily large size, which are nevertheless bisimilar to the singleton $\Omega = \{\Omega\}$. Limitation of size is hence contrary to the spirit of the "structure-forgetting" metaphor.

3.2 The Naive Concept of Structure: an Analytic Approach

In this paper we explore what happens if we take the structural notion of set at face value and consequently drop any size restriction. The problem of finding such a theory was explicitly stated by J. Barwise and L. Moss in their book **Vicious Circles** ([Barwise and Moss 1996]), in the chapter suggestively entitled *Wanted: A Strongly Extensional Theory of Classes*. Our system STS can be understood as an attempt to answer this call.

By renouncing to the limitation-of-size condition, we return to the pre-set-theoretical ("logical") notion of *structure*: a binary structure is just a binary relation R, given by some formula. Set-theoretically, this can be still represented as a *class of pairs*, but not necessarily as a set, be it small or not. As mentioned in the Introduction, this is a completely naive conception of structure. If we also fix a "root", we obtain the class-analogue of a pointed graph: the notion of *pointed system*. This concept has been studied in the context of ZFC and ZFA, but not as a central notion. We shall take it as the basic logical notion underlying the concept of set.

Next, we have to define a notion of *observational equivalence* between pointed systems. Aczel's bisimilarity relation captures this notion only for *small* systems (graphs). It is *not* appropriate for large systems, since it is based on the unwarranted assumption that we can use one of the systems as a whole to describe or simulate the other. Let us suppose we first observe the roots g_0, $g_0^{'}$ of the two systems; next, we are presented with an immediate successor of g_0' in the second system, and we have to match it with some immediate successor of g_0 in the first system. But this assumes one is given a complete picture or list of the collection of all the immediate successors of g_0. This is not a natural assumption when defining a notion of observational equivalence between large systems.

A more realistic assumption is that, at each step we only have access to a list of *partial descriptions* of all the immediate successors. The list can only use descriptions that have already been constructed; hence, even if the collection of all the successors is large, the collection of their descriptions will be small. But this means we are not matching a node with another node, but a description of a node with another description. Both might actually refer to many distinct nodes.

After ω many steps, we might have to continue this process of unfolding the structures: we now have available more descriptions of the immediate successors of the initial roots. We obtain a transfinite sequence of unfoldings, which can be interpreted as a series of analytical experiments performed on the initial object (set or pointed system). Two objects will be observationally equivalent if they have the same pattern of unfolding.

Observe the underlying *temporal metaphor*: as in the iterative picture, we need a "logical" concept of Time, given by an unending succession of *ordinal stages*; but these are no longer stages of construction, but *stages of discovery.* The idealized mathematician is no longer the builder of the mathematical universe, only its explorer: he just unwraps the gifts for all his transfinite life. Keith Devlin, in his book on **The Joy of Sets,** called this an *analytic approach* to set theory, contrasting it with the *synthetic* approach of the iterative conception.

On the other hand, one can see that the analytic approach presupposes the synthesis: the outcome of an "observation" or experiment is a partial description of the object in question. To analyze something is to actually construct a new object, as a (partial) unfolding of the initial one, a (possibly incomplete) representation of its structure. The explorer has to record somehow the results of his explorations, cumulating all his past and present information in a database: a box. Each of the data gathered in the box is itself a partial description of the intended object: another box. In this sense, the iterative universe is the "shadow" of the analytic universe, its trace of unfolding: when we explore the "real world" of sets-as-structures, we simultaneously build the wellfounded universe of boxes-upon-boxes, as a way to encode the partial information gained in the first process. The stages of discovery are also stages of construction (of "theories" or databases). The wellfounded sets play here the role of "linguistic objects", descriptions of the intended (possibly non-wellfounded) structures. But in the same time they are "real" objects in themselves, being a part of the intended universe.

3.2.1 Structural Unfoldings

Let us give a semi-formal implementation of the above concepts. For the rest of this section we shall assume the following: we are given a universe

U of objects, called *sets* and satisfying Extensionality; we are also given a collection $V \subseteq U$ of well-founded hereditarily small sets, satisfying the axioms of ZFC (actually this is a much stronger assumption than we need). The elements of V will be used to provide descriptions for arbitrary sets or structures. Inside V, one can define the collection On of all (von Neumann) ordinals in the usual manner. We shall also assume an informal Principle of Definition by Ordinal Recursion: basically, this says that we can define operations and relations on arbitrary sets by induction on ordinals.

This principle can be made precise: let us say that a variable y is *bounded by a variable x in the formula φ* if all the occurrences of y in ϕ are in the scope of some quantifier of the form $\forall y \in x$ or $\exists y \in x$. For a given variable x, we say that an n^{ary} relation symbol R is *bounded by x in the first variable in the formula φ* if, for every subformula of φ of the form $R(t_1, t_2, \ldots t_n)$, the first term t_1 is a variable bounded by x in φ.

The Recursion Principle we need can be stated now in the following form:

> Given an $n+1^{\text{ary}}$ relation symbol $R(x, x_1, \ldots x_n)$ and a formula in $n+1$ variables $\varphi(x, x_1, \ldots x_n; R)$ such that R is bounded by x in φ, there exists some class-relation $\mathcal{R}$ such that we have
>
> $$\forall \alpha \in On \forall x_1 \ldots \forall x_n (\mathcal{R}(\alpha, x_1, \ldots x_n) \Longleftrightarrow \varphi(\alpha, x_1, \ldots x_n; \mathcal{R}))$$

Of course, the existential claim of this principle refers to classes: $\mathcal{R}$ is a class-relation. In ZFC and ZFA, this principle can be proved, by explicitly defining the class $\mathcal{R}$. In our axiomatic system STS, we shall assume as an axiom some instance of this principle, namely the Satisfaction Axioms, and then $SAFA$ will allow us to *prove* the full Recursion Principle, by explicitly defining the relation $\mathcal{R}$. But we cannot do this yet, so *we prefer to keep this principle informal*. (The only way to formalize it directly would be to introduce some fixed-point operator μ to the language, allowing us to form a new formula $\mu R\varphi(\alpha, x_1, \ldots x_n)$, which defines the above class-relation $\mathcal{R}$; but this would of course lead us away from first-order logic to some kind of a fixed point logic.)

Having fixed our background assumptions, we proceed to make explicit the above notions of *structural unfolding* and *observational equivalence*. Given a set a, we want to successively "unfold" its structure, i.e. to construct a succession of well-founded sets, which will be called the *unfoldings of the set* a. The unfoldings will encode (in their membership structure) all the information about the structure of a which is available at some definite stage of analysis. So the unfoldings will play the role of partial descriptions of the given set. The unfoldings form a succession, in which later unfoldings offer better descriptions than earlier ones.

We that one can think of this succession of unfoldings of a given set as being well-ordered. In other words, the stages of unfolding can be identified with the ordinals. Indeed, suppose we are given some stage of unfolding, i.e. we have constructed partial descriptions for every set. These are the best descriptions available at this stage. But we can now construct a *better description* for each set a, by collecting all the available descriptions of its members. This is the natural way to further "unfold" the set a, thus forming its "next" unfolding: we simply go inside the set and collect the best available descriptions of its elements into a new set, which gives us a better representation of a. So for every stage of unfolding there is a "next" stage, at which we form the next unfolding. Suppose now that we are given, not one unfolding, but a succession of unfoldings of the same set, such that there is no best one. We can then accumulate all these descriptions to form a single, better one. Formally, this new unfolding is simply given by the sequence of all the previous unfoldings. We conclude that, for every set of unfoldings, there is a "next better" one.

So the stages of unfolding form an unending well-ordered succession, in correspondence with the ordinals. This means that, for each set a and each ordinal α, we can form a set a^α, called *the unfolding of rank α of the set a* (or, for short, *the α^{th} unfolding of the set a*). We shall define these unfoldings according to the above ideas:

Suppose we are given the α^{th} unfoldings a^α for all sets a; we define the next unfolding $a^{\alpha+1}$ as the set of all the α^{th} unfoldings of the members of a.

For limit stages λ, suppose we are given all the α^{th} unfoldings, for all ordinals $\alpha < \lambda$ and all sets a. We define the λ^{th} unfolding of the set a as the λ-sequence of all its unfoldings of ranks less than λ.

What about the "initial" 0^{th} unfolding a^0? We just consider 0 as a limit ordinal, which gives us that a^0 is always the empty sequence $\emptyset$.

To summarize, a "successor" unfolding (having a successor ordinal as its rank) is a way of analyzing the given set into components and representing it by the set of all the immediately preceding unfoldings of the components; a "limit" unfolding is a way of keeping track of all the previous unfoldings of the given set and representing it by the sequence of these preceding unfoldings. This means that, as expected, the process of unfolding has a *memory*: limit stages are moments of recollection, at which one accumulates the results of previous stages. By contrast, the successor stages are moments of active analysis, at which one goes inside each set, decomposing it into parts.

More precisely, we can define the notion of unfolding by recursion on the ordinals:

Definition 3.2.1 For every ordinal α and every set a, the *unfolding of rank α of the set a* is the set a^α, defined by:

$$\begin{aligned} a^{\alpha+1} &= \{b^\alpha : b \in a\} \\ a^\lambda &= \langle a^\alpha \rangle_{\alpha<\lambda} \text{ , for limit ordinals } \lambda \end{aligned}$$

where $\langle a^\alpha \rangle_{\alpha<\lambda}$ is the sequence s of length λ given by $s(\alpha) = a^\alpha$ for every $\alpha < \lambda$.

The notion of unfolding can be similarly defined for classes or pointed systems (Kripke structures). For a *classe* A, we put:

$$\begin{aligned} A^{\alpha+1} &= \{b^\alpha : b \in A\} \\ A^\lambda &= \langle A^\alpha \rangle_{\alpha<\lambda} \text{ , for limit ordinals, } \lambda \end{aligned}$$

For a given *system* (relation) R(or graphs) we define the unfoldings $(g, R)^\alpha$

of the pointed system (g, R), for all nodes g and all ordinals α:

$$\begin{aligned} (g,R)^{\alpha+1} &= \{(g',R)^{\alpha} : gRg'\} \\ (g,R)^{\lambda} &= \langle (g,R)^{\alpha} \rangle_{\alpha<\lambda} \text{ , for limit ordinals, } \lambda \end{aligned}$$

The informal character of this definition is due to the fact that it assumes we can define functions on the whole universe of sets by recursion on the ordinals; in ZFC this could be done using Foundation, which makes possible to consider the above definition as a double $\in$-recursion on both sets a and ordinals α. But in our possibly larger universe U, we cannot do $\in$-recursion, so the above definition cannot be justified in this way. In the next chapter, we shall set up an axiomatic system STS in which one can prove a Recursion Theorem that can be used to justify any such definition.

There are other ways of defining notions of unfolding, for which this distinction between the successor and the limit stages is not so relevant; one can simply do both at once: at each stage, both analyze into components and keep track of the previous stages. But all the natural alternatives are essentially equivalent to our notion. One such notion of mixed unfolding has been defined and studied by Ronald Fagin under the name of "α-worlds" (see Fagin 1994). Fagin's $\alpha - worlds$ are defined for Kripke structures (instead of sets) and they provide a way of partially describing structures up to bisimilarity. These descriptions are essentially equivalent with ours, in the case of Kripke structures with only one accessibility relation, understood as membership.

Examples: (In the examples below, the notation α^{β} is only used to denote the β^{th} unfolding of α, and should *not* be confused with ordinal or cardinal exponentiation.)

- *The unfoldings of the empty set* $\emptyset$:

$$\begin{aligned}\emptyset^{\alpha} &= \emptyset \text{ , for all successor ordinals } \alpha,\\ \emptyset^{\lambda} &= \langle\emptyset\rangle_{\alpha<\lambda} \text{ , for limit ordinals } \lambda,\end{aligned}$$

  where the sequence $\langle\emptyset\rangle_{\alpha<\lambda}$ is simply the function mapping every ordinal less than λ into $\emptyset$.

- *The unfoldings of rank* 1:

$$\begin{aligned}\emptyset^{1} &= \emptyset\\ a^{1} &= \{\emptyset\} \text{ , for all non-empty sets } a \neq \emptyset.\end{aligned}$$

- *The unfoldings of the natural numbers*: for a natural number $n \geq 2$, we have

$$\begin{aligned}n^{k} &= \min\{n,k\} \text{ , for all natural numbers } k\\ n^{\omega} &= \langle 0,1,\ldots,n-1,n,n,n,n,\ldots\rangle\\ n^{\omega+1} &= \{\langle 0,0,\ldots\rangle,\langle 0,1,1,\ldots,\rangle,\langle 0,1,2,2,\ldots\rangle,\ldots\langle 0,1,2,\ldots,n,n,\ldots\rangle\}\\ &= \{k^{\omega} : k<n\}\end{aligned}$$

  etc.

- *The unfoldings of the set* $\omega = \{0, 1, 2 \dots\}$ *of all natural numbers*:

$$\begin{aligned}
\omega^n &= n \text{ , for every natural number } n, \\
\omega^\omega &= \langle 0, 1, 2, 3, \dots, n, \dots \rangle \\
\omega^{\omega+1} &= \{\langle 0, 0, \dots \rangle, \langle 0, 1, 1, \dots \rangle, \langle 0, 1, 2, 2, \dots \rangle, \dots, \langle 0, 1, 2, \dots, n, n, n, \dots \rangle, \dots\} \\
&= \bigcup_{n \in \omega} n^{\omega+1} = \{k^\omega : k \in \omega\}
\end{aligned}$$

etc.

- *The unfoldings of Aczel's self-singleton set* $\Omega = \{\Omega\}$: We have not postulated the existence of any non-wellfounded sets. However, we have extended the definition of unfoldings to classes and pointed graphs (or systems). One can now consider a pointed graph $G = (g, R)$ having only one node g (say $g = \emptyset$) which is its own and only successor: $R = \{(g, g)\}$. Consider the unfoldings G^α. These will be the same as the unfoldings of Aczel's self-singleton set $\Omega = \{\Omega\}$, if such a set would exist. So we can describe the unfoldings of Ω, regardless of the question of its existence:

$$
\begin{aligned}
\Omega^0 &= \emptyset \\
\Omega^1 &= \{\emptyset\} \\
\Omega^2 &= \{\{\emptyset\}\} \\
&\vdots \\
\Omega^\omega &= \langle \emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\{\{\emptyset\}\}\}, \ldots \rangle \\
\Omega^{\omega+1} &= \{\Omega^\omega\} \\
\Omega^{\omega+2} &= \{\{\Omega^\omega\}\} \\
&\vdots \\
\Omega^{2\omega} &= \langle \emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \ldots \Omega^\omega, \{\Omega^\omega\}, \{\{\Omega^\omega\}\}, \{\{\{\Omega^\omega\}\}\}, \ldots \rangle \\
&\vdots
\end{aligned}
$$

- *The set of all possible unfoldings of a given rank*: One can see, by recursion, that for every ordinal α, there is only a small set of distinct α-unfoldings. Actually, the collection of all possible unfoldings of rank α is a (small) set in V.

- *The unfoldings of the Universe*: As mentioned above, we extended the definition of unfoldings to classes. Consider now the class U of all sets. One can see that its unfolding of rank $\alpha+1$ is just the above-mentioned set of all possible unfoldings of rank α. One can show that we have

$$
\begin{aligned}
U^0 &= \emptyset \\
U^1 &= \{\emptyset\} \\
U^2 &= \{\emptyset, \{\emptyset\}\} \\
&\vdots \\
U^\lambda &= \langle U^\alpha \rangle_{\alpha<\lambda} \text{ , for limit ordinals } \lambda \\
U^{\lambda+1} &= \{a^\lambda : a \in U\} \text{ (the set of all unfoldings of rank } \lambda) \\
U^{\alpha+2} &= \mathcal{P}(U^{\alpha+1})
\end{aligned}
$$

From this, one can prove by induction that, for each α, the α-unfolding U^α of the whole universe is just a (hereditarily) small set in V. So from the point of view of its successive unfoldings, the universe is just another structure, which we can analyze and describe up to any degree of accuracy, using only small well-founded sets as descriptions.

One can easily check that unfoldings have the following properties:

1. If $a \in b$ then $a^\alpha \in b^{\alpha+1}$ for every α.

2. If $a \subseteq b$ then $a^{\alpha+1} \subseteq b^{\alpha+1}$ for every α.

In the next chapter we shall prove the *converses* of these statements, assuming the axioms of our system STS. Property 1 above and its converse show that, in our system, the above-defined notion of unfolding is well-fit for the job it was intended to do: it indeed *captures all there is to know about a set, namely its members.*

Definition 3.2.2 Two objects (sets, classes or pointed systems) are *observationally equivalent* if all their unfoldings coincide; e.g. for sets:

$$a \equiv b \qquad \text{iff} \qquad a^\alpha = b^\alpha \;\text{ for all ordinals }\; \alpha;$$

We can obtain a bisimulation-type characterization of observational equivalence if we first define a notion of *observational equivalence up to rank* α, for every ordinal α, as identity of unfoldings of rank α:

$$a \equiv_\alpha b \qquad \text{iff} \qquad a^\alpha = b^\alpha .$$

Observe that we have:

$$a \equiv b \qquad \text{iff} \qquad a \equiv_\alpha b \text{ for all ordinals } \alpha$$

Then one can easily check that we have the following equivalences:

Proposition 3.2.3

$a \equiv_{\alpha+1} b$ *iff:* *for every* $a' \in a$ *there exists some* $b' \in b$ *s. t.* $a' \equiv_\alpha b'$

and, dually,

for every $b' \in b$ *there exists some* $a' \in a$ *s. t.* $a' \equiv_\alpha b'$ *;*

$a \equiv_\lambda b$ *iff:* *for every* $\alpha < \lambda$ *we have* $a \equiv_\alpha b$*,*

for limit ordinals λ*.*

This is the promised bisimulation-type characterization. One can state this in a uniform manner by observing that the above two equivalences can be combined into one, which states that $a \equiv_\alpha b$ is equivalent to

$$\forall \beta < \alpha \left[\forall a' \in a \exists b' \in b \; a' \equiv_\beta b' \;\&\; \forall b' \in b \exists a' \in a \; a' \equiv_\beta b' \right].$$

In words : *two sets are α-equivalent if for every ordinal $\beta < \alpha$, every member of one of the sets is α-equivalent to some member of the other set. The sets are observationally equivalent if they are α-equivalent for every α.*

One can use this characterization to check that for graphs that are "small" (i.e. in V), observational equivalence is the same with bisimilarity. One can also see that if a set is hereditarily small, then its membership graph is isomorphic to a graph in V; so for small sets, observational equivalence will coincide with bisimilarity. In particular, the two notions coincide in Aczel's universe.

3.2.2 Modal sentences as partial descriptions

Let us analyze more carefully the above-mentioned process of unfolding set structures. As defined above, unfoldings of rank α are partial descriptions of the given set, which are nevertheless relatively *maximal* from an informational point of view: they gather all the information that is available at stage α about a set and its members. In other words, they are designed to provide the best possible description of the set a that is available at stage α.

We want now to decompose the maximal descriptions given by unfoldings into simpler, more basic partial descriptions. Observe for instance that an unfolding gives a description for *every* member of the set; but it might happen we only know the description of some of its members. Also, at some stage, we might not know the real unfolding of a set, but only be able to narrow down the possibilities to a set of possible (alternative) unfoldings.

I call every such permissible partial description an *analytical description* or (anticipating things a bit) a *modal sentence.* Not every definable property is an analytical description. The main intuition is that we are never given the full structure, but we can only know its successive unfoldings. So a general guideline for constructing new partial descriptions is that they should be "bounded" by the unfolding process; they should not assume we know the full structure or the real identity of the object (set) or of its members, but should use only information that is available at some definite stage of unfolding. The adequacy of an analytical description should be decidable during the unfolding process: given a set a and a description φ, there must exist a stage α such that we can check whether or not φ does actually describe a by looking at the unfolding of rank α of the set a.

In other words, analytical descriptions should be "bounded" by unfoldings: every such description should be informationally weaker than some unfolding. We can only observe sets and structures through their unfoldings, and analytical descriptions should be only based on such "observations". This boundedness condition rules out languages like first-order logic, since the truth-value of sentences of the form $a = b$ or $a \in b$ depends on the *whole* $\in$-structure of the sets a, b. There is no obvious way to directly check at some definite ordinal stage a description of a set x saying that $x = a$; the same goes for descriptions like: $x \in a$, or $x \in x$, or "x has two distinct members". All these assume a full knowledge of the identity of x, that is of its structure, or a full knowledge of the structures of its members; this is because, as we have seen, the real identity of sets is not given beforehand, but is determined by their full structure. It may of course happen that for particular sets we can decide these properties only on the basis of some unfolding (which means that in these cases the property will be equivalent to some analytical description).

To make things precise, we introduce the following

Definition 3.2.4 Let P be a unary predicate (i.e. a class P, for which we write $P(a)$ instead of $a \in P$). P is an *analytical description* if there exists some ordinal stage α, such that one can decide whether or not P holds by looking at the α^{th} unfolding; that is, we have:

$$\forall a, b(a^\alpha = b^\alpha \Longrightarrow (Pa \leftrightarrow Pb)).$$

We also say that in this case the description P is *bounded by* α.

A careful analysis of analytical descriptions shows that they can all be generated by the following three operations:

(1) Negation: Given a possible description φ and an object a, we construct a new description $\neg\varphi$, to capture the information that φ does not describe a.

(2) Conjunction: Given a set Φ of descriptions of the object a, we accumulate all descriptions in Φ by forming their conjunction $\bigwedge \Phi$.

(3) Diamond: Given a description φ of some member (or members) of a set a, we construct the description $\Diamond\varphi$, which captures the information that a has a member described by φ.

Observe that the first two operations refer to sets-as-objects and generate the language of *infinitary propositional logic*. The third is the most basic operation involved in partial unfolding: we just unwrap the box and pick up (the description of) some thing inside. The language L_∞ generated by these three operations is called *the infinitary modal logic*. One can consider this logic as a fragment of $L_{\infty\omega}$, the standard first-order language with infinitary conjunctions and disjunctions. But it is easy to see that, unlike the sentences of the full $L_{\infty\omega}$, any sentence of the infinitary modal logic L_∞ is bounded by some ordinal, and so it is an analytical description.

One can similarly define these modal descriptions for pointed systems (and graphs). The resulting description relation coincides with the Kripke semantics for infinitary modal logic. We call this relation *satisfaction* and we define it recursively in the familiar way for Kripke structures. For *sets*, the corresponding recursive clauses correspond to the *set-semantics for modal logic*, defined in the more restricted case of ZFA by J. Barwise and L. Moss in *Vicious Circles*:

$$
\begin{array}{lll}
a \models \neg\varphi & \text{iff} & a \not\models \varphi \\
a \models \bigwedge \Phi & \text{iff} & a \models \varphi \text{ for all } \varphi \in \Phi \\
a \models \Diamond\varphi & \text{iff} & a' \models \varphi \text{ for some } a' \in a.
\end{array}
$$

We can also define the dual operators $\bigvee$, $\Box$ and some other useful operators:

$$
\begin{aligned}
\bigvee \Phi &=: \neg \bigwedge \{\neg\varphi : \varphi \in \Phi\} \\
\Box\varphi &=: \neg\Diamond\neg\varphi \\
\Diamond\Phi &=: \{\Diamond\varphi : \varphi \in \Phi\} \\
\Box\Phi &=: \{\Box\varphi : \varphi \in \Phi\} \\
\varphi \wedge \psi &=: \bigwedge\{\varphi, \psi\} \\
\varphi \vee \psi &=: \bigvee\{\varphi, \psi\} \\
\triangle\Phi &=: \bigwedge \Diamond\Phi \wedge \Box \bigvee \Phi.
\end{aligned}
$$

The last operator will be useful in constructing modal sentences that capture the above notion of unfolding.

Definition 3.2.5 A *sound description of a pointed binary structure,* or set, is just a modal sentence satisfied by that structure. A modal sentence is said to be *consistent* if it is a sound description of some structure, i.e. is satisfied by some pointed graph. A *modal theory* is a collection of (infinitary) modal sentences. The *(modal) theory of a structure* is the collection of all the modal sentences satisfied by that structure.

As expected, the information captured by our earlier notions of unfolding and maximal description of level α can be now expressed by certain modal sentences:

Definition 3.2.6 Define, for each set a and each ordinal α, a modal sentence

φ_a^α, by recursion on ordinals :

$$\begin{aligned}\varphi_a^{\alpha+1} &=: \triangle\{\varphi_b^\alpha : b \in a\} \\ \varphi_a^\lambda &=: \bigwedge\{\varphi_a^\beta : \beta < \alpha\}.\end{aligned}$$

Proposition 3.2.7

$$b \models \varphi_a^\alpha \textit{ iff } b^\alpha = a^\alpha.$$

In particular, we always have $a \models \varphi_a^\alpha$.

Proof: Easy induction on α. ■

This shows that our modal sentences φ_a^α capture indeed the same information as the corresponding unfoldings.

As a consequence, the relations of observational equivalence $\equiv$ and of equivalence $\equiv_\alpha$ up to rank α, as defined in the previous section, are easily expressible in terms of modal sentences:

$$\begin{aligned} a \equiv_\alpha b &\quad \text{iff} \quad \varphi_a^\alpha = \varphi_b^\alpha \\ a \equiv b &\quad \text{iff} \quad a, b \text{ satisfy the same modal sentences}\end{aligned}$$

So observational equivalence, as defined in the previous section, coincides with *modal equivalence*. So we can now give an alternative definition for observational equivalence:

Definition 3.2.8 Two sets, or structures, are *observationally equivalent* if they have the same modal descriptions, i.e. if their modal theories coincide.

One can also prove that the above notion of description is *maximal: infinitary modal language is the largest language which is bounded by unfoldings.* More precisely, one can prove:

Proposition 3.2.9 *Let P be an analytical description, bounded by some ordinal α. Then the predicate P is "essentially modal", i.e. it is logically equivalent to some modal sentence φ_P.*

Proof: Take

$$\varphi_P =: \bigvee\{\varphi_a^\alpha : P(a)\}.$$

We need to prove that for every set b, we have $P(b) \leftrightarrow b \models \varphi_P$.

For one direction, suppose that we have $P(b)$. Then $b \in \{a : P(a)\}$ and we also know that $b \models \varphi_b^\alpha$. These imply that $b \models \bigvee\{\varphi_a^\alpha : P(a)\}$, i.e. $b \models \varphi_P$.

For the other direction, suppose we have $b \models \varphi_P$. By the choice of the sentence φ_P, this means that there exists some a such that we have $P(a)$ and also $b \models \varphi_a^\alpha$. Then, by the above Proposition, it follows that $b^\alpha = a^\alpha$. From this, together with $P(a)$ and with the fact that P is bounded by α, we conclude that we have $P(b)$. ■

3.3 The Super-Antifoundation Axiom $SAFA$

The essence of the structural conception of set can be now captured by the following :

> *Sets are just arbitrary pointed systems modulo observational equivalence.*

In the system STS presented in the next chapter, this is expressed by the following AFA-like theorem:

Weak SAFA: *Every pointed system is observationally equivalent to a unique set.*

This was our initial formulation of our main axiom $SAFA$ ("Super-Antifoundation Axiom"), designed to replace AFA in the context of a theory of arbitrarily-sized strongly extensional classes. We shall call this statement *Weak* $SAFA$, since it is weaker than the final formulation of $SAFA$ (to be presented in the next section). The existential half of Weak $SAFA$ ("*every pointed system is equivalent to some set*") is a strengthening of the existential AFA. The uniqueness half is a strengthening of strong extensionality, which can be called Super-Strong Extensionality: *observationally equivalent sets are identical.*

Stated explicitly in terms of unfoldings, the meaning of Weak $SAFA$ is this:

> *Given any structure (pointed system) S, the transfinite pattern (sequence) of unfoldings of S can be always seen as the pattern of unfolding of some* set*; and this pattern of unfolding uniquely determines the set.*

There also exists an explicit formulation of Weak $SAFA$ in terms of modal theories. First, recall that a modal sentence is *consistent* iff it is satisfied by some pointed graph. Observe now that the existential half of Aczel's AFA is equivalent to the statement that *every consistent modal sentence is satisfied by some set.*

To generalize this formulation, we need the following

Definition 3.3.1 A modal theory is *consistent* if all its members are satisfied by a single structure (pointed system). The *total analytical pattern of a structure* is nothing but its modal theory, i.e. the collection of all its sound descriptions.

One can see that the total analytical pattern of every set is a consistent modal theory, and that moreover it is *maximally consistent*. Then one can check that the converse of this statement is equivalent to the above-mentioned formulation of *Weak* $SAFA$:

Proposition 3.3.2 *Weak* $SAFA$ is equivalent to the assertion that:

> *Every maximally consistent modal theory is satisfied by a unique set.*

This is a nice way to see that *Weak* $SAFA$ is indeed a strengthening of AFA. This formulation is strong enough for a lot of purposes. In particular, its uniqueness half is enough to give sets a clear structural identity. Its existential half is also enough to define almost all interesting sets in our theory (e.g. the universal set U), and to prove the main closure properties: the universe of sets U can be thus seen to be closed under power-set, arbitrary intersections, unions of small families of sets, pairs, Cartesian products etc. But one cannot prove the Union axiom: if a is some set, then $\bigcup a$ cannot be proved to be a set.

We shall see this failure is not due to some incompleteness or lack of closure of the universe of *sets*, but rather with the "incompleteness" of our universe of definable *classes*. The deep reason is the lack of strong class-forming principles in all theories which reject (as does ours, as well as ZFC and ZFA) arbitrary classes-as-objects, accepting them only as ways-of-speaking about unary formulas. Observe that the above definition of consistency for modal theories involves a quantifier over pointed systems. In practice, a consistent modal theory will be given together with a pointed system satisfying it, so there will not be any problems in stating properties that quantify universally over consistent modal theories. But *there will be problems* when we will have to prove the *existence* of some consistent theory. It might happen that we cannot prove the consistency of the theory, because we cannot define a corresponding pointed system, despite the fact that we can "see" it is "locally consistent"; indeed, we can express the theory as a sequence of partial unfoldings (modal sentences), which might not be strong enough to allow us to put our hands on a pointed system, but could still be seen to be consistent up to any ordinal stage of unfolding. The AFA-like theorem mentioned above would not apply to it. Such an analytical pattern would give a potentially consistent description of a possible structure, but it would remain unrealized both in the realm of sets and in the one of pointed systems. The reason is that *we cannot show it to be consistent*: we cannot prove the existence of a system that satisfies it. This is exactly what happens when we try to prove the existence of the union-set $\bigcup a$.

The intention behind Weak $SAFA$ was to generalize the existential half of AFA, seen as a *maximality principle*. We wanted to ensure that the universe is as large as it is consistent with strong extensionality, by postulating that every possible structural pattern is realized by a set. But we can now see that, if we limit ourselves to descriptions that are maximal (like the unfoldings and the total analytical pattern), we lose some of the possibilities. The main advantage of introducing modal sentences as partial descriptions instead of the unfoldings is that we can recover these lost possibilities.

But, for this, we need an alternative definition of consistency for modal theories, one that would not involve quantification over pointed systems.

Definition 3.3.3 A modal theory is *weakly consistent* if all its sub*sets* are consistent. More generally, a *partial analytical pattern* is any collection of sound descriptions *which is closed under infinitary conjunction.*

The last condition is needed if we want to think of our collection as encoding the knowledge we might have about a particular structure at a certain moment: for, given a set of sound partial descriptions of the same object, we "know" that their conjunction is also a sound description of that object. As a consequence, *every partial analytical pattern is weakly consistent*, though it might not be provably consistent, because of the above-mentioned weakness of our theory of classes. The price to be paid for not having classes-as-objects but only as predicates is that we can sometimes "see" from inside the theory that a certain pattern is possible (weakly consistent) and so it describes a possible structure, without being able to define a structure that realizes it inside the theory. The assumption that every weakly consistent theory is consistent is equivalent to a "strong infinity axiom" (namely to the hypothesis that the cardinal of the universe is *weakly compact*).

But denying sethood to weakly consistent structures is a restriction to the structure paradigm, which comes again as an artifact of the way we formalize pointed systems (as classes) inside set theory. There is no reason to throw away collections of descriptions that we can easily see to be possible patterns of analytical behavior. The spirit of the structural conception points toward a maximal notion of set, and hence to the following strong version of the (existential half of) Super-Antifoundation Axiom:

Every partial analytical pattern describes a set.

This formulation is stronger than the previous one (in terms of unfoldings), since it "corrects" the above-mentioned lack of strong class-comprehension principles. Assuming this postulate one can actually *prove* the mentioned strong infinity axiom.

The uniqueness half of $SAFA$ (super-strong extensionality) can be expressed now as: *sets are uniquely determined by their total analytical pattern*). It is easy to check that this is equivalent to the previous formulation (in terms of unfolding) and also with *the last formulation* of strong extensionality. Putting the two halves together, one obtains our central postulate $SAFA$, which embodies our set-theoretical philosophy : loosely speaking, *a set is any possible pattern of observable behavior under iterative unfolding.*

Chapter 4

Axiomatic Theory and Developments

4.1 The System STS

I present an axiomatic system STS (*Structural Theory of Sets).* I shall use the standard language of Set Theory, with *variables for sets.* I shall also use *classes* as a "manner of speaking" extensionally about unary formulas. The *class of all sets* (defined by the formula $x = x$) is denoted by U. I also postulate the existence of two classes V and Sat. The first is intended to be the class of all wellfounded sets and so a model of ZFC; the second is the relation of *satisfaction* of a modal formula by a set.

I shall assume that the class U of all sets is an extensional universe satisfying some very mild closure conditions (closure under singletons and finite unions). As far as I know, the existence of the sets formed by these operations has never been doubted.

The class V is needed for constructing descriptions to record the results of the analytic process. As we have seen in the previous chapter, we need to construct a synthetic-iterative universe V in order to be able to talk about large sets and structures. The wellfounded sets in V will play the role of partial descriptions for sets, and the ordinals (defined inside V in the usual way) will play the role of *stages (levels) of description*. So we think of the wellfounded elements of V as "linguistic objects", that we successively build in a quasi-constructive manner, and use in exploring the sets of the "real world". But at the same time the elements of V are themselves parts of the real world.

The intuitive picture for V is the basic iterative picture: we think of V as being built up from below, in transfinitely many stages; at each stage, we form all the *sets* made out of previously constructed elements. At limit stages we accumulate the results of our set-formation. We keep doing this "as long as we can", i.e. until we have formed all the sets that can be formed in this way. It is natural to assume this will take a long time, i.e. the entire process of set-formation is "large". This can be understood in at least two ways. First, the process is infinite, so that we can form the set of natural numbers at some stage; this justifies the axiom of Infinity. Secondly, the process is larger than any of its outputs, so that there are more stages than there are members in any set formed in this way; to put it the other way around: sets in V are "small", compared to the length of the formation process . This justifies Replacement.

The resulting class V will have to be transitive and closed under set-formation: a set will be in V if and only if its members are in V. Notice that we do not claim this process can be actually carried out in full: V is just a class, i.e. is simply a way of describing this process of formation.

Notice also that we *do not assume we can form power-sets* inside V. We just accumulate in the class V all the objects that can be formed as sets of previously-formed objects. Nothing in the iterative picture warrants that the collection of all the subsets of a set can be itself unified into a set. This makes more intuitive the above "smallness" assumption.

I should mention though that, as a consequence of our axiom $SAFA$, the Power-Set axiom *will* in fact be valid. $SAFA$ does imply that the power set can be unified into a set, but this has nothing to do with V, with the iterative conception or with limitation of size: the sets postulated by $SAFA$ can be both non-wellfounded and large. In fact, if we define (as von Neumann did) "smallness of a set" as "being smaller than the universe" (where by universe we mean U), then it is not true that the power-set of a small set is itself small: as we shall see, there exist sets smaller than U whose power-sets are as large as the universe. But once, by $SAFA$, every power-set is accepted as a set, the class V will also satisfy the Power Set axiom, because V is closed under set formation. (In fact, V will be proved to be a model of ZFC^-.) If we choose to define smallness as "smaller than some set in V" (or, in von Neumann style, as "smaller than V itself") then it follows that *the power-set of a small set is also small.* But this rather non-intuitive claim is now a *theorem*, not an axiom. As in the case of Cantor's theorem on the nondenumerability of the continuum, we just accept the existence of the power set as a provable fact, and not because of its self-evidence. This solves the above-mentioned "power-set mystery": *SAFA provides us with a separate justification for power sets.*

Basic Axioms:

(A1). Extensionality

(A2). Closure of the universe U under singletons and finite unions: if a, b are sets then $\{a\}, a \cup b$ are sets.

(A3). A set is in V iff all its elements are in V. In other words, the class V is transitive and closed under subsets: $\mathcal{P}V = V$.

(A4). V is a model of the axioms of: *Infinity, Replacement, Union and Choice.*

We define smallness, systems and graphs, modal sentences and satisfaction:

Definition 4.1.1 A set is *small* if it is of the same size as some set in V

Definition 4.1.2 A *(pointed) system* is (a pair of a "top" set and) a class of pairs (i.e. a binary relation).

Definition 4.1.3 A *(pointed) graph* is a small (pointed) system. A *(pointed) V-graph* is a (pointed) graph that belongs to V.

In the rest of this dissertation, we assume that we are given some encoding of the symbols $\neg, \bigwedge$ and $\Diamond$ as well-founded sets in V. We also use $\neg\varphi$ as abbreviation for the pair $(\neg, \varphi)$, and similarly for the other modal formulas. (Technically, to make this work, we have to assume that none of the basic connectives is encoded as a pair.)

Definition 4.1.4 Define, by recursion inside V, the class L_∞ of *infinitary modal sentences* as the least subclass of V such that:

1. if $\varphi \in L_\infty$ then $(\neg\varphi) \in L_\infty$ and $(\Diamond\varphi) \in L_\infty$
2. if $\Phi \subseteq L_\infty$ is a *set* of modal sentences then $(\bigwedge \Phi) \in L_\infty$.

We define the other modal connectives $\vee, \rightarrow, \top, \bot, \Box$ by the usual abbreviations.

Definition 4.1.5 We define, by recursion in V, *satisfaction for modal formulas on pointed V-graphs $S \models \varphi$*:

$$\begin{array}{lll} (g, R) \models \neg\varphi & \text{if} & (g, R) \not\models \varphi \\ (g, R) \models \bigwedge \Phi & \text{if} & (g, R) \models \varphi \text{ for all } \varphi \in \Phi \\ (g, R) \models \Diamond\varphi & \text{if} & (g', R) \models \varphi \text{ for some } g' \text{ s.t. } gRg'. \end{array}$$

Intuitively, we would like to define satisfaction on a *set*, by considering its $\in$-structure as a pointed graph. But, for large sets, their $\in$-structures will not be graphs in V, so we cannot use recursion in V to define satisfaction on an *arbitrary set.* To do this, we would need a principle of recursion of the type mentioned in the previous chapter. Instead of this, we shall just postulate the corresponding inductive conditions as axioms about the undefined predicate Sat. In the following, we use the notation $a \models \varphi$ for $(a, \varphi) \in Sat$.

Satisfaction Axioms:

(A5). $a \models \neg\varphi$ iff $a \not\models \varphi$

(A6). $a \models \bigwedge \Phi$ iff $a \models \varphi$ for all $\varphi \in \Phi$

(A7). $a \models \Diamond\varphi$ iff $a' \models \varphi$ for some $a' \in a$.

We can now define *observational equivalence* $\equiv$ as modal equivalence: two pointed systems are observationally equivalent if they have the same descriptions, i.e. they satisfy the same infinitary modal sentences. (Technically, we first define it for objects, i.e. sets and pointed graphs, and then extend it to pointed systems and classes.) All the notions defined in the section on modal descriptions, in particular the relations $\equiv_\alpha$, the notions of sound description, partial and total analytical pattern, the modal theory of a set $th(a) = \{\varphi : a \models \varphi\}$, can be now rigorously defined, and the results mentioned in the previous section can be proved.

The most straightforward way to rigorously define the above notions of observational equivalence up to rank α is not by the recursive definition from the previous section, but in a more indirect manner. This is because we have not yet proved a Recursion Theorem for functions having the whole universe as domain; this will have to wait, since it will make essential use of our axiom $SAFA$. But one can still use recursion inside V to define the following notions:

Definition 4.1.6 The *degree* of a modal sentence $\varphi \in L_\infty$ is an ordinal $deg(\varphi) \in On$, defined by recursion in V:

$$\begin{aligned} deg(\bigwedge \Phi) &= sup\{deg(\varphi) : \varphi \in \Phi\} \\ deg(\neg\varphi) &= deg(\varphi) \\ deg(\Diamond\varphi) &= deg(\varphi) + 1. \end{aligned}$$

For each ordinal α, the *observational equivalence up to rank* α is a relation $\equiv_\alpha$, defined by:

$a \equiv_\alpha b$ iff a and b satisfy the same modal sentences of degree $\leq \alpha$.

Then one can prove that these equivalence relations satisfy the recursive "bisimulation"-type condition mentioned in the previous section: namely, $a \equiv_\alpha b$ is equivalent to

$$\forall\beta < \alpha\, [\forall a' \in a \exists b' \in b\ a' \equiv_\beta b'\ \&\ \forall b' \in b \exists a' \in a\ a' \equiv_\beta b'] \,.$$

We cannot yet define the above notions of *unfolding* and description of a set in full generality (again because we do not have a general Recursion Theorem yet), but we can only define them for sets in V (by the very definition from the previous section, which can be made into a recursive definition for sets $a \in V$).

We can state now the main axiom of the system:

(A8). Super-Anti-Foundation Axiom ($SAFA$):

(i). *Existence (Structural Completeness)*: Every partial analytical pattern describes a set.

(ii). *Uniqueness (Super-Strong Extensionality):* Sets are uniquely determined by their total analytical pattern, i.e. if $a \equiv b$ then $a = b$.

4.2 Correspondence between Sets and Modal Theories

Recall that a *weakly consistent theory* is a modal theory (i.e. a class of modal sentences) such that all its subsets are satisfiable on pointed graphs. A theory is *complete* if for every modal sentence φ, the theory contains φ or its negation. A theory is *maximally weakly consistent* if it is weakly consistent and has no weakly consistent proper extension. One can check that the maximally weakly consistent theories coincide with the complete weakly consistent theories.

The existential half of $SAFA$ is equivalent to the following "compactness"-type statement:

> *Every weakly consistent theory T is satisfied by some set a, i.e.*
> $T \subseteq th(a)$.

But it is easy to see that every theory of the form $th(a)$ is maximally weakly complete. So, by the above statement and by maximality, we conclude that *the maximally weakly consistent theories are exactly the ones of the form* $th(a)$ *for some set* a.

On the other hand, super-strong extensionality says that *sets are characterized by their modal theories*: if $th(a) = th(b)$ then $a = b$. So the function th that associates to each set a its modal theory $th(a)$ is a one-to-one correspondence. This is a generalization of a theorem about sets in ZFA, proved by Barwise and Moss: a set in Aczel's hyperuniverse can be characterized by a single modal sentence.

Putting these together, we can see that, taken as a whole, $SAFA$ implies the following statement:

> *The function* th *gives a bijective correspondence between sets and maximally weakly consistent modal theories.*

The last statement is weaker than $SAFA$ and is actually equivalent to what we called "weak $SAFA$" in the previous chapter: the claim that every pointed system is observationally equivalent to a set.

It is useful to know there exists a modal system $\mathbf{K}_\infty$ for infinitary modal logic, which is *sound* for pointed systems (and graphs) and *weakly complete* for pointed graphs. One can show that, in the context of the other axioms of our system, the consistent theories of this proof system are precisely the weakly consistent theories.

So we conclude that th is a bijective correspondence between sets and maximally consistent theories in $\mathbf{K}_\infty$. One can make this bijection into an isomorphism, by defining an accessibility relation between theories, as in the canonical model construction in standard modal logic:

Definition 4.2.1 For T, S maximally weakly consistent modal theories, put

$$T \to S \quad \text{iff} \quad \forall\varphi(\varphi \in S \Rightarrow \Diamond\varphi \in T)$$
$$\text{iff} \quad \forall\varphi(\Box\varphi \in T \Rightarrow \varphi \in S)\ .$$

One can easily check now that we have the following

Proposition 4.2.2 *For all sets a, b:*

$$a \in b \qquad \textit{iff} \qquad th(b) \to th(a)\ ,$$

i.e. the above-mentioned bijection is an isomorphism between the $\in$-structure of the universe of sets and the accessibility structure between maximally weakly consistent modal theories.

Proof:

For one direction, assume $a \in b$. To show that $th(b) \to th(a)$, let $\varphi \in th(a)$. Then $a \models \varphi$, so $b \models \Diamond\varphi$, i.e. $\Diamond\varphi \in th(b)$.

For the other direction, assume $th(b) \to th(a)$ and let set $b' = b \cup \{a\}$. To show $a \in b$, we need to prove that $b = b'$. By Super-Strong Extensionality, it is enough to show that b and b' satisfy the same modal sentences. We prove this by induction on the complexity of sentences. The only non-trivial case is the induction step for sentences of the form $\Diamond\varphi$:

Suppose that $b \models \Diamond\varphi$. Then there is some $c \in b$ such that $c \models \varphi$. But $b \subseteq b'$, so we have $c \in b'$. Hence $b' \models \Diamond\varphi$.

For the converse, suppose that $b' \models \Diamond\varphi$. Then there is some $c' \in b'$ such that $c' \models \varphi$. If $c' \in b$, then we obtain that $b \models \Diamond\varphi$, as desired. If $c' \in b' \setminus b$, then we must have $c' = a$, and so $a \models \varphi$, i.e. $\varphi \in th(a)$. But we assumed $th(b) \to th(a)$, which by definition implies that $\Diamond\varphi \in th(b)$. Hence $b \models \Diamond\varphi$. ■

So one could say that *a set is just a maximally (weakly) consistent theory in infinitary modal logic.* This gives the idea for proving the consistency of our system STS: working in ZFC with an appropriate large cardinal assumption (the existence of a weakly compact cardinal κ), we shall define the modal logic L_κ having only conjunctions of size less than κ, construct its canonical model and interpret it as a universe of sets, with membership defined by the accessibility relation. This construction gives a model of the system STS.

4.3 The Canonical Model of STS

We shall sketch here the proof of consistency of our system STS. We only give here the construction of the model, the main definitions and lemmas and sketch the proof of the validity of our axioms for this model.

As mentioned in the previous section, we work in ZFC with an *extra-assumption*, namely that of the *existence of an infinite weakly compact cardinal* κ. For our purposes, it is convenient to *define weak compactness* by the following:

Definition 4.3.1 Consider, for each cardinal κ, the first-order language $L_{\kappa\omega}$ with infinitary conjunctions and disjunctions of length strictly less than κ.

We say that $L_{\kappa\omega}$ is *strongly incompact* if there exists some signature μ and some set Γ of sentences of $L_{\kappa\omega}(\mu)$ with the following properties:

- Γ has cardinal power κ
- Γ has no model
- Every subset of Γ of power $< \kappa$ has a model.

If this is not the case, we say that $L_{\kappa\omega}$ is *weakly compact*.

A cardinal κ is said to be *strongly inaccessible* if the corresponding universe V_κ in the iterative hierarchy is a model of ZFC. The cardinal κ is said to be *weakly compact* if it is weakly inaccessible and $L\kappa\omega$ is weakly compact.

It is well-known that the above definition is equivalent to the more commonly used definition of weak compactness in terms of trees. The hypothesis of the existence of infinite weakly compact cardinal is known to be unprovable in ZFC, but is nevertheless generally considered by most set-theorists to be *consistent* with ZFC. Actually, the existence of infinite weakly compact cardinals is nowadays seen as a rather mild large cardinal assumption. Throughout this section, we shall assume this hypothesis.

Historical Note: To the best of our knowledge, the first to have considered this assumption to construct a model for a universal set theory satisfying Positive Comprehension was Weydert [Weydert1989]. As already mentioned, Forti and Hinnion [Forti1989] have succeeded in proving the consistency of the Generalized Comprehension Principle, in the assumption of the existence of an infinite weakly compact cardinal. The model in [Foti1989] can be proved to be isomorphic to the "canonical model" construction given below.

Let now κ be a (fixed) infinite weakly compact cardinal. We consider now the *modal logic* L_κ, defined exactly as infinitary modal logic $L\infty$, but *with conjunctions and disjunctions restricted to sets of sentences of size strictly smaller than* κ. Clearly, L_κ is a subfragment of $L_{\kappa\omega}[\mu]$, where $\mu = R$ is the signature consisting of one binary relation symbol R. By the strong inaccessibility of κ, the set L_κ of all these modal sentences has size κ. A *theory* of L_κ is a set $\Phi \subseteq L_\kappa$.

We define *satisfaction* of a modal sentence or theory by a pointed graph (g, R) in the usual way. We can also define *satisfaction for sets a (in our ZFC-universe)* as in STS, by considering the pointed graphs given by their membership structures. We use the notations:

$$\begin{aligned} th_\kappa(g, R) &= \{\varphi \in L_\kappa : (g, R) \models \varphi\} \\ th_\kappa(a) &= \{\varphi \in L_\kappa : a \models \varphi\} \end{aligned}$$

A theory Φ is said to be *consistent* if it is satisfied by some pointed graph. The theory Φ of L_κ is *weakly consistent* if every subset of Φ of power $< \kappa$ is consistent. It is clear that the weak compactness of κ implies the following

Lemma 4.3.2 *Every weakly consistent theory of L_κ is consistent.*

A theory is said to be *maximally consistent* if it is consistent and has no consistent proper extension. It is easy to see that:

Lemma 4.3.3 *A theory Φ of L_κ is maximally consistent iff it is of the form $th(g,R)$, for some pointed graph (g,R).*

Proof: By the consistency of Φ, there exists some (g,R) satisfying it, i.e. $\Phi \subseteq th(g,R)$. But $th(g,R)$ is also consistent (being satisfied by (g,R)); by the maximality of Φ, we have $\Phi = th(g,R)$. The converse is trivial. ■

From this we can easily prove that:

Lemma 4.3.4 *Every weakly consistent theory of $L\kappa$ is included in some maximally weakly consistent theory.*

Proof: Let Φ be weakly consistent; by our first lemma above, Φ must be consistent. Let (g,R) some pointed graph satisfying Φ. Then $\Phi \subseteq th(g,R)$, and $th(g,R)$ is maximally consistent by the previous lemma. ■

Notations: We put $Th_\kappa =: \{\Phi \subseteq L_\kappa : \Phi \text{ is maximally consistent } \}$. We denote the elements of Th_κ by letters $T, S \ldots$, to distinguish them from arbitrary theories (denoted by $\Phi, \Psi \ldots$).

Definition 4.3.5 For $T, S \in Th_\kappa$, we put (as in the case of L_∞)

$$
\begin{aligned}
T \to S \quad &\text{iff} \quad \forall\varphi(\varphi \in S \Rightarrow \Diamond\varphi \in T) \\
&\text{iff} \quad \forall\varphi(\Box\varphi \in T \Rightarrow \varphi \in S)\ .
\end{aligned}
$$

The graph $\mathbf{M} =: (Th_\kappa, \to)$ is an infinitary analogue of the *canonical model construction for modal logic.* As usual, we write $T \models \varphi$ when we have $(T, \to) \models \varphi$.

It is easy now to check that

Lemma 4.3.6 *For $T \in Th\kappa$, $\varphi \in L_\kappa$, and $\Phi \subseteq L_\infty$ having size $card(\Phi) < \kappa$, we have the following:*

1. *$(\neg\varphi) \in T$ iff $\varphi \notin T$*

2. *$\bigwedge \Phi \in T$ iff $\Phi \subseteq T$*

3. *$\Diamond\varphi \in T$ iff there exists $S \in Th_\kappa$ s.t. $T \to S$ and $\varphi \in S$*

Proof: Induction on the complexity of sentences:

The steps for negation and conjunction are trivial.

For $\Diamond$, in one direction: suppose that $\Diamond\varphi \in T$. Let Φ be a theory given by:

$$\Phi =: \{\psi : \Box\psi \in T\} \cup \{\varphi\}.$$

It is easy to check that Φ is weakly consistent: if $\Psi \subseteq \Phi$ is a subtheory of size $< \kappa$, then Ψ can be seen to be equivalent either to a sentence of the form ψ or to a sentence of the form $\psi \wedge \varphi$, where $\Box\psi \in T$. (This is because T is maximally consistent, so it is closed under conjunction, and because, in general, $\Box \bigwedge \Theta$ is equivalent to $\bigwedge\{\Box\theta : \theta \in \Theta\}$.) If Ψ were inconsistent then one of the above two sentences would be inconsistent. In both cases, the stronger of the two sentences, namely $\psi \wedge \varphi$, would be inconsistent. Then $\Diamond(\psi \wedge \varphi)$ would also be inconsistent. But it is easy to see that $\Diamond(\psi \wedge \varphi) \in T$ (using modal reasoning, the fact that T is maximally consistent and the facts that $\Box\psi \in T$ and $\Diamond\varphi \in T$). So it follows that T would be inconsistent. Contradiction.

This shows that every such Ψ is consistent. So Φ is weakly consistent. By the previous lemma, Φ is included in some maximally consistent theory S. One can easily check now that we have $T \to S$ and $\varphi \in S$.

The converse direction is trivial. ■

As an easy corollary we obtain the

Lemma 4.3.7 ***(Truth Lemma for Canonical Model)***

For every $T \in Th_\kappa$ and every $\varphi \in L_\kappa$, we have:

$$T \models \varphi \text{ iff } \varphi \in T.$$

Proof: : Easy induction on the complexity of sentences, using the previous lemma. ■

The canonical model $\mathbf{M} = (Th_\kappa, \to)$ offers a natural interpretation to the language of set theory, in which the universe is taken to be the set Th_κ and the membership relation is interpreted as the converse of $\to$. We want to prove that this is a model of the axioms of STS. For this, we first need to give an interpretation to the basic classes (formulas) V and Sat of the system STS.

We put

$$\begin{aligned} V^{\mathbf{M}} &=: \{th_\kappa(x) : x \in V_\kappa\} \\ Sat^{\mathbf{M}} &=: \{(T, \varphi) : T \in Th_\kappa, \varphi \in T\} \end{aligned}$$

Now we can state the main theorem of this section:

Theorem 4.3.8 $(\mathbf{M}, V^{\mathbf{M}}, Sat^{\mathbf{M}})$ *is a model of the axioms of STS. As a consequence, STS is consistent, if we assume the axioms of ZFC and the existence of an infinite weakly compact cardinal.*

Proof: (sketch)

- *Extensionality* is easy to check: suppose that we have $T, T' \in Th_\kappa$ s.t. $\forall S \in Th_\kappa(T \to S \Leftrightarrow T' \to S)$. By the definition of $\to$, this implies that we have $\{\varphi : \Diamond\varphi \in T\} = \{\varphi : \Box\varphi \in T'\}$. This provides the essential step (for $\Diamond$) for proving that $\forall\varphi(\varphi \in T \Leftrightarrow \varphi \in T')$, by induction on the complexity of φ; the other two steps, for negation and conjunction, are trivial.

- *Closure of the Universe under pairing and binary union*: Let $T, S \in Th_\kappa$. Check that the theories

$$\Phi =: \{\Diamond\varphi : \varphi \in T \cup S\} \cup \{\Box(\varphi \vee \psi) : \varphi \in T, \psi \in S$$

and

$$\Psi =: \{\Diamond\varphi : \Diamond\varphi \in T \cup S\} \cup \{\Box\varphi : \Box\varphi \in T \cap S\}$$

are weakly consistent. Then, check that the unique maximal consistent theory $\{T, S\}^{\mathbf{M}}$ which extends Φ formally satisfies (in the model $\mathbf{M}$ the defining property of the unordered pair; similarly, the maximally consistent theory extending Ψ satisfies the defining property of the "$\mathbf{M}$-union" of the sets corresponding to T and S.

- *The properties of $V^{\mathbf{M}}$:*

We shall use the following internalization of (a special case of) the Barwise-Moss Modal Characterization Theorem for sets in V_{AFA}:

Every set $a \in V_\kappa$ is characterized by some sentence $\theta_a \in L_\kappa$.

This can be obtained by first restricting the above-mentioned theorem to the well-founded sets in V, and then by applying reflection (using the fact that κ is an accessible cardinal, so that $V_\kappa \models ZFC$.

It is easy to check now that $(V^{\mathbf{M}}, \rightarrow)$ *is isomorphic to* $(V_\kappa, \ni)$: the isomorphism is given by the restriction of the function th_κ to V_κ, which is onto definition (since $V^{\mathbf{M}}$ was defined as the range of this function) and one-to-one by the above characterization result. This isomorphism shows that $V^{\boldsymbol{M}}$ *is indeed a model for* ZFC.

The only properties of $V^{\mathbf{M}}$ that remain to be checked are *transitivity* and *closure under subsets.* Transitivity is trivial. For closure under subsets, let us suppose that $T \in Th_\kappa$ is "**M**-included" in $V^{\mathbf{M}}$, i.e. that

$$\forall S \in Th_\kappa(T \rightarrow S \Longrightarrow S \in V^{\mathbf{M}}.$$

By the definitions of $\rightarrow$ and $V^{\mathbf{M}}$, this means that every maximally consistent theory S, with the property that $S \supseteq \{\varphi : \Box\varphi \in T\}$, is satisfied by some set $a_S \in V^{\mathbf{M}}$. But the above characterization result, a_S is characterized by its sentence $\theta_S \in L_\kappa$, so that S (being the maximal theory of some a_S) must contain this sentence.

It follows that the theory

$$\Phi =: \{\varphi : \Box\varphi \in T\} \cup \{\neg\theta_a : a \in V_\kappa\}$$

is inconsistent. By our first lemma, there must exist some inconsistent $\Psi \subseteq \Phi$ of power $< \kappa$. But it is easy to see that such a Ψ will be equivalent to a theory of the form

$$\Psi' = \{\varphi\} \cup \{\neg\theta_a : a \in A\},$$

where $\neg\varphi \in T$ and A is a subset of V_κ of size $< \kappa$, and hence $A \in V_\kappa$ (since κ is inaccessible). But this means that $\{\varphi\} \cup \{\bigwedge_{a\in A} \neg\theta_a\}$ is inconsistent, so φ logically entails $\bigvee_{a\in A} \theta_a$. Hence $\Box\varphi$ ($\in T$) entails $\Box \bigvee_{a\in A} \theta_a$. By the maximality of θ, it follows that $\Box \bigvee_{a\in A} \theta_a \in T$. From this one can easily deduce that:

$$T = th(\{a \in A : \Diamond\theta_a \in T\}.$$

But $\{a \in A : \Diamond\theta_a \in T\} \in V_\kappa$ (since $A \in V_\kappa$), and so $T \in V^{\mathbf{M}}$ (by the definition of $V^{\mathbf{M}}$).

- *Satisfaction Axioms:*

First, we remark that the interpretation of (the formula defining) L_∞ in our model is L_κ:

$$L_\infty^{(\mathbf{M})} = L_\kappa.$$

Similarly, we remark that the notion of satisfaction of a formula of L_κ by a pointed graph in $V\kappa$ is *absolute* for our model. Finally, we make the trivial observation that our choice for $Sat^{\mathbf{M}}$ was the right one:

$$Sat^{(\mathbf{M})} = Sat^{\mathbf{M}}.$$

This implies that our model satisfies the Satisfaction Axioms.

- *SAFA:*

  The *uniqueness half* (Strong extensionality) follows trivially from the Truth Lemma above, which basically says that $th_\kappa(T, \rightarrow) = T$.

  The *existential half*: we first check that the notion of "weak consistency" of a theory $\Phi \in L_\kappa$ is absolute for our model. Then, let Φ be a "weakly consistent" theory, in the sense of the model **M**. By absoluteness, Φ is really weakly consistent (in the sense above). By the lemma above, there exists a maximally consistent theory $T \supseteq \Phi$, $T \in Th_\kappa$. By the Truth Lemma, this implies that $T \models \Phi$, so Φ is indeed satisfied by some "set" of our model.

■

4.4 The Meaning of Classes

As mentioned above, we only consider classes (and in particular pointed systems and modal theories) as ways of talking about sets. Classes are syntactic objects (predicates), which can be used to denote real objects, i.e. sets. There are various ways a class can be used to do this. For instance, a class can be thought to denote collectively any of its members. In this sense, a class may not have a denotation (if it is empty), but in general will have multiple denotations. We are not concerned here with this notion of denotation, but rather with the non-ambiguous ones, in which a class is used as a *name*, denoting a unique object. So all the denotation relations we shall consider will be *functions* (possibly partial). Note that they will not be "functions" in the set-theoretical sense, since they are defined on proper classes; we shall use them as convenient notations.

4.4.1 The Denotation Function

We are looking for a uniform way to give a *meaning* (a denotation) to every definable class. This denotation has to be an object, i.e. a *set*, which can be thought as the intended "model" of the class. We start with particular families of classes, for which there is an obvious meaning.

1. **Literal Denotation of a Set-Class:** This is the most straightforward, naive, notion of denotation. A class is thought to denote itself, in case it is a set. A class C has a *literal denotation* $d_{\mathrm{lit}}(C)$if and only if C is a set. In this case, C denotes itself: $d_{\mathrm{lit}}(C) = C$.

   Unfortunately, this denotation function is partial, because of the set-theoretical paradoxes.

2. **Modal Denotation (Characterization):** By Super-Strong Extensionality, we know that every set a is characterized by its total analytical pattern, i.e. its modal theory $th(a)$. As mentioned above, this is the class of all the sound analytical descriptions of the set a. So it is a natural idea to use the class $th(a)$ as a name for the set a.

   The *modal denotation (or characterization) function* is the inverse th^{-1} of the function th. Namely, th^{-1} is defined on maximally weakly consistent modal theories and

$$th^{-1}(T) =: \text{ the unique set } a \text{ such that } th(a) = T.$$

3. **Generalized Denotation of a Class:** As we have seen, proper classes cannot have literal denotations. But nevertheless, $SAFA$ gives us a way to assign a unique "generalized denotation" to every class.

   Proposition 4.4.1 *Every class C is observationally equivalent to a unique set $d(C)$.*

   Proof:

   Let $d(C) = th^{-1}(th(C))$ be the modal denotation of the theory of C. By definition, this is the unique set a such that $th(a) = th(C)$, i.e. the unique set $a \equiv C$. ■

   Definition 4.4.2 The function d, given by $d(C) = th^{-1}(th(C))$, will be called *the generalized denotation* (or simply *the denotation*) of C. Obviously, when C is a *set*, the generalized denotation coincides with the literal denotation: $d(C) = d_{\text{lit}}(C) = C$.

   In the context of our system, this denotation operation seems to be the canonical candidate for playing the role of a formal correspondent of the Cantorian operation of "unifying a collection into one thing". The denotation of a class is the only object (set) which is observationally equivalent to that class. So the denotation function d provides us with an *internal representative* (inside the universe of sets) for each class. The uniqueness of this representative shows that there is no other candidate for the Cantorian function of unifying a class into a set. We shall later prove that the denotation of C can be seen as the *closure* $\overline{C}$ of the class C in a certain topology. We shall also see what are the denotations of some of the well-known "paradoxical" classes (e.g. Russell's class).

But, for now, the very existence of the denotation function provides the beginning of *an explanation* for the classical set-theoretical paradoxes. In ZFC and ZFA, the paradoxes were understood as proving that some classes, though definable, are nevertheless *meaningless*, since they do not denote any object. But the proper classes in STS, while not having a literal denotation, still have a meaning, given by their generalized denotation. The lesson of the paradoxes is that *we can freely define sets-as-classes only up to observational equivalence*. So the Comprehension Principle is limited only by the Super-Strong Extensionality Principle: when we "form" or define a set $d(C)$ by comprehension, denoting it by the class $C = \{x : P(x)\}$ defined by a predicate P, we have only identified the set up to observational equivalence. There might be many classes observationally equivalent to C, and the denotation function picks only one of them to represent all. Denotation is preserved by observational equivalence. But this means that $d(C)$ cannot be always equal to C itself (as for the literal denotation), because distinct observationally equivalent classes must have the same denotation. The proper classes (or the "paradoxical" classes) are those classes C for which $d(C) \neq C$.

The classical set-theoretical paradoxes can be thus understood as *paradoxes of denotation.* This idea was proposed, in a slightly different context, by J. Barwise and L. Moss in **Vicious Circles**. The way we understand this proposal in STS is by the generalized denotation function. The classical paradoxes prove that the denotation of a class cannot be always "literal". This is now explained by the fact that observational equivalence puts a restriction on our power to control the actual structure of the denotation of a class.

In other words: our capacity to define sets by predicates, or to unify classes into wholes, is subject to the limitations associated with our capacity to *observe* (in principle) the intended set. No definition can help us identify objects beyond the limits set by the relation of observational equivalence. The Naive Comprehension Principle is true *modulo observational equivalence.* The (generalized) denotation function gives us a canonical way to assign to each definable class some reference, which is an object reflecting all the analytical (i.e. infinitary modal) properties of that class.

The denotation function can be further extended to pointed systems:

4. **The Denotation of a Pointed System**

We can use pointed systems as names for sets, by the following generalization of the above proposition:

Proposition 4.4.3 ***(Weak*** $SAFA$***):*** *Every pointed system* $S = (g, R)$ *is observationally equivalent to a unique set* $d(S)$.

Proof: Take $d(S) = th^{-1}(th(S))$. ■

As we have mentioned in the previous chapter, this proposition is a particularly useful and natural weakening of $SAFA$, which was my initial proposal for the axiom $SAFA$. Its uniqueness half is just super-strong extensionality, while its existential half (every pointed system is $\equiv$ to some set) is called *structural completeness.* If we think of $\equiv$ as modal equivalence, then this can be seen as a *strong Reflection Principle for infinitary modal logic.*

So we can think of a pointed system S as a way to denote the unique set $d(S)$ which is observationally equivalent to S. Thus, we define the *denotation function* d *for pointed systems* by

$$d(S) =: th^{-1}(th(S)).$$

There is a slight ambiguity produced by the fact that pointed systems are classes, and the denotation of S as a class is not in general the same as its denotation as a pointed system. But we shall still use the same notation d for both functions, and decide which function is meant by notation used for its argument (S or (g, R) for systems, and A, B, C... for classes).

We can recover the class-denotation from the pointed-system-denotation, by identifying classes with particular kinds of pointed systems. Namely, we associate to a given class C the pointed system S_C defined by its membership-structure: just take an arbitrary object as the root and take as its immediate successors all the members of C (or some copies of them), and in rest take membership as the successor (edge) relation. Then one can easily show that our two denotation functions "agree" in the following sense:

$$d(C) = d(S_C).$$

This shows that the notion of denotation for pointed systems can be indeed understood as an extension of the class-denotation function.

As in the case of classes, the denotation of a pointed system acts as the *internal representative* (inside the universe of sets) of a family of observationally equivalent pointed systems. The denotation function is the formal implementation of my earlier notion of "actualization" or "realization" of a potential structure. Denotation takes a potential structure (pointed system) S and returns an actual $\in$-structure (set) $d(S)$, which is observationally equivalent to S.

4.5 Decorations and approximations

One can understand the denotation of a pointed system as a generalization of Aczel's notion of *decoration*. Recall that a *decoration* of a graph (i.e. small system) R is a function d_R mapping nodes to sets such that

$$d_R(g) = \{d_R(g') : gRg'\} \text{ , for all nodes } g.$$

For large systems, this notion of decoration is not so useful, since the collections $\{d_R(g' : gRg'\}$ on the right side will usually not be sets, while the left side $d_R(g)$ is by definition a set. So most large systems will not have decorations in Aczel's sense. But there is a more natural generalization of this notion for systems:

Definition 4.5.1 A *decoration* of a system (binary relation) R is a definable class-function d_R, having the property that

$$d_R(g) \equiv \{d_R(g') : gRg'\} \text{ , for all sets } g.$$

The *decoration* of a pointed system $S = (g, R)$ is the set $d_R(g)$.

We use the same name (decoration) for the more general notion. This is consistent, since one can easily see that, if a system R has a decoration in Aczel's sense, then it has a decoration in our generalized sense, and the two notions coincide for S. In particular, they coincide for the small systems (graphs), which form the specific domain of application of Aczel's definition. Moreover, generalized decorations are unique (by super-strong extensionality). So there is no possibility of confusion, and we can use the same term to denote both notions.

We do not use capital letters to denote decorations, since we shall prove later that they are always *sets*. But the relevant part of a decoration is its restriction to the class $\{g : gRg' \text{ for some } g'\}$ of all the nodes of the system R. And this restriction is not always a set (but only when the class of all nodes is itself a set).

An easy consequence of $SAFA$ is the following generalization of Aczel's formulation of AFA:

Proposition 4.5.2 *Every system has a unique decoration. The decoration of a pointed system S is its denotation $d(S)$.*

As a result of this proposition, *the notions of decoration and denotation of a pointed system coincide.* Aczel's decorations were conceived as generalizations of Mostowski's collapsing functions, that map well-founded structures to set-structures in ZFC, and Aczel's AFA was a generalization of Mostowski's well-known Collapsing Lemma to non-wellfounded structures. Now we can understand the denotation of a pointed system as a way of collapsing it to a set in STS, and the weak version of $SAFA$ as a further generalization of the Collapsing Lemma.

Proof of the proposition: For a system R, define the function d_R by:

$$d_R(g) =: d(g, R),$$

where d is the denotation function for pointed systems. One can check that d_R is indeed a definable class-function (the actual definition is given by $d_R(g) = th^{-1}(th(g, R))$). To show that d_R is a decoration, we need to prove that the denotation function has the following property:

$$d(g, R) \equiv \{d(g', R) : gRr\},$$

for every pointed system (g, R).

To show this, we use the fact that $d(g, R) \equiv (g, R)$, to prove that, for all modal sentences φ, we have:

$$d(g, R) \models \varphi \text{ iff } \{d(g', R) : gRr\} \models \varphi.$$

The proof of this is an easy induction on the complexity of modal formulas φ. ■

Applications:

1. **Aczel's Universe.** One can easily use the concept of decoration to construct a standard model of Aczel's ZFA inside our theory: just take the decorations of all the pointed graphs in V:

   $$V_{AFA} =: \{d_R(g) : \ (g, R) \in V \text{ is a pointed } \ V - \text{graph}\}.$$

   This coincides with the class HS of all hereditarily small sets and can be proved to be a model of ZFA. It is the natural model of Aczel's theory inside our universe of sets.

2. **Unfoldings.** As another application, we can now formally define the *unfoldings* of rank α, by transforming the informal definition by recursion into a corresponding system (relation), whose decorations are the unfoldings.

Formally, we first put, for every ordinal α and every set x:

$$a_x^\alpha =: (\alpha, x, 0)$$

$$b_x^\alpha =: (\alpha, x, 1)$$

$$c_x^\alpha =: (\alpha, x, 2)$$

Then we define a class-relation R by the following clauses:

(i) $a_x^{\alpha+1} R a_y^\alpha$ iff $y \in x$

(ii) $a_x^\lambda R b_x^\alpha$ iff $\alpha < \lambda$ and λ is a limit ordinal

(iii) $b_x^\alpha R \{\alpha\}$, $b_x^\alpha R c_x^\alpha$

(iv) $\{\alpha\} R \alpha$

(v) $\alpha R \beta$ iff $\beta < \alpha$

(vi) $c_x^\alpha R \alpha$, $c_x^\alpha R a_x^\alpha$

(vii) no other pairs are in the relation R but the ones listed above (in cases **(i)**-**(vi)**).

We can then define the unfoldings by:

$$x^\alpha =: d_R(a_x^\alpha)$$

and check by induction that they satisfy the desired recursive equations.

We can similarly define the modal sentences φ_a^α and φ_α^a informally defined above (to capture the information given by unfoldings and descriptions). It is easy to check that they have all the properties mentioned in section **4**, in particular that, for every set b, we have:

$$b \models \varphi^a_\alpha \text{ iff } b \equiv_\alpha a.$$

Moreover, one can easily see that φ^a_α is semantically equivalent to the conjunction of all the modal sentences of rank $\leq \alpha$. We shall call the sentence φ^a_α the *modal description of rank α of the set a.*

3. Boffa Approximations We shall show now that *every class can be approximated by hereditarily small sets* in a canonical way. In other words, we show that, for every α, every possible unfolding of rank α is realized in Aczel's universe $V_{AFA} = HS$. More precisely,

Proposition 4.5.3 *Every class is approximable by sets in V_{AFA}, in the following sense : for every class C and every ordinal α there exists some $c_\alpha \in V_{AFA}$ s.t. $c_\alpha \equiv_\alpha C$.*

We first prove the proposition for sets. For this, we introduce the *Forti-Boffa approximations*:

Definition 4.5.4 A *Forti-Boffa approximation of rank α of a set a* is a set a_α, satisfying the following conditions:

$$\begin{aligned} a_{\alpha+1} &= \{b_\alpha : b \in a\} \\ a_\lambda &= \{b_\lambda : b \text{ is a set s.t. } b_\alpha \in a_{\alpha+1} \text{ for every } \alpha < \lambda\}, \text{ for } \lambda \text{ limit ordinal.} \end{aligned}$$

Observe that this is not a simple definition by recursion: the clause for limit-ordinals is not inductive. But one can prove the existence of the

Boffa approximations by using decorations. Using the same notations a_x^α, we define a class-relation R' by the following inductive clauses

(i') $a_x^{\alpha+1} R' a_y^\alpha$ iff $y \in x$

(ii') $a_x^\lambda R' a_y^\lambda$ iff $a_x^{\alpha+1} R' a_y^\alpha$, for every $\alpha < \lambda$, (λ limit ordinal)

Then we define

$$x_\alpha =: d_R(a_x^\alpha)$$

and check by induction that these approximations are *hereditarily small* and satisfy the conditions of the above definition. Then we show that they indeed approximate the given set, in the sense of the proposition above:

Lemma 4.5.5 *For all sets a and all ordinals α, we have*

$$a_\alpha \equiv_\alpha a.$$

Proof: The proof goes by induction on α. The successor step is trivial, since the inductive clause for approximation in this case matches the recursive bisimulation-like clause for $\equiv_{\alpha+1}$. For the limit case, we need to show that $a_\lambda \equiv_\lambda a$. For this, we check by induction on β that:

$$a_\lambda \equiv_\beta a' \text{ for all } \beta < \lambda \text{ and all sets } a'.$$

Both the successor and the limit cases for this second induction are trivial. ■

Now we can extend this to classes:

Proof of Proposition above: Let C be a given class. Put $c = d(C)$. This is a set having the property that $c \equiv C$, and so $c \equiv_\alpha C$ for every α. By the above lemma, the hereditarily small approximations c_α of the set c satisfy $c_\alpha \equiv \alpha c$. By transitivity we obtain $c_\alpha \equiv C$. ■

Observe the similarity between approximations and unfoldings. The only difference is at the limit stages. The unfoldings approximate the given set in an informational sense, while the Boffa approximations are real "set-theoretical" approximations: their $\in$-structure is more and more alike the structure of the given set. They have all the useful properties of the unfoldings, and moreover they provide canonical hereditarily small representatives for every set, up to observational equivalence of rank α. But the price for this is that they are not necessarily well-founded: they live in Aczel's universe V_{AFA}, while the unfoldings live in the well-founded universe WF.

As far as I know, the above notion of approximation was first defined inside $SAFA$ by Forti [Forti1989], in his construction of a model for the Generalized Positive Comprehension Principle. It was then formalized by Boffa [Boffa1989], who gave it a direct axiomatization, independent of AFA.

4.5.1 Paradoxical Sets

The classical set-theoretical paradoxes show that certain classes are "paradoxical", i.e. they cannot be sets, but proper classes. In our system, this means they do not denote themselves: they are not their own (generalized) denotations. But what are the denotations of these paradoxical classes? We have understood these denotations as "actualizations" of the corresponding classes, and characterized them as decorations of the associated pointed systems. But, in the case of the specific classes that generate the classical paradoxes, we would like to have some concrete descriptions for their denotations.

(1.) The Universe: The class U of all sets is a set (and so it is its own denotation). What about Cantor's paradox? Well, U is an exception to Cantor's diagonal argument: $\mathcal{P}U = U$ is not bigger than U. Cantor's argument cannot be carried out, since the collections obtained by diagonalization are not sets, but proper classes. U is the largest fixed-point of the power-set operator.

(2.) The set of all ordinals: We start with the class On of all the von Neumann ordinals. An ordinal is a set which is well-ordered by $\in$. Alternatively, the class On can be defined as the least class closed under the two Cantorian operations of *successor* and *limit* (i.e. union): if α is an ordinal then its successor $\alpha \cup \{\alpha\}$ is an ordinal; if A is a set of ordinals then its supremum $\bigcup A$ is an ordinal.

One can show that ordinals look more and more alike as they grow bigger:

indeed, we have $\alpha \equiv_\alpha \beta$ for all $\beta > \alpha$. Moreover, they look more and more alike the class of all ordinals: $\alpha \equiv_\alpha On$ for every α.

Now, if On were a *set* then we could form the set $On \cup \{On\}$. But from the above observation (and the recursive bisimulation-type properties of $\equiv_\alpha$) would follow that $On \equiv_\alpha On \cup \{On\}$ for all α, and so $On \equiv On \cup \{On\}$. By super-strong bisimilarity we obtain $On = On \cup \{On\}$, so $On \in On$. This is a contradiction: the ordinals are well-ordered by $\in$, i.e. well-founded. So On *cannot be a set.*

The above argument closely resembles the classical Burali-Forti argument. But this version of the argument can also be applied to the set $d(On)$, the denotation of the class On. This is the smallest set closed under successors and limit(set-union). The same argument above, starting with "if On were a set... " can be now applied to $d(On)$, which is indeed a set and is observationally equivalent to On. But now the argument does not lead to a contradiction: $d(On)$ is not necessarily well-founded. The conclusion is that we just have:

$$d(On) = On \cup \{d(On)\}.$$

So the final outcome of the Cantorian operations of successor and limit is a fixed point of these operations. This reveals the structure of the the set $d(On)$, the denotation of the class of all ordinals; this "largest ordinal" contains as members all the well-founded ordinals, and one more thing: itself. We shall sometimes consider $d(On)$ as a non-standard "ordinal", bigger than all the standard ones. For this reason we shall use the symbol ∞ to denote it. So we put

$$\infty =: d(On) = On \cup \{\infty\}.$$

(3.) The Universe of the Limitation-of-Size Conception: The natural universe for the limitation of size conception is the class HS of all hereditarily small sets, which coincides with Aczel's universe V_{AFA}. This is not a set. To see this, recall the Proposition in the previous section, saying that every set can be approximated by hereditarily small sets. As a consequence, we obtain that *the denotation of Aczel's universe V_{AFA} is the whole universe U*:

Corollary 4.5.6 *$V_{AFA} \equiv U$, and so $d(V_{AFA}) = U$.*

Proof: We show that $V_{AFA} \equiv_{\alpha+1} U$ for all $\alpha \in On$: one direction is easy, since $V_{AFA} \subseteq U$; for the other direction, we use the above-mentioned Proposition, saying that $\forall\alpha \in On \forall a \in U \exists a' \in V_{AFA} a \equiv_\alpha a'$. ■

So the real universe, which is reflexive ($U \in U$), is indistinguishable (at any stage of unfolding) from Aczel's non-reflexive universe V_{AFA}. This seems weird: what is the use of U then? The answer is that V_{AFA} is not an actual object. Its actualization is our universe U.

Remark: $V_{AFA} \neq U$ (since by Russell's Paradox we have $V_{AFA} \notin V_{AFA}$, while $U \in U$), and *so V_{AFA} is not a set* (since otherwise $V_{AFA} \equiv U$ would imply $V_{AFA} = U$, by **SAFA**).

(4.) The Russell Set: A set is reflexive if it is a member of itself. By Russell's Paradox, the Russell class of all non-reflexive sets $R = \{x : x \notin x\}$ cannot be set. What are the missing elements? It turns out that every reflexive set which is not hereditarily small (not in V_{AFA}) can be approximated by non-reflexive sets. So we conclude that we have:

$$d(R) = R \cup (U \setminus V_{AFA})$$

(5.) The set of all wellfounded sets: Mirimanoff 's paradox shows that the class WF of all wellfounded sets cannot be a set. Its denotation is

$d(WF) = \{x :$ every hereditarily small element of the transitive closure of x is wellfounded$\}$.

Observe that WF is generated by the well-known iterative process: if we put $V_0 = \emptyset$, $V_{\alpha+1} = \mathcal{P}V_\alpha$ and $V_\lambda = \bigcup_{\alpha<\lambda} V_\alpha$ for λ limit, then we have $WF = \bigcup_\alpha V_\alpha$ Its denotation $d(WF)$ is the *least fixed point of this process* (among *sets*):

$$d(WF) = \mathcal{P}(d(WF)).$$

Examples **(1.), (2.) and (5.)** can be generalized to a *theory of monotonic operators, having least and greatest fixed points.* Our universe has very interesting properties in this respect: unlike what happens in ZFC (or ZFA), the usual iterative-inductive way of constructing the least fixed point can be dualized to get a construction of the largest fixed point (see below the chapter on Applications).

4.5.2 The Size of the Universe

Proposition 4.5.7 *If $A \neq B \subseteq V_{AFA}$ are distinct subclasses of V_{AFA} then their denotations are distinct: $d(A) \neq d(B)$.*

Proof: Let $a \in A \setminus B$ (say). Then $a \in V_{AFA}$, so a is characterizable by a modal formula θ^a, by the Modal Characterization Theorem for sets in V_{AFA} proved by Barwise and Moss. Hence $A \models \Diamond\theta^a$, while $B \models \neg\Diamond\theta^a$. But $A \equiv d(A)$ and $B \equiv d(B)$, so $d(A) \models \Diamond\theta^a$, $d(B) \models \neg\Diamond\theta^a$, and hence $d(A) \neq d(B)$. ■

Note that the above Proposition gives us an injection of the "external" powerset of V_{AFA}(the family of all subclasses of V_{AFA}) into U, and hence that $|U| \geq 2^{|V_{AFA}|} > |V_{AFA}|$. Using **SAFA**, one can show that we also have:

Proposition 4.5.8 $|U| \leq 2^{|V_{AFA}|}$. *Hence* $|U| = 2^{|V_{AFA}|}$.

Recall the notation $\infty =: d(On) = On \cup \{\infty\}$. This can be considered as being "the last ordinal", although a non-standard one (since it is not wellordered by $\in$, but is still wellordered by $\subset$). It has the size of V_{AFA} (and of On), so we can identify it with "the cardinal of V_{AFA}" (which was until now just an informal notation !). Thus we *define*:

$$|C| =: \infty \text{ iff } C \text{ has the same size as } V_{AFA}.$$

Hence we have:

$$|V_{AFA}| = |WF| = |On| = |\infty| = \infty.$$

Let us also define:

$$|C| = U \text{ iff } C \text{ has the same size as } U.$$

Then the above Proposition says that

$$|U| = 2^{\infty}.$$

A nice equation indeed !

Chapter 5

Comprehension and Topology

5.1 Comprehension for Modal Theories

In this section, we prove the Comprehension Principle for Modal Theories: *every modal theory defines a set*. We also prove that this Modal Comprehension is enough to generate all the sets: *every set is definable by an (infinitary) modal theory*. We also show that *the denotation function for classes has all the properties of a topological closure operator*. We do this by introducing a closure operator on classes, dual to the notion of deductive closure of a modal theory. We later prove that this closure operator coincides with the denotation function on classes.

5.1.1 Closure and Modal Definability

It is well-known that every relation between two sets (or classes) induces a "Galois connection" between the respective subsets (or subclasses). I investigate here the Galois connection induced by the *satisfaction relation* between the class U of all sets and the class L_∞ of all infinitary modal sentences.

Definition 5.1.1 For a modal theory T , the *class defined by* T is the class $Mod(T) =:\{a : a \models \varphi$ for every $\varphi \in T\}$ consisting of all the ””models” of the theory Φ, i.e. all the sets satisfying Φ. We say a class C is *modally definable* if it is defined by some modal theory, i.e. $C = Mod(\Phi)$ for some $\Phi \subseteq L_\infty$.

Dually, for a class $C \subseteq U$, consider its modal theory $Th(C) =: \{\varphi : c \models \varphi$ for every $c \in C\}$, containing the modal sentences that satisfied by every member of C. We say that a theory is *deductively closed*if it is of the form $Th(C)$. These theories are the ones that are closed under (semantical) entailment.

Notice the difference between $Th(C)$ and $th(C)$.

Consider now the family $Class$ of all classes. Again, this is just a way of speaking: everything we shall define or prove can be stated without reference to families of classes. But nevertheless, from an intuitive point of view, it is useful to consider the family $Class$ and observe that it is a *complete lattice* with respect to inclusion. The supremum and infimum of any family of classes is given by their union and their intersection, respectively. The least element of $Class$ is the empty set $\emptyset$, while the greatest element is the universe U. We can also consider the family $Theory$ of all modal theories, which is also a complete lattice with inclusion. Then we can understand the operators Mod and Th as functions between these complete lattices.

Proposition 5.1.2 *The pair* (Mod, Th) *is a* Galois connection *between classes and theories (both ordered by inclusion), i.e:*

$$C \subseteq Mod(T) \text{ iff } T \subseteq Th(C).$$

Consequently, the operators $T \longmapsto Th(Mod(T))$ *(on theories) and* $C \longmapsto Mod(Th(C))$ *are topological* closure *operators on the corresponding complete lattices.*

Proof: Easy. ■

Definition 5.1.3 Let $\overline{T} =: Th(Mod(T))$ (the deductive closure of T) and $\overline{C} =: Mod(Th(C))$. The second will be called *the closure* of C.

As mentioned above, both have the properties of a topological closure operator. In particular, we have:

- every class is included in its closure: $C \subseteq \overline{C}$;
- the empty-set and of the universe are their own closures: $\overline{\emptyset} = \emptyset$, $\overline{U} = U$. This implies that the empty set $\emptyset$ and the universe U are both *sets.*
- Closure permutes with arbitrary intersections: if $\mathcal{F}$ is a (definable) family of classes then

  $$\overline{\bigcap \mathcal{F}} = \bigcap \{\overline{C} : C \in \mathcal{F}\}.$$

  We shall see later how can we state this directly, without any reference of families of classes.
- Closure permutes with binary union: $\overline{A \cup B} = \overline{A} \cup \overline{B}$. This will be later generalized to *small unions*, i.e. unions of small families of classes.

Observe that a class C is modally definable if and only if $C = \overline{C}$. The closure of a class $\overline{C}$ will be studied in the next section (when it will be given an alternative definition, in terms of decorations). We shall see that the closure of C is the (generalized) denotation of C. So the closure operation on classes can be understood as the Cantorian operation of unifying a class into a set: to unify all the objects having some property into a whole one might have to add some extra-objects, which cannot be separated from the class. In topological terms, these are the "limits" or "accumulation points" of the process that generates all the members of the class.

We know that every set can be characterized by a modal theory. We show now that every set is also *definable* by a modal theory. Moreover, the converse is true: modally definable classes are sets. So we obtain our first characterization of sets-as-classes:

Theorem 5.1.4 ***(Modal Comprehension Principle and its Converse)***

A class C is a set if and only if it is modally definable, i.e. iff $C = \overline{C}$. In particular, $\overline{C}$ is always a set (since $\overline{\overline{C}} = \overline{C}$, by the topological properties of closure).

Proof: To show that every set is modally definable: let a be set, $T = th(a)$ be its characteristic theory. Then check that the theory $T^{-\,\Box} =: \{\varphi : \Box\varphi \in \Phi\}$ defines the class a.

To show modally definable classes are sets: Let Φ be a modal theory and let $C = \{c : c \models \Phi\}$ be the class defined by it. We can safely assume that Φ is the largest theory with this property, i.e. that $\Phi = \{\varphi : c \models \varphi \text{ for all } c \in C\}$.

Now take the theory $D\Phi =: \{\Box\varphi : \varphi \in \Phi\} \cup \{\Diamond\varphi : (\neg\varphi) \notin \Phi\}$. Check this theory is weakly consistent, take some set a satisfying it and check that $a = C$. ■

The above theorem can be restated in the following form:

> *The restriction of the operator Th to the universe U of sets gives a bijective correspondence between sets and deductively closed modal theories. The inverse of this correspondence is given by the operator Mod.*

5.1.2 Applications of Modal Comprehension

As an easy consequence we prove:

Proposition 5.1.5 *The union axiom and the power set axiom are true.*

Proof: Let a be a set and let $\Phi = Th(a) =: \{\varphi : x \models \varphi \text{ for every } x \in a\}$. Since a is a set, we have $a = Mod(Th(a)) = \{x : x \models \Phi\}$, so that a is defined by the theory Φ. Then it easy to check that$\mathcal{P}a$ is defined by the theory $\Box\Phi =: \{\Box\varphi : \varphi \in \Phi\}$.

Similarly, $\bigcup a$ is defined by the theory $\Phi^{-\ \Box} =: \{\varphi : \Box\varphi \in \Phi\}$, but this fact is harder to check: it makes essential use of the *strong* formulation of $SAFA$. To see this, let a' be some set satisfying the theory $\Phi^{-\ \Box}$. We need to show that $a' \in \bigcup a$, i.e. that there exists some $b \in a$, such that $a' \in b$. To construct the set b, we take the theory

$$\Psi = \Phi \cup \Diamond th(a') = \Phi \cup \{\Diamond\varphi : a' \models \varphi\}.$$

Now, it is easy to see that any set b which satisfies this theory must have the desired properties that: $b \in a$ and $a' \in b$. So *it is enough to show there exists some set b satisfying the theory Ψ*. But there is no direct way to show that Ψ is consistent (i.e. to construct a pointed system which satisfies it). The theory Ψ is a proper class, defined in terms of the maximal theory $th(a')$ of a', for which we do not have any control, except that we know that a' satisfies $\Phi^{-\ \Box}$ (so that we must have $\Phi^{-\ \Box} \subseteq th(a')$).

The trick is to use *strong* $SAFA$: in this way, we only need to check that Ψ is *weakly consistent*. For this, let $\Psi' \subseteq \Psi$ be a sub*set* of Ψ. Then, by the definition of Ψ, it follows that we can write Ψ' as a union of two sets $\Psi' = \Phi' \cup \Diamond\Theta$, with $\Phi' \subseteq \Phi$ and $\Theta \subseteq th(a')$. Then we must have $(\bigwedge \Theta) \in th(a')$ (here $\bigwedge \Theta$ is a sentence since Θ is a set). But $th(a')$ is a consistent theory, so $(\neg \bigwedge \Theta) \notin th(a')$. But, as mentioned above, we know that $\Phi^{-\ \Box} \subseteq th(a')$, and hence $(\neg \bigwedge \Theta) \notin \Phi^{-\ \Box}$. But this means that $(\Box\neg \bigwedge \Theta) \notin \Phi$. By the definition of $\Phi = Th(a)$, it follows that there must exist some $b \in a$ which falsifies $(\Box\neg \bigwedge \Theta)$. In other words, $b \models \Diamond \bigwedge \Theta$. From this follows that $b \models \Diamond\theta$ for every $\theta \in \Theta$, i.e. $b \models \Diamond\Theta$. But we also have $b \in a = \{x : x \models \Phi\}$, so $b \models \Phi$ and hence $b \models \Phi'$ (since $\Phi' \subseteq \Phi$). Putting these together, we see that $b \models \Phi' \cup \Diamond\Theta$, and so the theory $\Psi' = \Phi' \cup \Diamond\Theta$ is consistent. But Ψ' is just an arbitrary subset of Ψ. Hence Ψ is weakly consistent. By $SAFA$, there exists some b which satisfies the theory Ψ. (As pointed out above, any such b will have the desired properties.) ■

One can similarly prove the universe of sets is closed under small unions, arbitrary intersections, singletons etc.

In particular, it follows that V is in fact a model of full ZFC^-. We have not assumed Foundation for V, so we cannot prove it to be a model of full ZFC. But we can of course construct the class $WF \subseteq V$ of all well-founded sets, which is indeed a model of ZFC.

We shall now prove a version of the topological property stating that closure permutes with arbitrary intersections. Taking into account the fact that the closure of a class is always a set and that a class is a set if and only if it closed, we can restate this topological property in a form that does not mention families of classes:

Proposition 5.1.6 *The universe of sets U is closed under arbitrary intersections, i.e. if A is a collection of sets then $\bigcap A$ is a set.*

Proof: Let A be a collection (class) of sets. Its intersection $\bigcap A$ is a definable class. Its closure $\overline{\bigcap A}$ is a set, which by the topological properties of closure coincides with the intersection of all the closures of the members of A. But the members of A are sets, so they coincide with their own closures. Hence:

$$\overline{\bigcap A} = \bigcap \{\overline{a} : a \in A\} = \bigcap A.$$

So $\bigcap A$ is a set. ■

Corollary 5.1.7 *The universe of sets U coincides with the family of all its subclasses that are closed with respect to the topology induced by the denotation function. As a consequence, U is a complete lattice with respect to* *inclusion. The infimum of a collection A of sets in this lattice is its intersection $inf(A) = \bigcap A$, and the supremum of C is the closure of its union $sup(A) = \overline{\bigcup A}$.*

5.2 Characterizations of Sets and Closures

As we have seen above, the operator $C \longmapsto \overline{C}$ is a topological closure operator. In particular, *$\overline{C}$ is a superset of C: $C \subseteq \overline{C}$.* To unify a collection you may have to add some objects (the "limits" of the process that generates the collection). The way our system formalizes the incompleteness of proper classes is by saying they are not closed in this topology. This gives a second characterization of sets:

> *A class is a set if and only if it is closed in the topology given by the denotation function.*

So the universe U coincides with the family of all its closed subclasses. Small classes play the role that *finite* sets have in topology: finite sets are closed in every well-behaved (Hausdorff) topology; while all small sets are closed in our topology.

The following theorem gives a list of several characterizations for closure:

Theorem 5.2.1 *The closure $\overline{C}$ of a class C can be characterized by any of the following statements:*

1. *The closure is the denotation: $\overline{C} = d(C)$. In other words, $\overline{C}$ is the* unique set *observationally equivalent to C.*

2. *The closure* $\overline{C}$ *is the* largest class *observationally equivalent to* C. *So to actualize a class you have to realize* all *its potentialities: out of all the possible observationally equivalent classes, the actualization function (closure) picks up the largest one.*

3. *The closure* $\overline{C}$ *is the* least set *that includes* C. *It is the "best upper approximation" of* C *inside the universe of sets.*

4. *An object is a member of* $\overline{C}$ *if it cannot be distinguished from the members of* C *by a modal formula; i.e. if it satisfies all the modal formulas true everywhere on* C.

5. *An object is a member of* $\overline{C}$ *if it cannot be distinguished from the members of* C *at any stage of unfolding:*

$$\overline{C} \quad = \quad \{c : \textit{ for every } \alpha \textit{ there is some } c_\alpha \in C \textit{ s. t. } c \equiv_\alpha c_\alpha\}$$

Proof:

1. By the definition of denotation, we see that in order to prove that $\overline{C} = d(C)$, we need to show that $\overline{C}$ is observationally equivalent to C: $C \equiv \overline{C}$. In other words, we need to show that C and $\overline{C}$ satisfy the same modal sentences φ. We can prove this by induction on the complexity of φ. The interesting case is when φ is of the form $\Diamond\psi$.

   Suppose that $C \models \Diamond\psi$. Then there exists some $c \in C$ such that $c \models \psi$. But we know that $C \subseteq \overline{C}$, by the topological properties of closure. So $c \in \overline{C}$, and hence $\overline{C} \models \Diamond\psi$.

Conversely, suppose that $\overline{C} \models \Diamond\psi$. Then there exists some $c \in \overline{C}$ such that $c \models \psi$. By definition $\overline{C} = Mod(Th(C))$, so that we must have $c \models Th(C)$. But this means that we cannot have $(\neg\psi) \in Th(C)$ (else c would satisfy both ψ and $\neg\psi$). So $(\neg\psi) \notin Th(C)$, from which we conclude, by the definition of Th, that not all the members of C satisfy $\neg\psi$. Hence there must exist some $c' \in C$ such that $c' \models \psi$. But this implies that $C \models \Diamond\psi$.

2. To show that $\overline{C}$ is the largest class observationally equivalent to C, let B be any class such that $B \equiv C$. Then $\overline{B} = \overline{C}$ (by **SAFA**). But $B \subseteq \overline{B}$, so $B \subseteq \overline{C}$.

3. To show $\overline{C}$ is the least set observationally equivalent to C, let a be any set such that $C \subseteq a$. Then, by the topological properties of the closure, we have $\overline{C} \subseteq \overline{a} = a$.

4. This one follows directly from the definition of $\overline{C} =: Mod(Th(C))$.

5. This one follows from the previous one, by observing that an object c satisfies all the sentences in $Th(C)$ iff, for every ordinal α, c satisfies all modal sentences of rank $\leq \alpha$ in $Th(C)$. But this is equivalent to saying that, for every α there exists some c_α such that $c \equiv_\alpha c_\alpha$.

■

From this we obtain two new characterizations of sets:

A class is a set if and only if it is maximal among all the classes which are observationally equivalent to it.

A class is a set if and only if it contains every object which cannot be distinguished from all its members at any stage of unfolding.

5.3 Limits

The way hypersets in ZFA relate to wellfounded sets has been compared by many to the construction of the complex numbers from the reals or to the construction of the rational numbers (as pairs of integers). The way we described the sets in STS (as ON-long sequences of unfolding) using wellfounded sets resembles the construction of the real numbers as sequences of rationals. The analogy can be pursued by studying the notion of *limit* that comes with the topology induced by modal descriptions. I only mention briefly some definitions and some properties.

5.3.1 Convergence, Compactness and Completeness

Two sets a, b are said to be *equivalent with respect to a modal sentence* φ if we have: $a \models \varphi$ iff $b \models \varphi$. In this case we write $a \equiv_\varphi b$.

Definition 5.3.1 Given an On-long sequence of sets indexed by ordinals $\langle a_\alpha \rangle_{\alpha \in On}$ and a set a, we write

$$\lim_{\alpha \to \infty} a_\alpha = a$$

if for every infinitary modal sentence φ there exists some ordinal $\delta \in On$ such that for all $\alpha > \delta$ we have $a_\alpha \equiv_\varphi a$. Equivalently, if for every ordinal $\varepsilon \in On$ there exists some ordinal $\delta \in On$ such that for all $\alpha > \delta$ we have $a_\alpha \equiv_\varepsilon a$. In this case we say the sequence is *convergent*.

Definition 5.3.2 The sequence $\langle a_\alpha \rangle_{\alpha \in On}$ is said to be a *Cauchy* sequence if for every modal sentence φ there exists some ordinal $\delta \in On$ such that for all $\alpha, \beta > \delta$ we have $a_\alpha \equiv_\varphi a_\beta$. Equivalently, if for every ordinal $\varepsilon \in On$ there exists some ordinal $\delta \in On$ such that for all $\alpha, \beta > \delta$ we have $a_\alpha \equiv_\varepsilon a_\beta$.

Examples of convergent sequences are:

$$\lim_{\alpha \to \infty} \alpha =: \overline{On} = On \cup \{\overline{On}\}$$

and

$$\lim_{\alpha \to \infty} V_\alpha = \overline{WF} = \overline{\bigcup V_\alpha},$$

where $\{V_\alpha\}_\alpha$ is the usual iterative hierarchy.

As a consequence of $SAFA$, one can show that the universe U is compact, as a topological space:

Theorem 5.3.3 *(**compactness**) Every On-long sequence of sets $\langle a_\alpha \rangle_{\alpha \in On}$ has a convergent subsequence.*

Proof: Unlike most others, this proof uses the strong formulation of $SAFA$ in an essential manner. The theorem is not provable if we only assume the "weak version" of $SAFA$.

Assume as given the sequence $\{a_\alpha\}_{\alpha \in On}$. It is enough to show that the sequence has an *accumulation point*, i.e. there exists some set a such that $\forall \beta \exists \gamma > \beta \; a_\gamma \equiv_\beta a$. It is easy to construct a subsequence convergent to such an accumulation point.

The following should be understood as a recursive definition: for each ordinal β, consider the sentence

$$\psi_\beta =: \bigvee_{\gamma > \beta} \varphi^\beta_{a_\gamma},$$

where each sentence $\varphi^\beta_{a_\gamma}$ is the (above defined) modal description of rank β of the set a_γ.

Observe that, for each β, ψ_β is a modal sentence, despite the fact that it was defined by a disjunction of a family indexed by all the ordinals bigger than β. The reason is that there is only a small set of distinct possible (modal) descriptions of a given rank. Observe also that, in general, we have

$$a_\gamma \models \psi_\beta \text{ for every } \beta < \gamma.$$

Now take the theory $T =: \{\psi_\beta : \beta \in On\}$. This theory is weakly consistent: every *subset* $T' \subseteq T$ is small (since $T' \subseteq T \subseteq V$ and V is closed under set-formation, so the set T' must be in V), and so is contained in some subtheory of the form $T_\alpha = \{\psi_\beta : \beta < \alpha\}$, for some ordinal α. But each such subtheory is consistent, since it is satisfied by a_α.

So, by $SAFA$, we conclude that T must have a model a. This set a is easily seen to be an accumulation point: let β be a given ordinal; then a is a model of T, so in particular $a \models \psi_\beta$. But by the definition of phi_β, this implies that there exists some $\gamma > \beta$ such that $a \models \varphi^\beta_{a_\gamma}$, and hence $a \equiv_\beta a_\gamma$. So we have $\forall\beta\exists\gamma > \beta\ a_\gamma \equiv_\beta a$, i.e. a is an accumulation point. ■

Corollary 5.3.4 ***(completeness):*** *Every Cauchy sequence is convergent.*

Proof: Trivial, given the previous theorem: it is easy to see that every two accumulation points of a Cauchy sequence must be in the relation $\equiv \alpha$, for every α, so they must be observationally equivalent. By super-strong extensionality, they have to be identical. But this means the sequence has exactly one accumulation point, i.e. it has a limit. ■

5.3.2 Properties of limits

Proposition 5.3.5 *Closure is closure under limits:*

$$\overline{C} = \{a : \exists \{a_\alpha\}_{\alpha \in On} \subseteq C s.t. \lim_{\alpha\to\infty} a_\alpha = a\}.$$

This gives a new characterization of set: A class C is a set iff it is closed under limits.

Proposition 5.3.6

1. *The limit operator permutes with the following operations: singleton, powerset, binary union, infinitary union operator, Cartesian product.* But not with intersections !

2. *The following relations are preserved under limits: $\in, \subseteq, \models, \equiv_\alpha, \equiv_\varphi$.*

3. *If $\{a_\alpha : \alpha \in On\}$ is convergent then*

$$\lim_{\alpha\to\infty} a_\alpha = \{\lim_{\alpha\to\infty} b_\alpha : \{b_\alpha : \alpha \in On\} \text{ is a convergent sequence s.t. } b_\alpha \in a_\alpha \text{ for every } \alpha\}.$$

4. *If $\lim_{\alpha\to\infty} a_\alpha = a$ and φ is a modal formula then the following are equivalent:*

$$(i). \qquad a \models \varphi$$

$$(ii). \qquad \exists\alpha\forall\beta > \alpha \ a_\beta \models \varphi$$

$$(iii). \qquad \forall\alpha\exists\beta > \alpha \ a_\beta \models \varphi$$

In the next section, we shall generalize this property of modal sentences to an important class of infinitary first-order sentences, namely the extended positive infinitary formulas ($\mathbf{EPF}_\infty$:

Proposition 5.3.7 *Suppose* $\lim_{\alpha\to\infty} a_\alpha = a$ *and* $\varphi(x) \in \mathbf{EPF}_\infty$*, such that* $\forall\alpha\exists\beta > \alpha \ \varphi(a_\alpha)$*. Then we also have* $\varphi(a)$*.*

.

5.4 Strong Comprehension and Closure Properties

The so-called Generalized Positive Comprehension Principle is a very strong comprehension schema, that has been considered by Malitz (in [Malitz 1976]) and Weydert(in [Weydert 1989]) and proved to be consistent by Forti [Forti 1989]. The model constructed by Forti is a topological structure, which can be proved to satisfy all our axioms. So our system STS can be considered an axiomatization of Forti's model. The axiomatization, though necessarily incomplete, is strong enough to prove the Generalized Comprehension Principle. Moreover, $SAFA$ is much stronger than this principle and it decides most set-theoretical questions regarding the shape and features of the intended universe.

In this section we shall prove a strengthening of the Generalized Positive Principle in our system STS.

Definition 5.4.1 The class of *infinitary extended positive formulas* (EPF_∞) is the least class containing all atomic formulas $x \in y$, and closed under the following operations

1. infinitary conjunctions and disjunctions: if Φ is a set of EPF_∞-formulas then $\bigwedge \Phi$ and $\bigvee \Phi$ are also EPF_∞-formulas
2. existential quantifiers: if φ is an EPF_∞-formula then $\exists x \varphi$ is also an EPF_∞-formula
3. bounded universal quantifiers of three different types: if φ is an EPF_∞-formula, and $\theta(x)$ is an arbitrary formula in $\mathbf{L}_{\infty\omega}$ which has x as its only free variable, then the following formulas are also in EPF_∞: $\forall x \in y\ \varphi$, $\forall x \subseteq y\ \varphi$, $\forall x (\theta(x) \to \varphi)$.

5.4.1 The Extended Positive Comprehension Theorem

Theorem 5.4.2 *($\mathbf{EPF}_\infty$-Comprehension): For every $\varphi \in \mathbf{EPF}_\infty$, having x as a free variable, we have*

$$\exists y \forall x\, (x \in y \Leftrightarrow \varphi)\, .$$

As a consequence, the universe of sets U is closed under the following operations:

- unordered pairs (and singletons): is a, b are sets then $\{a, b\}$ is a set

- pairs: if a, b are sets then (a, b) is a set
- small unions: if $\{a_\beta\}_{\beta<\alpha}$ is a small set of sets then $\bigcup_{\beta<\alpha} a_\beta$ is also a set
- set-unions: if a is some set then $\bigcup a$ is also a set
- powerset: if a is some set then $\mathcal{P}a$ is also a set
- dual powerset: if a is some set then $\{x : x \cap a \neq \emptyset\}$ is a set
- arbitrary intersections: if C is a class of sets then $\bigcap C$ is also a set
- small Cartesian products: if $\{a_\beta\}_{\beta<\alpha}$ is a small set of sets then $\prod_{\beta<\alpha} a_\beta$ is a set; in particular, if a, b are sets then $a \times b$ is a set
- inverse: if a is some set then $a^{-1} = \{(x, y) : (y, x) \in a\}$ is also set
- domain: if a is some set then $dom(a) = \{x : \exists y \text{ s.t. } (x, y) \in a\}$ is a set
- codomain (range): if a is some set then $cod(a) = \{y : \exists x \text{ s.t. } (x, y) \in a\}$ is a set
- (relational) composition: if a, b are sets then $a \circ b = \{(x, z) : \exists y \text{ s.t. } (x, y) \in a \text{ and } (y, z) \in b\}$. The same goes for functional composition.
- image, projections and permutations.

Another consequence is that the following relations and operations are *sets* (in U):

- identity: $I = \{(x, y) : x = y\}$

- membership: $E = \{(x, y) : x \in y\}$
- inclusion: $C = \{(x, y) : x \subseteq y\}$
- singleton operator: $S = \{(x, \{x\}) : x \text{ is a set}\}$
- pair operator: $P = \{(x, y, (x, y)) : x, y \text{ are sets}\}$
- binary union operator $Un_2 = \{((x, y), x \cup y) : x, y \text{ are sets}\}$
- set-union operator: $Un = \{(x, \bigcup x) : x \text{ is a set}\}$
- powerset operator: $\mathcal{P} = \{(x, \mathcal{P}x) : x \text{ is a set}\}$
- dual powerset operator
- (binary) Cartesian product operator: $\prod = \{(x, y), x \times y : x, y \text{ are sets}\}$
- the domain operator on relations: $dom = \{(x, dom(x)) : x \subseteq U \times U\}$
- the codomain operator on relations: $cod = \{(x, cod(x)) : x \subseteq U \times U\}$
- the inverse operator on relations: $inv = \{(x, x^{-1}) : x \subseteq U \times U\}$
- the relational composition operator: $comp = \{((x, y), x \circ y) : x, y \subseteq U \times U\}$
- the image operator, the projections, the permutations and groupings of n-tuples.

The rest of this chapter is devoted to the proof of the above Theorem on $\mathbf{EPF}_\infty$-Comprehension.

Lemma 5.4.3 *If , then we have: For all sets $a_1, \dots, a_n, b_1, \dots, b_n$ ($n \in N$), the following are equivalent:*

- $a_i \equiv_\alpha b_i$ *for every* $i = 1, \dots, n$
- $\{a_1, \dots, a_n\} \equiv_{\alpha+1} \{b_1, \dots, b_n\}$
- $(a_1, \dots, a_n) \equiv_{\alpha+n} (b_1, \dots, b_n)$

Proof: Easy, using the recursive conditions satisfied by the relations $\equiv_\alpha$. ■

Definition 5.4.4 For every class (unary formula) C and every ordinal α, define a class

$$C_\alpha =: \{x : \exists y \equiv_\alpha x \text{ s.t. } y \in C\}.$$

More generally, for every n-ary formula $\varphi(x_1, \dots, x_n) \in L_{\infty,\omega}$ and every ordinal α, define a formula

$$\varphi_\alpha(x_1, \dots, x_n) =: \exists y_1 \equiv_\alpha x_1 \dots \exists y_n \equiv_\alpha x_n \; \varphi(y_1, \dots, y_n).$$

Lemma 5.4.5 *For every class C, every n-ary formula $\varphi(x_1, \dots, x_n) \in L_{\infty,\omega}$ and every ordinal α, we have the following:*

1. *the class C_α is a set, which moreover is definable by a single modal sentence. More generally, the formula φ_α is "essentially modal", in the following sense: there exists some modal sentence θ, such that the formula $\varphi_\alpha(x_1, \dots, x_n)$ is equivalent to the formula $(x_1, \dots, \dots) \models \theta$.*

2. *$C \subseteq C_\alpha$. More generally, φ implies each φ_α.*

3. $\overline{C} = \bigcap_{\alpha \in On} C_\alpha$.

4. *C is a set iff $C = \bigcap_{\alpha \in On} C_\alpha$. More generally, $\{(x_1, \ldots, x_n) : \varphi(x_1, \ldots, x_n)\}$ is a set iff the formula $\varphi(x_1, \ldots, x_n)$ is equivalent to the conjunction of all φ_α's (i.e. with the theory $\{\varphi_\alpha : \alpha \in On\}$).*

This suggests the following

Definition 5.4.6 A formula $\varphi \in L_{\infty\omega}$ is called *comprehensible* if it is equivalent to the theory $\{\varphi_\alpha : \alpha \in On\}$.

Corollary 5.4.7 *Let $\varphi(x_1, \ldots, x_n) \in L_{\infty,\omega}$ be a formula having all its free variables among $x_1, \ldots, x_n$. Then the following are equivalent:*

1. *φ is comprehensible*

2. *there exists some set a such that we have*

$$\forall x_1, \ldots, x_n((x_1, \ldots, x_n) \in a \iff \varphi(x_1, \ldots, x_n)).$$

So to prove the above Theorem, it is enough to prove the following

Lemma 5.4.8 *Every formula $\varphi(x_1, \ldots, x_n) \in \mathbf{EPF}_\infty$ is comprehensible.*

Proof: We need to show that each φ is implied by the conjunction of all the φ_α's (since the converse implication is always true). We do it by induction on the complexity of formulas in $\mathbf{EPF}_\infty$:

- For formulas $\varphi(x_1, \dots, x_n) = x_i \in x_j$: The interesting case is when $i, j \in \{1, , \dots, n\}$. Fix some sets $b_1, \dots, b_n$. Suppose that all the formulas $\varphi_\alpha(b_1, \dots, b_n)$ are true. But this means for every α there exist some $b_{\alpha,i} \equiv_\alpha b_i$, $b_{\alpha,j} \equiv_\alpha b_j$ such that $b_{\alpha,i} \in b_{\alpha,j}$. In particular, for every $\alpha + 1$, we have $b_{\alpha,j} \equiv_{\alpha+1} b_j$. By the bisimulation-like recursive conditions for $\equiv_{\alpha+1}$ and from the fact that $b_{\alpha+1,i} \in b_{\alpha+1,j}$ we conclude that there exists some $c_{\alpha,i} \in b_j$, such that $c\alpha, i \equiv_\alpha b_{\alpha+1,i}$. So, by transitivity, we have $c_{\alpha,i} \equiv_\alpha b_i$. Hence we have:

$$\forall \alpha \exists c_{\alpha,i} \in b_j \text{ s.t. } c_{\alpha,i} \equiv_\alpha b_i.$$

  This means there exists a sequence $\{c_{\alpha,i}\}_\alpha \subseteq b_j$ such that $\lim_{\alpha \to \infty} c_{\alpha,i} = b_i$. But b_j is a set, and so it is closed under limits, hence $b_i \in b_j$, i.e. the formula $\varphi(b_1, \dots, b_n)$ is true, as desired.

- The inductive steps for infinitary conjunctions and disjunctions are trivial.

- The step for formulas of type $\exists x_i \varphi(x_1, \dots, x_n)$: The interesting case is when $i \in \{1, \dots, n\}$. By induction hypothesis, we assume φ is comprehensible. So, by the above Corollary, there exists some set a such that

$$\forall x_1, \dots, x_n(\varphi(x_1, \dots, x_n) \iff (x_1, \dots, x_n) \in a).$$

  Fix some sets $b_1, \dots, b_n$ and assume that all formulas $\varphi_{(\alpha)}(b_1, \dots, b_n)$ are true. This means that for each α there exist some $b_{\alpha,1} \equiv_\alpha b_1, \dots, b_{\alpha,n} \equiv_\alpha$

b_n, such that we have $\exists x_i \varphi(b_{\alpha,1}, \ldots, x_i, \ldots, b_{\alpha,n})$. So there must exist some c_α such that the formula $\varphi(b_{\alpha,1}, \ldots, c_\alpha, \ldots, b_{\alpha,n})$ is true. By the above hypothesis for φ, it follows that $(b_{\alpha,1}, \ldots, c_\alpha, \ldots, b_{\alpha,n} \in a$. Let now c be an accumulation point of the sequence $(c_\alpha)_\alpha$. This means there exist arbitrarily large α such $c \equiv_\alpha c_\alpha$. So we obtain that, for arbitrarily large α, we have

$$b_{\alpha,1} \equiv_\alpha b_1, \ldots, b_{\alpha,n} \equiv_\alpha b_n \text{ and also } c_\alpha \equiv_\alpha c,$$

hence $(b_{\alpha,1}, \ldots, c_\alpha, \ldots, b_{\alpha,n}) \equiv_\alpha (b_1, \ldots, c, \ldots, b_n)$. So $(b_1, \ldots, c, \ldots, b_n)$ is a limit of a sequence of tuples in a, and hence it must belong to a (since a is a set). By the above hypothesis on φ, this implies that $\varphi(b_1, \ldots, c, \ldots, b_n)$ is true, and so we have $\exists x_i \varphi(b_1, \ldots, b_n)$.

- The steps for formulas of types $\forall x_i \in x_j \varphi(x_1, \ldots, x_n)$, $\forall x_i \subseteq x_j \varphi(x_1, \ldots, x_n)$ and $\forall x\, (\theta \rightarrow \varphi)$ are similar, but easier, since we do not need an accumulation point.

■

This finishes the proof of the Extended Positive Comprehension Theorem. As mentioned above, this a strengthening of the Generalized Positive Comprehension Principle, proposed by Malitz and Weydert, and proved to be consistent by Forti ([Forti1989]), using a topological model.

Corollary 5.4.9 *Every unary formula in* $\mathbf{EPF}_\infty$ *is equivalent in STS to some infinitary modal theory.*

The converse of this corollary is trivially true: every definable modal theory T has the form $T = \{\varphi(x) : x \in C\}$, where C is a definable class and φ is a formula such that each $\varphi(a)$ is a modal sentence, for each $a \in C$. Let $\theta(x)$ be the unary formula that defines the class C. Then the theory T is equivalent to the formula $\forall x(\theta(x) \rightarrow \varphi(x))$.

Corollary 5.4.10 *The converse of the Extended Positive Comprehension Principle is also true: every set is definable by a formula in* $\mathbf{EPF}_\infty$.

Proof: Use the Converse of Modal Comprehension and the above observation about the converse of the above corollary. ■

5.4.2 Relations and Functions

We turn now to the study of those relations and functions that are *sets*. We skip the proofs for the results in this section: all of them are easy applications of EPF-Comprehension and of the above topological characterizations of sets.

Definition 5.4.11 An n^{ary} *relation* is an element of $\mathcal{P}(U^n)$. An n^{ary} *class-relation* is a class of n-tuples, i.e. a subclass of U^n.

Not any class-relation is a relation. All the small ones are, of course. But, as observed above, the Generalized Comprehension Principle implies that many other natural set-theoretical class-relations are relations: in particular, we have the *membership relation*$\in$, the *inclusion relation* $\subseteq$ and the *identity relation* $=$.

Proposition 5.4.12 *Let $R \subseteq U^n$ be a class-relation. The following are equivalent:*

1. *R is a relation (i.e. a set)*
2. *R is defined by a $\mathbf{EPF}_\infty$-formula $\varphi(x_1, \ldots, x_n)$*
3. *R is closed as a class with the set topology on U*
4. *R is closed in the product topology induced by the set topology on $U \times U$*
5. *R is closed under limits: if $\lim_{\alpha\to\infty} a_{\alpha i} = a_i$ for every $i \in \{1, \ldots, n\}$ and $R(a_{\alpha 1}, \ldots, a_{\alpha n})$ holds for every α, then $R(a_1, \ldots, a_n)$ holds*
6. *R is closed under approximation: if $a_{\alpha i} \equiv_\alpha a_i$ for every $i \in \{1, \ldots, n\}$ and $R(a_{\alpha 1}, \ldots, a_{\alpha n})$ holds for every α, then $R(a_1, \ldots, a_n)$ holds*

Proof: Easy consequence of the corresponding results for sets. ■

As an application, we get another important example of relations:

The *indistinguishability relations* $\equiv_\alpha$ are sets.

Another application is:

Proposition 5.4.13 *The universe of relations $Rel = \bigcup_{n\in N} \mathcal{P}(U^n)$ is closed under the following operations: composition, inverse, restriction to a set, cylindrifications, small unions, arbitrary intersections; the domain and codomain of a relation are sets. Given an equivalence relation R, any equivalence class $[x]_R$ is a set. For every relation R and every object a, the class of all predecessors $\{x : xRa\}$ and the class of all successors $\{x : aRx\}$ of a are sets.*

Also, observe that the class $Rel_n = \mathcal{P}(U^n)$ of a given arity is a set. The class $Rel = \bigcup_n Rel_n$ of all relations is a set.

Functions and Operators

Definition 5.4.14 An *operator* is a class-function, i.e. a binary class relation R which is functional: if Rab and Rac then $b = c$. A *function* is an operator which is a set in U. An operator is *total* if its domain is the universe U, i.e. if for every set a there exists some b s.t. aRb.

We use the functional notation for operators and functions.

Examples: From the results on $\mathbf{EPF}_\infty$-Comprehension above, it follows that the following operators are functions: powerset $\mathcal{P}$, union, singleton, identity, pairing, Cartesian product, domain, codomain, relational composition are all functions. We denote by id the identity relation $=$, seen as a function:

$$id(x) = x \text{ for every set } x.$$

Examples of non-functions: As we shall prove later (in the section on L-Extended Comprehension), the intersection operator $a \longmapsto \bigcap a$ is not a function. Moreover, the binary intersection operator $(a, b) \longmapsto a \cap b$ is not a function.

Proposition 5.4.15 *Let Γ be an operator. The following are equivalent:*

1. *Γ is a function*

2. *the domain of* Γ *is a set and* Γ *is continuous in the set topology, i.e. for every* $x \in dom(\Gamma)$ *and every ordinal* $\varepsilon \in On$ *there exists some ordinal* $\delta \in On$ *such that for all* $y \in dom(\Gamma)$, *if* $x \equiv_\delta y$ *then* $\Gamma(x) \equiv_\varepsilon \Gamma(y)$

3. *the domain of* Γ *is a set and* Γ *is sequential-continuous, i.e. if* $\lim_{\alpha\to\infty} a_\alpha = a$ *and all* a_α*'s are in the domain of* Γ, *then* $\lim_{\alpha\to\infty} \Gamma(a_\alpha) = \Gamma(a)$

4. *the domain of* Γ *is a set and* Γ *is uniformly continuous, i.e. for every* $\varepsilon \in On$ *there exists some* $\delta \in On$ *such that for all* $x, y \in dom(\Gamma)$, *if* $x \equiv_\delta y$ *then* $\Gamma(x) \equiv_\varepsilon \Gamma(y)$.

Corollary 5.4.16 *The composition of two functions is a function. The inverse of a bijective function is a bijective function.*

Proposition 5.4.17 *Relational composition between binary relations*

$$(R, R') \longmapsto R \circ R'$$

is a function with domain $Rel_2 \times Rel_2$ *and codomain* Rel_2. *The relational inverse operator on binary relations* $R \longmapsto R^{-1}$ *is a function. Similarly for relations of arbitrary arity.*

If F *is a* set of functions *then the restrictions to* F *of the functional composition and functional inverse operators are (possibly partial) functions.*

The problem with the last result is that we don't know yet what classes of functions are sets.

Definition 5.4.18 The *degree of an operator* Γ is an operator $Deg_\Gamma : On \longrightarrow On$ on ordinals, defined by:

$$Deg_\Gamma(\varepsilon) =: \text{ the least } \delta \in On \text{ s.t. } \forall x, y(x \equiv_\delta y \Longrightarrow \Gamma(x) \equiv_\varepsilon \Gamma(y)).$$

Observe that an operator Γ has a degree iff is uniformly continuous in the set topology. As a consequence we have:

Proposition 5.4.19 *Every function f has a degree operator Deg_f. Moreover, an operator has a degree iff it is a function. The degree operator Deg_f is always monotonic and continuous in the order-theoretic sense, i.e. it permutes with supremum (union). Hence, the degree operator has arbitrarily large fixed points.*

We consider degrees with the natural order given by

Definition 5.4.20 For operators $F, G : On \longrightarrow On$, we write $F \leq G$ if we have $F(\alpha) \leq G(\alpha)$ for every ordinal α.

Definition 5.4.21 Given some operator $F : On \longrightarrow On$, we say that a function f is *bounded by F* if it has degree $Deg_f \leq F$.

A function f is *non-expansive* if it is bounded by the identity operator id, i.e iff

$$x \equiv_\alpha y \text{ implies } f(x) \equiv_\alpha f(y)$$

for all sets x, y.

A function is said to be *bounded* if it bounded by the operator $id + \omega$, where

$$(id + \omega)(\alpha) =: \alpha + \omega.$$

So a function is bounded iff

$$x \equiv_{\alpha+\omega} y \text{ implies } f(x) \equiv_{\alpha} f(y),$$

for all sets x, y. One can check that this is equivalent to the condition that:

$$x \equiv_{\alpha+\omega} y \text{ implies } f(x) \equiv_{\alpha+\omega} f(y),$$

for all sets x, y. Clearly, every non-expansive function is bounded.

Proposition 5.4.22 *The following functions are non-expansive:*

- *identity*
- *singleton*
- *powerset*
- *binary union* $(a, b) \longmapsto a \cup b$
- *Cartesian product* $(a, b) \longmapsto a \times b$
- *all permutations of* n*-tuples are non-expansive*

Also, the composition of two non-expansive functions is non-expansive. The restriction of a non-expansive function to a set is non-expansive.

Proposition 5.4.23 *The following functions are bounded:*

- *all the above*
- *union* $a \longmapsto \bigcup a$

- *domain* $a \longmapsto dom(a) = \{x : \exists y \ s.t. \ (x, y) \in a\}$
- *codomain*
- *projections*
- *transitive closure*

The composition of two bounded functions and the restriction of a bounded function to a set are bounded.

Moreover, the composition operator restricted to bounded functions is bounded.

Definition 5.4.24 A bounded function f is said to be *boundedly injective* if it has a bounded inverse; i.e. it exists some bounded function g such that $g \circ f = id_{dom(f)}$. Equivalently, f is boundedly injective iff

$$f(x) \equiv_{\alpha+\omega} f(y) \text{ implies } x \equiv_\alpha y,$$

for all sets x, y.

A function is a *bounded bijection* if it is bounded, surjective and boundedly injective. A function is a *bounded permutation* of some set a if it is a bounded bijection from the set a onto itself.

5.4.3 The Problem of the Exponential

Definition 5.4.25 Given two sets a, b, the *exponential* a^b is the class of all functions having b as domain and a as codomain:

$$a^b =: \{f : f \text{is a function} : b \longrightarrow a\}.$$

Theorem 5.4.26 (Forti) a^b is a set iff b is small.

So, unlike most other set-theoretical operations, the exponential is not always defined: our universe of sets U is not closed under this operation. There is no set U^U of all total functions; moreover, there is no set $U \rightharpoonup U$ of all partial functions. Even worse, the exponential does not exist except in the trivial case in which the exponent is small.

So we have to look for substitutes for exponentiation. An obvious candidate would be *the denotation* (closure) of the exponential $d(a^b) = \overline{a^b}$. Of course, this would also contain some non-functional relations between b and a. But one might hope these relations are somehow "quasi-functional" or particularly well-behaved. Unfortunately, this is *not* the case in general, and in perhaps the most important example: $d(U^U)$.

Proposition 5.4.27 *Every binary set-relation $r \in \mathcal{P}(U \times U)$, whose restriction to V_{AFA} is functional, belongs to the closure $\overline{U^U} = d(U^U)$ of the class of all total functions.*

So "anything goes outside V_{AFA}", with the proviso that we need to ensure sethood (so our relation r cannot be an arbitrary class-relation, but needs to be continuous). This is very bad: for instance, $d(U^U)$ is closed under infinitary unions. Hence there is no way to think of them as being "functions" in any reasonable sense.

Observe though that there are cases in which the closure of the exponential works better. One such nice example is $d(On^{On}) = \overline{On^{On}}$. Its members are relations in $\mathcal{P}(\overline{On} \times \overline{On})$, which are functional except maybe for the point $\infty = \overline{On}$; and they are are always non-functional for a reason, namely the multiple small values at ∞ express the fact that the function takes those values for arbitrarily large ordinals (while the presence of infinite value ∞ at infinity expresses the fact function takes arbitrarily large ordinal values).

But we are looking now for a more general solution to the exponential problem.

Theorem 5.4.28 *Let F be a monotonic, order-continuous operator on ordinals such that $F \geq id$ (i.e. $F(\alpha) \geq \alpha$ for every α). The class of all total functions (on U) that are bounded by F is a set. Similarly, the class of all partial functions that are bounded by F is a set. More generally, for all sets a, b, the class of all total (or partial) functions from a to b which are bounded by F is a set.*

Proof: We prove it for the most general case, that of the exponential a^b, and the others will follow. We show the mentioned class is closed under sequential limits. For this, let $(f_\alpha)_{\alpha \in On}$ be a convergent sequence of functions from a to b that are bounded by F. Let r be the limit of this sequence. We can safely assume that we have $f_\alpha \equiv_\alpha r$ for all α (else, we can take a subsequence). It is easy to see now that this implies that r is a relation in $\mathcal{P}(b \times a)$.

Fix some arbitrary ordinal α. Let $(a, b), (a', b') \in r$ such that $a \equiv_{F(\alpha)} a'$. We shall prove this implies that $b \equiv_\alpha b$. Observe that would give both functionality and boundedness by F.

Since $f_{F(\alpha)+3} \equiv_{\alpha+3} r$, we can match the pairs $(a,b),(a',b') \in r$ by some $(a_\alpha, b_\alpha),(a'_\alpha, b'_\alpha) \in f_{\alpha+3}$ such that $(a_\alpha, b_\alpha) \equiv_{F(\alpha)+2} (a,b)$ and $(a'_\alpha, b'_\alpha) \equiv_{F(\alpha)+2} (a',b')$. But this implies that we have: $a_\alpha \equiv_{F(\alpha)} a$, $b_\alpha \equiv_{F(\alpha)} b$, $a'_\alpha \equiv_{F(\alpha)} a'$, $b'_\alpha \equiv_{F(\alpha)} b'$. But, by the way chose our pairs in $f_{F(\alpha)+3}$, we have $f_{F(\alpha)+3}(a_\alpha) = b_\alpha$,$f_{F(\alpha)+3}(a'_\alpha) = b'_\alpha$. By the fact that all f's are bounded by F, we obtain that

$$b \equiv_{F(\alpha)} b_\alpha \equiv_\alpha b'_\alpha \equiv_{F(\alpha)} b'.$$

Using the fact that $F(\alpha) \geq \alpha$, we conclude that $b \geq b'$. ■

In particular, there is a set of non-expansive functions and a set of bounded functions. We propose the latter as a good enough approximation of the exponential.

Definition 5.4.29 For sets a, b, define the *bounded exponential* to be the set of all bounded (i.e bounded by $id+\omega$) functions from b to a. We use both the set-theoretical notation $exp_{bd}(a,b)$ and the domain-theoretic notation $b \rightarrow_{bd} a$ to denote the bounded exponential.

We also consider the set of all bounded *partial* functions from b to a and we denote it by $b \rightharpoonup_{bd} a$.

Also, define the *bounded symmetric group* $S_{bd}a$ of a set a as the set of all bounded permutations of the set a.

The set $Bdf =: U \rightharpoonup_{bd} U = \bigcup_{a,b} exp_{bd}(a,b)$ is the set of all bounded functions. It coincides with the set of all restrictions to sets of the functions in $U \rightarrow_{bd} U$.

The (bounded) exponential operator $exp : U \times U \longrightarrow U$ is defined by:

$exp(a, b) =: b \rightarrow_{bd} a$.

The *bounded application operator app* is a partial function on $Bdf \times U$, such that $app(f, x) =: f(x)$, whenever $x \in dom(f)$.

We have seen that a lot of set-theoretical functions are bounded. Now we can add more:

Proposition 5.4.30 *The (bounded) exponential operator exp_{bd} is bounded (and hence is a function). The bounded application operator is bounded (and hence is a function). The composition operator restricted to bounded functions $Comp : Bdf \times Bdf \longrightarrow Bdf$, given by $Com(f, g) = f \circ g$, is bounded. The inverse operator restricted to boundedly injective functions is bounded. The bounded symmetric group $S_{bd}a$ of any set a is indeed a group, if considered with the composition and inverse operators.*

So the notions of "bounded function" and "bounded exponential" seem to be very well-behaved approximations of their unbounded versions. The bounded exponential contains most natural set-theoretical functions. Moreover, as we shall see in the model theory section, one can show that all the functions used outside set-theory will have bounded "copies": every first-order structure is isomorphic to a "bounded" structure (endowed only with bounded functions).

One can observe though that, in case one's preferred function f happens to be unbounded, the above theory can be easily extended to cover it: just replace the above notion of "bounded exponential" with "the set of all functions bounded by F", where is any monotonic, order-continuous operator on the ordinals, which takes as values only limit-ordinals. (One can take, for instance $F = deg_f + \omega$, to include our desired function f in our new exponential.) All the results above hold in this more general context.

5.5 Further Extension of the Comprehension Theorem

In this section, we extend our "Extended Comprehension Theorem" to allow defined terms and relations in the comprehensible formulas. We only state here the theorem and present some examples. The proof is an easy application of the Extended Comprehension Theorem above and of the above results on functions and relations.

Let L be a *vocabulary*, given by a set $Const$ of constant symbols, a set Var of variables, a set $Funct$ of function symbols and a set Rel of relation symbols. Each of the function and relation symbols comes with an *arity*, which is a natural number. We denote by $Funct_n$ ($Pred_n$) the set of function (relation) symbols of arity n.

The *terms* of L are defined in the usual recursive manner: variables and constants are terms; if $t_1, \dots, t_n$ are terms and $f \in Funct_n$, then $f(t_1, \dots, t_n)$ is a term.

The *class* of *set-theoretic infinitary formulas* of L is defined in the metatheory by the usual conditions: if $t_1, \dots, t_n$ are terms and $R \in Rel_n$, then $t_1 \in t_2, Rt_1 \dots t_n$ are formulas; if Φ is a set of formulas then $\bigwedge \Phi$ is a formula; if φ is a formula and x is a variable, then $\neg\varphi$ and $\forall x \varphi$ are formulas.

One defines inclusion $\subseteq$, disjunction $\bigvee$, implication and logical equivalence, existential quantifier $\exists$, bounded quantifiers $\forall x \in y, \exists x \in y, \forall x \subseteq y$ as abbreviations, in the usual way. One can also define the notions of *occurrence of a variable in a term or formula, free occurrence, bounded occurrence* etc. As usually, we restrict ourselves to formulas having only *finitely many free variables* and we denote the class of all these formulas by $L_{\infty\omega}$.

An *interpretation* for L is an operator I such that: for every $c \in Const$, $I(c)$ is defined (as a set $I(c) \in U$); for every function symbol $f \in Funct_n$, $I(f)$ is defined and is an n^{ary} *function*, i.e. a *set* $I(f) \in U^{(U^n)}$; for every relation symbol $R \in Rel_n$, $I(R)$ is defined and is an n^{ary} *(set-)relation*, i.e. a set $I(R) \in \mathcal{P}(U^n)$.

A *valuation* for L in the universe of sets is a function $v : Var \to U$. Given a valuation v, we can extend any interpretation I to an *interpretation function* v^I defined on all the terms of L, in the following way:

$$\begin{aligned} v^I(x) &= v(x) \text{ for } x \in Var, \\ v^I(c) &= I(x) \text{ for } c \in Const, \\ v^I(f(t_1, \ldots, t_n) &= I(f)(v^I(t_1), \ldots, v^I(t_n)) \text{ for } f \in Funct_n \text{ and } t_1, \ldots, t_n \text{ terms.} \end{aligned}$$

The *truth* of a formula φ, relative to a given interpretation I and to a given valuation v, is defined in the usual manner:

$$
\begin{array}{rcl}
t_1 \in t_2 \text{ is true} & \text{iff} & v^I(t_1) \in v^I(t_2) \\
R(t_1, \ldots, t_n) \text{ is true} & \text{iff} & (v^I(t_1), \ldots, v^I(t_n)) \in I(R) \\
\neg\varphi \text{ is true} & \text{iff} & \varphi \text{ is not true (for the same } I \text{ and same } v) \\
\bigwedge \Phi \text{ is true} & \text{iff} & \text{every formula in } \Phi \text{ is true} \\
 & & \text{(for the same } I \text{ and same } v) \\
\forall x\varphi \text{ is true} & \text{iff} & \varphi \text{ is true for the same } I \text{ and for every} \\
 & & v' \text{ s.t. } v'(y) = v(y) \text{ for all } y \neq x.
\end{array}
$$

A formula is *valid* if it is true for every interpretation I and every valuation v.

The class of *L-extended positive formulas* is defined by recursion in the metatheory, as the least subclass of $L_{\infty\omega}$ that contains all the atomic sentences and is closed under infinitary conjunction and disjunction, existential quantifier and the following three kinds of bounded universal quantifiers: $\forall x \in t\varphi, \forall x \subseteq t\varphi, \forall x(\theta(x) \to \varphi)$ (where t is a term in which the variable does not occur, $\theta(x)$ is an arbitrary formula having x as its only variable and φ is some L-extended positive formula).

Theorem 5.5.1 ***(L-Extended Positive Comprehension):*** *Let varphi be some L-extended positive formula and t be some term, both having the variables $x_1, \ldots, x_n$ free and not containing any occurrence of the variables x, y. Then the formula*

$$\exists y \forall x(x \in y \Leftrightarrow \exists x_1 \ldots x_n(x = t \wedge \varphi))$$

is valid.

This theorem allows us to introduce the notation $\{t : \varphi\}$ to denote the set z above, for every L-positive formula φ, every term t and every interpretation I. So, once we have checked that some operators are *functions* and some relations are *set-relations*, we can use them to build terms that are allowed to appear in comprehensible formulas. So, in particular, one can introduce terms of the form $\mathcal{P}t, \bigcup t, \{t, t'\}, (t, t'), t \cup t', t \times t', dom(t), cod(t), TC(t), t \circ t', exp_{bd}(t, t')$ and relations like $\subseteq, =, \equiv_\alpha$.

Examples:

$\{(x, y) : \mathcal{P}x \in y\}$

$\{\mathcal{P}x : \forall y \in \bigcup x \ (y, x) \in \mathcal{P}y\}$

$\{(f, g, f \circ g) : \exists x, y, z(f \in exp_{bd}(x, y) \text{ and } g \in exp_{bd}(y, z))\}$.

Using this theorem, we can also prove that some classes are *not* sets.

Corollary 5.5.2 *The intersection operator* $a \longmapsto \bigcap a$ *is not a function. Moreover, the binary intersection operator* $(a, b) \longmapsto a \cap b$ *is not a function.*

Proof: Suppose that the binary intersection operator were a function. We know that the singleton operator is also a function. Then it would follow that the term $x \cap \{x\}$ can be allowed in extended positive formulas, so that one can form the set

$$\{x : x \cap \{x\} = \emptyset\}.$$

But this cannot be a set, since it coincides with the Russell class $\{x : x \notin x\}$ (which is not a set, by Russell's Paradox). This contradiction shows binary intersection is not a function.

Similarly, one can show the infinitary intersection operator is not a set, since the class $\{x : \bigcap\{x, \{x\}\} = \emptyset\}$ coincides with the Russell class. ■

Similarly, one can show that

Corollary 5.5.3 *The* difference relation $D = \{(x, y) : x \neq y\}$ *is not a set. In general, no Frege cardinal* $\{x : |x| = \kappa\}$ *is a set, for any* $\kappa \geq 2$.

Proof: We only show D is not a set (the proof for Frege cardinals is similar). Suppose, towards a contradiction, D were a set. Then, by the Extended Comprehension Theorem, the class $\{(x, y) : \forall y \in x (x, y) \in D\}$ would be a set. But it is easy to see that this coincides with the Russell's class. ■

Chapter 6

Applications

6.1 Solving Equations

6.1.1 Flat Systems of Equations

One of the goals of our theory is to be able to solve systems of equations involving set operations. We start with the simplest possible case.

Definition 6.1.1 A *(generalized) flat system of equations* is a triplet $\mathcal{E} = \langle \mathcal{X}, \mathcal{A}, \mathcal{E} \rangle$ of classes, such that X and A are disjoint classes and E is a class-function $E : X \longrightarrow \mathcal{P}(\mathcal{X} \cup \mathcal{A})$. In the context of a flat system, the elements of A are called *atoms* and the elements of X are called variables For every $x \in X$, we denote by E_x the image $E(x)$. We also put: $B_x =: E_x \cap X$ and $C_x =: E_x \cap A$

A *weak solution* to the flat system $\mathcal{E}$ is a class-function S with domain $dom(S) = X$, such that for all $x \in X$ we have:

$$S_x \equiv \{S_y : y \in B_x\} \cup C_x.$$

A *strong solution* is defined similarly, by requiring equality $=$ instead of observational equivalence $\equiv$, in the above system.

Proposition 6.1.2 *Every flat system with no atoms has a unique weak solution. If the system is small (i.e. if X and A are small) then the weak solution is also a strong solution.*

Proof: Define a binary relation (system) R by:

$$yRx \text{ if } y \in E_x \text{ , for } x \in X, \text{ and}$$

$$yRx \text{ if } y \in x \text{ , for } x \in A.$$

We can immediately see that the restriction of the decoration function d_R to the set A of atoms is the identity function. Then it easily follows that the restriction of the decoration function d_R to the set of variables X gives the solution function. Uniqueness follows from super-strong extensionality. If the system is small, we can use Replacement to show that, for every $x \in X$, the class $\{S_y : y \in B_x\} \cup C_x$ is a set. But S_x is also a set, and they are observationally equivalent, so they have to coincide.

6.1.2 Fixed Point Theorems: the Greatest Fixed Point

Proposition 6.1.3 *If $\Gamma : U \to U$ is a monotonic operator then Γ has a least fixed point $\Gamma_\infty \in U$ and a greatest fixed point $\Gamma^\infty \in U$.*

Proof: We know that $(U, \subseteq)$ is a complete lattice. One can easily check that Tarski's proof of the existence of fixed points in a complete lattice still works in our setting (despite the fact that U is a *large* lattice). This is the proof that constructs the least fixed point Γ_∞ as the greatest lower bound of the Γ-sound points $inf\{x : x \subseteq \Gamma x\}$ and the largest fixed point as the least upper bound of the Γ-correct points $sup\{x : x \subseteq \Gamma x\}$. ■

Note that the usual, more "constructive", proof of the existence of fixed points, using recursion on the ordinal does not work in general, for an arbitrary monotonic operator. But we shall see that it does work for functions (i.e. operators which are *sets*). But first, let us compare our "internal" fixed points (which are sets) with the "real" or external class-fixed points.

Proposition 6.1.4 *If $\Gamma : U \to U$ is a monotonic operator, then:*

1. Γ can be extended to classes, by defining

$$\Gamma =: \bigcup\{\Gamma(x) : x \text{ is a set} \in U \text{ and } x \subseteq C\}.$$

   and the resulting operator is still monotonic.

2. Γ has a *least class-fixed point* $lfp(\Gamma)$ and a *greatest class-fixed point* $gfp(\Gamma)$.

3. $\Gamma_\infty = \bigcap\{x : \Gamma x \subseteq x\}$

4. $\Gamma^\infty = gfp(\Gamma) = \bigcup\{x : x \subseteq \Gamma x\}$.

This means that the largest fixed point inside our universe of sets is the real one:

Corollary 6.1.5 *The largest (class-)fixed point of a monotonic operator is always a set.*

This is not true for least fixed points, e.g. for the powerset operator: $lfp(\mathcal{P}) = WF$, while $\mathcal{P}_\infty = \overline{WF} \neq WF$. So, *in general, the greatest fixed point is better-behaved than the least fixed point.*

For monotonic *functions*, one can construct both the least and the greatest fixed points in On steps:

Theorem 6.1.6 *Let* $\Gamma \in U$, $\Gamma : U \to U$ *be a monotonic* set-*function. Define by recursion on ordinals:*

$$\Gamma_\alpha =: \bigcup_{\beta<\alpha} \Gamma\left(\Gamma_\beta\right),$$

$$\Gamma^\alpha =: \bigcap_{\beta<\alpha} \Gamma\left(\Gamma^\beta\right).$$

Then:

1. $lfp(\Gamma) = \bigcup_\alpha \Gamma_\alpha$
2. $\Gamma_\infty = \overline{lfp(\Gamma)}$
3. $\Gamma^\infty = gfp(\Gamma) = \bigcap_\alpha \Gamma^\alpha$.

We can see that, for monotonic set-functions, the natural recursive process of approximation of the fixed points converges in On steps. The reason this recursive process converges to a fixed point is *not* that the recursion would close off at some ordinal (as in ZFC). Our set-function Γ might be large (e.g. the powerset function $\mathcal{P}$), in which case no ordinal will suffice. But, as long as Γ is a set, the recursive process will reach a fixed point in On steps. Another interesting thing is that, unlike in the case of ZFC and ZFA, our theory relates recursion and corecursion in a simple, symmetrical manner: given a monotonic operator on sets, one can approximate the greatest fixed point by a descending sequence, dual to the one that approximates the least fixed point. This is only possible because of the presence of very large, "over-comprehensive" sets, which nevertheless remain "well-behaved" from a set-theoretical point of view.

6.1.3 Domain Equations

In the semantics of programming languages, one needs to solve *reflexive domain equations* of the following form:

$$X = F(X).$$

Usually, F is assumed to be built by composing elementary operations, like union, Cartesian product, powerset, exponential etc.

There are several known approaches to this problem: Scott domains, de Bakker-Rutten metric domains etc. In all of them, the "real" operations of powerset and exponentiation are replaced by "internal" operators. This is because of the size barrier imposed by Cantor's theorem: in ZFC (and ZFA) we always have $|A| < |\mathcal{P}A|$ and $|A| < |B^A|$ (for $|B| > 2$). This barrier has been lifted from our theory, which makes possible to find "real" solutions to reflexive equations. (This has already been observed by Forti, in the frame of his “hyperuniverses”.)

Indeed, our fixed point above can be used to solve equations of the form $X = F(X)$, for functions F composed of: binary (and small) unions, set-unions, powersets $\mathcal{P}$, dual powersets $\mathcal{P}^{\mathrm{dual}}$, inverse X^{-1}, domain $dom(X)$, codomain $cod(X)$, relational composition $X \circ Y$, projections, image operator, bounded exponentiation with a fixed base $B \to_{bd} X$, bounded partial exponentiation $X \rightharpoonup_{bd} Y$. All these operators are monotonic functions in our universe, and so their compositions are also monotonic functions.

Some Examples: If A, B are sets then the following equations have solutions X that are *sets*:

$$
\begin{aligned}
X &= \mathcal{P}X \text{ (both } \overline{WF} \text{ and } U \text{ are solutions)} \\
X &= \mathcal{P}^{\mathrm{dual}} X \text{ (the set } NWF \text{ of all non-wellfounded sets is the largest solution)} \\
X &= A \cup (B \times X) \\
X &= A \cup \mathcal{P}(B \times X) \\
X &= \mathcal{P}(A \times X^{-1}) \\
X &= A \cup (B \to_{bd} X) \\
X &= X \rightharpoonup_{bd} X \\
X &= A \cup (X \rightharpoonup_{bd} X) \\
X &= A \cup dom(X \cup B) \cup (X \rightharpoonup_{bd} B) \\
X &= (cod(X) \cup (X \cup A)^{-1}) \times \mathcal{P}X \\
X &= X \circ (A \cup (B \times X))
\end{aligned}
$$

6.2 Model theory

In this section we explore some elementary notions of model theory inside STS. We plan to develop the subject in a future paper.

6.2.1 Tarski's Paradox: The Universal Model

We proceed to formalize model theory inside our system STS. Let $L = (Var, Pred, Funct, Const) \subseteq V$ be some signature, where Var is a set of variables (usually assumed to be countable), $Pred$ is a set of predicate symbols, $Funct$ a set of function symbols and $Const$ a set of constants. All these sets are assumed to be in V (which is our variable universe in STS, which can denote either WF or V_{AFA}). This doesn't make any difference, apart from the smallness assumption, since our symbols are actually "codes" of the linguistic ones. The elements of $Funct$ and $Pred$ come equipped with some natural number, called their arity. We denote by $Funct_n$, and $Pred_n$, the set of function (relation) symbols of arity n. One can define inside STS the (codes of) the sentences of the first-order language with signature L, in the usual way.

A *model* for the language L is defined in the usual way, as a pair $\mathbf{M} = (M, d)$, where: M is a set and d is a function mapping each $R \in Pred_n$ into some relation $d(R) \in \mathcal{P}(M^n)$, each $f \in Funct_n$ into some function $d(f) \in M^{(M^n)}$ and each $c \in Const$ into some $d(c) \in M$.

A *valuation* on $\mathbf{M}$ is a function $v \in M^{Var}$.

We have some degree of freedom: as is usually done, we considered valuations as being external to the model; we could have included the in the definition of a model; or we can consider the function d as external to the model and instead include in the model a set Rel of relations, in bijective correspondence (via d) with the set $Pred$, etc. We shall sometimes use subscripts d_M, Rel_M etc.

The problem is to define the *satisfaction relation*. As we shall see, there cannot be a unique formula in STS, to define truth or satisfaction of an arbitrary sentence in an arbitrary model. But we can define a formula Sat_n that defines satisfaction for formulas of complexity length n. Sat_n is a ternary relation between models, formulas and valuation, which we shall usually write as $\mathbf{M} \models \varphi[v]$. The definition is by induction on n, with the obvious inductive clauses, starting with:

$$(M) \models R(x_1 \ldots x_n)[v] \Longleftrightarrow: (v(x_1), \ldots, v(x_n)) \in d_{\mathbf{M}}(R),$$

for formulas of complexity 0, etc.

We write informally $\mathbf{M} \models \varphi[v]$, and that "$\varphi$ is satisfied (or true) in the model M by the valuation v", whenever we have $\mathbf{M} \models_n \varphi[v]$ for $n = lh(\varphi)$ being the complexity length of φ. But we should keep in mind that this is not a first-order formula in STS, but a metatheoretic device to refer to a disjunction of infinitely many formulas. For particular classes of models (e.g. small models), this is actually equivalent to some first-order formula. But not in general. One can check that all the usual ways to make this inductive definition into a single formula fail in our setting, because of the failure of

Separation and Replacement for "large" models. Moreover, we shall soon prove that this is not an accident, since the "real" satisfaction relation is not definable in STS.

Given a theory T, i.e. a set of (codes of) sentences in first-order logic, we say that "$\mathbf{M}$ *is a model of the theory* T" if $\mathbf{M} \models \varphi$ for every $\varphi \in T$. Again, this is not a definable formula in STS, but a metaformula. Nevertheless, we can still use it.

In particular, we can take now the language of set theory $L = (Var, \in)$ and we can write the set of (codes of) axioms for STS. We shall also denote this theory by STS. Notice that STS is an infinite system of axioms, since we have stated it using *axiom schemes*. This is similar to the case of Zermelo's ZFC. We can consider models $\mathbf{M} = (M, R)$ of this language, with M some arbitrary set and $R \in \mathcal{P}(M^2)$. And then we can prove the following:

Theorem 6.2.1 ***Strong Reflection Theorem for*** STS: *The theory* STS *has a model. Namely the set* $(U, \in)$ *is a model of our theory.*

This seems to contradict Tarski's theorem. In reality, it does not: we cannot use it to define truth inside the system as "satisfaction in $(U, \in)$", because we have not defined any single satisfaction relation. But then what is the content of the Strong Reflection Theorem?

Recall that ZFC had also infinitely many axioms. In ZFC there is a single definable "satisfaction relation", but no single internal model; by the Reflection Theorems, there were infinitely many "partial internal" models, for any given finite subset of the axioms. The situation in STS is completely dual: there is a single internal model for all the axioms, but no definable general notion of satisfaction; but for every n we can define a formula Sat_n, which gives the satisfaction relation for all formulas of complexity less than n; as we saw, the definition looks pretty uniform from outside the system (it is an induction), and it is clear that it agrees with the external, meta-theoretical, notion of satisfaction for all formulas of lower complexity. But there is no way to write it in a uniform way inside the system.

So the actual content of the above Theorem is a *schema* saying that:

> *For each finite subset T of the axioms of STS, let n be some (externally given) natural number, larger than the complexity of all the sentences in T. Then we have $(U, \in) \models_n \bigwedge T$.*

So we conclude that the system STS has the amazing property that *it provides a definable model for itself.* This model can be seen to be a model from inside the theory: the system can prove that each of its axioms holds in the universal model $(U, \in)$. What the system cannot do is to say that the universal model is a model of *all* its axioms; and this is because there is no uniform notion of "satisfaction (truth) of an arbitrary formula in a model". The non-existence of such a uniform definition can be shown as an application:

Corollary 6.2.2 *Satisfaction is not definable in STS. More precisely, let $\varphi(x, y)$ be some formula such that, for every (external) natural number n, we have*

$$\forall x, y(Sat_n(x, y) \Longrightarrow \varphi(x, y)).$$

Then there exists some pair a, b such that $\varphi(a, b)$ is true, but for every given n the formula $Sat_n(a, b)$ is false.

The proof uses Tarski's Theorem and the above Strong Reflection Theorem.

Another application is the following:

Corollary 6.2.3 *If consistent, the system STS is* not finitely axiomatizable.

Proof: If it were, we could define satisfaction by the formula Sat_n, for some n larger than the complexity of all the axioms. ■

6.2.2 Sets of Models of a Given Theory

We would like to have the classes of models $Mod(T)$ of important first-order theories T as *sets* in our universe. For purely relational languages, this is possible:

> The class of all models of a relational language (no functions) is a set.

But it is clear that we cannot expect the functional languages to have the same property, because of the problems with the exponential operator. So we need to restrict ourselves to *bounded* models.

Definition 6.2.4 A model is *bounded* if the interpretations of all the function-symbols are bounded functions. We denote by $Mod_{bd}(T)$ the class of all models of the theory T.

It is easy to see that, by restricting ourselves to bounded models, we do not lose anything from the model-theoretic point of view, as far as small models are concerned:

Proposition 6.2.5 *Given a small model M, there exists a bounded model isomorphic to M. Moreover, there exist many such isomorphic copies: the class of all bounded models which are isomorphic to M is large.*

Proposition 6.2.6 *By* $\mathbf{EPF}_\infty$*-Comprehension, the following classes of models are sets:*

1. *the set of all bounded models of an arbitrary first-order language;*
2. *the set of all bounded models of an equational theory;*
3. *the set of all bounded models of a positive theory. Here a "positive theory" is a theory whose formulas are built from atoms of the form $Rt_1 \ldots t_n$ and $t_1 = t_2$ (equations), using conjunction, disjunction, quantifiers $\forall, \exists$ and bounded universal quantifiers of the form $\forall x_1 \ldots x_n(Rx_1 \ldots x_n \to \ldots)$ (but no negation).*

Examples:

- the set of all bounded monoids
- the set of all bounded groups
- the set of all bounded rings
- the set of all bounded vector spaces

- the set of all bounded lattices (defined in terms of join and meet, not as partial orders)
- the set of all bounded distributive lattices
- the set of all bounded Boolean algebras
- the set of all binary (pointed) graphs (Kripke structures)
- the set of all (labelled) transition systems
- the set of all pairs (A, R), such that R is a reflexive binary relation on the set A
- the set of all pairs (A, R), such that R is a reflexive symmetric relation on A

Observe that these sets of models contain copies of all the small models of the given type, but also large natural models: the model $(U, \subseteq)$ belongs to the last set above; the models $(U, \cup)$, $(U, \times)$ are bounded monoids etc.

Unfortunately, *transitivity and antisymmetry* are not expressible by $\mathbf{EPF}_\infty$-formulas, and so the class of all partial orders and the class of all equivalence relations are not sets. (But one can define notions of "bounded partial order" and "bounded equivalence relation" for which these classes are sets.)

Similarly, the definitions of important notions like "field" and "integral domain" involve negative assertions (e.g. "$0 \neq 1$"). But in such cases, the negativity is eliminable, by simply fixing some of the constants. If, for instance, we agree to consider *only fields or integral domains having the empty* *set $\emptyset$ as the zero element* 0, then the classes of these algebraic structures are *sets*.

We plan to pursue the study of model theory in STS in a future paper. An interesting subject is finding analogues for the classical model-theoretical constructions for arbitrary (possibly large) models in STS: e.g. one can define a notion of "bounded ultraproduct", by replacing functions with bounded functions in the classical definition.

Other sets of mathematical structures:

One can consider natural classes of mathematical structures which are not first-order definable. Usually, these structures involve functions, and the corresponding class will not be a set. But its bounded version will usually be a set. An important example is:

The set of all bounded topological spaces: Define a *bounded topological space* to be a pair (S, C), such that $C : \mathcal{P}(S) \rightarrow_{bd} \mathcal{P}(S_)$ is a bounded function, called the *closure operator* and satisfying the well-known Kuratowski conditions:

1. $A \subseteq C(A)$ for all $A \subseteq S$
2. $C(A) = C(C(A))$ for all $A \subseteq S$
3. $C(\emptyset) = \emptyset$
4. $C(A) \cup C(B) = C(A \cup B)$.

It is easy to see, by $\mathbf{EPF}_\infty$-Comprehension, that the class of all bounded topological spaces is a set.

6.2.3 Circular Model Theory and Semantical Paradoxes

Why would one be interested in whether or not a natural class of models is a set? One particular reason for this is *reflexivity* or *circularity*: there are special first-order theories, for which the class of all their models can be informally seen to satisfy the theory. If this class is a set, then it will be itself a model of the theory, namely a *circular model*: one whose domain contains itself.

Example. The "(bounded) monoid of all (bounded) monoids" Mon: on the above-mentioned set of all bounded monoids, one can define Cartesian product or direct sum, and check that both these operations are bounded. Hence, for each of these operations, we obtain a circular monoid of all monoids: $Mon_\times$ and Mon_+. We have both $Mon_\times, Mon_+ \in Mon_\times$ and $Mon_\times, Mon_+ \in Mon_+$. Moreover, other natural monoids belong to these models: e.g. the monoid $(U \rightarrow_{bd} U, \circ)$ of all total functions with composition.

One can find other interesting algebraic examples, in which we do not have reflexivity, but "cycles" of length bigger than 1: take the set Grp of all bounded groups; take the group $Aut(Grp)$ of all the bounded automorphisms of the group Grp, with functional composition and inverse as its operations. Obviously, Grp is in the transitive closure of $Aut(Grp)$ (since we consider a homomorphism $f : G \rightarrow H$as a triplet (G, H, f), so in particular an automorphism of Grp has the form (Grp, Grp, f). But $Aut(Grp)$ is itself a *bounded group*, so it is a member of Grp. This gives rise to cycles of length

3: e.g., if we denote by Id_{Grp} the identity automorphism on Grp, we have $Grp \in Id_{Grp} \in Aut(Grp) \in Grp$.

Circular models have been used in the context of ZFA to give new insights into the *classical semantical paradoxes* (see Barwise and Etchemendy 1987, Barwise and Moss 1996). We plan to explore in a future paper the natural analogues of their results in our frame.

6.3 Category Theory

A well-known problem in the foundations of category theory is to find a way to make sense of "large categories" (e.g. the category of all sets) and of "super-large, reflexive categories" (e.g. the category of all categories). Both are forbidden by the limitation-of-size assumption built into ZFC and ZFA.

The second case (reflexive categories) is completely intractable in ZFC. The first case (large categories) is tractable in an indirect way, by using classes; nevertheless, this treatment makes impossible some natural categorial constructions on large categories: e.g only for small categories A, B, we are able to define the exponential category $[\mathbf{A}, \mathbf{B}]$ (having as objects all functors F from $\mathbf{A}$ to $\mathbf{B}$, as morphisms from F to G all natural transformations from F to G, as identities the identity natural transformations, and as composition the composition of natural transformations). For large categories, the exponential category is said to be *illegitimate*.

Sometimes, these problems are solved by adding to ZFC one more layer of objects ("families") on top of "classes" and "sets". A natural thing to do would be to add transfinitely many more layers, and assuming they also satisfy the axioms of ZFC; this is equivalent to asserting the existence of an inaccessible cardinal in ZFC. But of course, this only lifts the problem, without actually solving it. In particular, there will be *no reflexive categories* in any such set theory.

The above-mentioned discussion on circular models suggests that our theory is strong enough to deal with these problems. The resulting category theory will be universal with respect to its *objects*, but it will be restricted with respect to its *morphisms*: only bounded functions are allowed as morphisms and functors.

Definition 6.3.1 A *bounded category* is a sextuple $\mathbf{A} = (\mathcal{O}, hom, id, \circ, dom, cod)$, consisting of:

1. a set $\mathcal{O}$, whose members are called $\mathbf{A}$-*objects*;

2. a bounded function $hom : \mathcal{O} \times \mathcal{O} \to U$; the members of each $hom(A, B)$ are called $\mathbf{A}$-*morphisms from A to B*;

3. a bounded function $id : \mathcal{O} \to U$, such that $id(A) \in hom(A, A)$ for every object $A \in \mathcal{O}$; the morphism $id(A)$ will be denoted by id_A and called the $\mathbf{A}$-*identity on A*;

4. a bounded binary partial function $\circ$ (composition), such that, for all $\mathbf{A}$-objects A, B, C and for all $\mathbf{A}$-morphisms $f \in hom(A, B)$, $g \in hom(B, C)$, the function $\circ$ is defined and $g \circ f \in hom(A, C)$; the morphism $g \circ f$ is called the *composite of f and g*;

5. bounded partial functions *dom* and *cod*, defined on all **A**-morphisms into the set $\mathcal{O}$ of objects; the object $dom(f)$ is called the *domain of* f, while $cod(f)$ is called the *codomain of* f.

The above sets and functions are required to satisfy the following equations:

(a) composition is associative: $h \circ (g \circ f) = (h \circ g) \circ f$, whenever both are defined;

(b) **A**-identities act as identities with respect to composition: $id_B \circ f = f$ and $f \circ id_A = f$, for every $f \in hom(A, B)$;

(c) $dom(f) = A$ and $cod(f) = B$, for every $f \in hom(A, B)$.

Observation: By $\mathbf{EPF}_\infty$-Comprehension, *the class of all bounded categories* Cat *is a set.*

Important Examples

1. The *(bounded) category of all sets* **Set** has: *sets* as objects $\mathcal{O} = U$, *bounded partial functions* as morphisms $hom(A, B) = exp_{bd}(B, A) = A \rightharpoonup_b dB$, *functional composition* as composition $\circ$, the *identity function on* A as the identity id_A, the *functional domain and codomain* as its *dom* and *cod* functions. Notice that the function *hom* is just the bounded exponential function exp_{bd}.

2. The *category of all bounded groups* **Grp** has *bounded groups* as objects and *bounded group homomorphisms* as its morphisms.

3. The *category of vector spaces* **Vect** has vector spaces as objects and *bounded linear maps* as morphisms.

4. The *category of bounded topological spaces* **Top** has bounded topological spaces as objects and *bounded continuous functions* as morphisms.

All these "large" categories are members of the set Cat of all bounded categories.

Definition 6.3.2 If $\mathbf{A}$ and $\mathbf{B}$ are bounded categories, then a *bounded functor from* $\mathbf{A}$ *to* $\mathbf{B}$ is a bounded function F that assigns to each $\mathbf{A}$-object A some $\mathbf{B}$-object $F(A)$ and to each $\mathbf{A}$-morphism $f \in hom(A, A')$ some $\mathbf{B}$-morphism $F(f) \in hom(F(A), F(A'))$, in such a way that:

1. F preserves composition: $F(f \circ g) = F(f) \circ F(g)$, and

2. F preserves identities: $F(id_A) = id_{F(A)}$.

The *identity functor* and the *composition of functors* are defined in the usual way.

Most natural functors between bounded categories are bounded functors: e.g. the forgetful functor, the identity functor, the (covariant and contravariant) hom-functors, the (covariant and contravariant) power-set functors etc.

Definition 6.3.3 The *category of all bounded categories* **Cat** has as objects all the bounded categories, as morphisms from $\mathbf{A}$ to $\mathbf{B}$ all the bounded functors from $\mathbf{A}$ to $\mathbf{B}$, and as composition the usual composition of functors.

Proposition 6.3.4 *The category* $\mathbf{Cat}$ *contains as objects all the above-mentioned categories* $\mathbf{Set}, \mathbf{Grp}, \mathbf{Vect}, \mathbf{Top}, \mathbf{Cat}$. *In particular,* $\mathbf{Cat}$ *is a* reflexive category*:* $\mathbf{Cat} \in \mathcal{O}_{\mathbf{Cat}}$ *is an object of itself.*

One can go on and define a notion of *bounded natural transformation* between bounded functors, and define the *bounded exponential category* $[\mathbf{A}, \mathbf{B}]$, for any two bounded categories $\mathbf{A}, \mathbf{B}$.

Chapter 7

Conclusions

This paper is an attempt to build a set theory on a purely structural view on the concept of set. I make a distinction between a *potential* structure and its actualization into a set (via decoration or closure). I propose an analytical picture, in which objects are analyzed in stages and all we can know about them are their *unfoldings* or *partial descriptions*. A *set* is what is left from this process of analysis: it is the *trace of unfolding* of some possible object, its pattern of analytical behavior. I have a notion of *observational equivalence* between structures, defined as identity of analytical behavior. Sets can be understood as *arbitrary structures modulo observational equivalence*. As collections, sets are *closed*, completed classes, which are as large as their pattern of unfolding allows them. They contain every object which cannot be separated from all their elements at any stage of unfolding. This gives them well-defined boundaries and a clear-cut identity.

I explore the connection between this notion of set and modal logic. Sets can be identified with the maximally consistent theories that characterize them. Sets can also be understood as modally definable classes. This provides a proof (and so a justification) for the Power Set axiom on different grounds than the ones of the classical conceptions.

The universe of sets described has nice fixed-point and closure properties. Recursion and corecursion are related in a simpler manner over this universe than over Aczel's hyperset universe. Some category-theory notions can be stated as objects (sets), not just as classes. The topological aspect comes from the underlying presence of a notion of observational approximation (structures can be "almost bisimilar" up to any ordinal depth). This universe provides models for many recursive and corecursive domains, which could be used as uniform frameworks for giving denotational semantics. This universe of sets seems also to be a good candidate for a general framework to study semantical paradoxes.

Bibliography

Aristotle. 1983. *Physics, Books III and IV.* Trans. Edward Hussey. Oxford University Press.

Abramski, Samson. 1988. A Cook's Tour of the Finitary Non-Well-Founded Sets. Unpublished manuscript.

Aczel, Peter. 1988. *Non-Well-Founded Sets.* CSLI Lecture Notes Number 14. Stanford: CSLI Publications.

van Aken, J. 1986. Axioms for the set-theoretic hierarchy. *Journal of Symbolic Logic* 51(4):992-1004.

Baltag, Alexandru. 1995. Modal Characterizations for Sets and Kripke Structures. Unpublished manuscript.

Barwise, Jon, and John Etchemendy. 1987. *The Liar: An Essay in Truth and Circularity.* Oxford:Oxford University Press.

Barwise, Jon and Lawrence Moss. 1991. Hypersets. In *The Mathematical Intelligencer* 13(4): 31-41.

Barwise, Jon and Lawrence Moss. 1996. *Vicious Circles: On the Mathematics of Non-Wellfounded Phenomena.* CSLI Lecture Notes Number 60. Stanford: CSLI Publications.

Boffa, Maurice. 1968. Graphes extensionneles et axiome d'universalité. *Zeitschrift für Math. Logik und Grundlagen der Math.* 14:329-334.

Boffa, Maurice. 1989. A Set Theory with Approximations. *Jahrubuch der Kurt Gödel Gessellschaft.*

Dauben, J. Warren. 1990. *Georg Cantor: His Mathematics and Philosophy of the Infinite.* Princeton University Press.

Devlin, Keith. 1993. *The Joy of Sets.* Undergraduate Texts in Mathematics. Berlin: Springer Verlag.

Fagin, Ronald. 1994. A Quantitative Analysis of Modal Logic. *Journal of Symbolic Logic* 59(1):209-252.

Forster, T. E. 1995. *Set Theory with a Universal Set*, second edition. Oxford Logic Guides, No. 31. Oxford: Oxford Science Publications.

Forti, Mario and Furio Honsell. 1983. Set Theory with Free Construction Principles. *Annali Scuola Normale Supeiore di Pisa, Classe di Scienze* 10:493-522.

Forti, Mario and Furio Honsell. 1985. Axioms of Choice and Free Construction Principles, *Bull. Soc. Math. Belg.*, series B, 36:69-79.

Forti, Mario and Furio Honsell. 1992. Weak foundation and antifoundation properties of positively comprehensive hyperuniverses. In *L'Antifondation en Logique et en Theeorie des Ensembles*, ed. R. Hinnion. 31-43. Cahiers du Centre de Logique.

Forti, Mario and Furio Honsell. 1992. A General Construction of Hyperuniverses. *Quadermi dell' Istituto di Matematiche Applicate "U. Dini"*, no. 9.

Forti, Mario and R. Hinnion. 1989. The Consistency Problem for Positive Comprehension Principles. *Journal of Symbolic Logic* 54(4):1401-1417.

Garciadiego, A. R. 1992. *Bertrand Russel and the Origins of the Set-theoretical Paradoxes.* Birkhauser Verlag.

Hallet, Michael. 1984. *Cantorian Set Theory and Limitation of Size.* Oxford: Clarendon Press.

Malitz, R. J. 1976. Set Theory in which the Axiom of Foundation Fails. Ph. D. thesis. University of California, Berkley, California. Unpublished.

Moore, A. W. 1990. *The Infinite.* Series: Problems of Philosophy. Routledge.

Moschovakis, Yiannis N. 1994. *Set Theory Notes.* Undergraduate Texts in Mathematics. Amsterdam:North Holland.

Weydert, F. 1989. How to Approximate the Naive Comprehension Principle Inside Of Classical Logic.*Bonner Math. Schriften*, no. 194.

Alexandru Baltag

Date and Place of Birth: January 15 1967, Bucharest, Romania

Degrees awarded:

- Master of Science, Department of Mathematics, University of Bucharest, August 1989
- Doctor of Philosophy, Department of Mathematics, Indiana University, July 1998