7

Non-Well-Founded Set Theory

The approach to set theory that has motivated and dominated the study presented so far in this book has essentially been one of *synthesis*: from an initial set of axioms, we build a framework of sets that can be used to provide a foundation for all of mathematics. By starting with pure sets provided by the Zermelo–Fraenkel axioms, and progressively adding more and more structure, we may obtain all of the usual structures of mathematics. And then, of course, we may make use of those mathematical structures to model various aspects of the world we live in. In this way, set theory may be used to provide ways to model 'mathematical' aspects of our world.

But there is an alternative way to approach set theory, namely in an *analytic* fashion, where we start with all of the various 'mathematical' structures we observe in the world and progressively *strip away* structure until all that is left are pure sets.

As you might expect, there is no *a priori* reason that these two approaches will lead to the same theory of sets. Indeed, some very familiar real-world structures give rise to a dramatically different conception of set from the now-familiar Zermelo–Fraenkel notion.

For example, suppose I try to model set-theoretically the items of information in some information-storage device, say this very book. Let $\mathcal{B}$ be the set of all sets explicitly referred to in this book. Clearly, since $\mathcal{B}$ is referred to in this book (I am just now referring to it), we have

$$\mathcal{B} \in \mathcal{B}.$$

More generally, it is not hard to think up examples of 'real world' sets having closed loops of membership:

$$a_1 \in a_2 \in \ \ldots \ \in a_n \in a_1.$$

Such sets are said to be *circular*. With the growing tendency to apply set-theoretic methods in computer and information science, it is getting steadily harder to avoid having to deal with such sets in a formal and rigorous manner.

Now, in Zermelo–Fraenkel set theory, the Axiom of Foundation explicitly rules out the formation of circular sets or sets having themselves as members. So at the very least, if we are to approach set theory in an *analytic* fashion, in a manner that will, for instance, allow us to capture some of the self-referential structure that arises in information systems, we will have to dispense with this particular axiom. But just how significant a step will this be? Will it, for instance, mean that we shall be working within a framework quite unlike that used in other parts of mathematics?

The answer turns out to be 'no'. Simply dropping the Axiom of Foundation from the axioms of set theory results in practically no change in almost all of present day mathematics (or its applications). The reason is that this axiom is totally irrelevant as far as most applications of set theory are concerned. The kinds of sets that arise in, say, Analysis or Algebra, simply *are*, as a matter of fact, noncircular. No axiom is required to guarantee this. It is really only within set theory itself that the Axiom of Foundation is important.

Thus, in contemplating the introduction of a set theory that violates the Axiom of Foundation, which is what this chapter is all about, we are not starting out along a path that will bring us into conflict with the bulk of current mathematical practice. We shall simply find ourselves using sets of a different nature than those used elsewhere (for different purposes).

Of course, in developing a set theory as a conceptual abstraction from, say, information structures in the world, there may turn out to be other features that *do* conflict with the set theory used elsewhere in mathematics. But as far as is known, this is not the case. Indeed, it is possible to regard the universe of sets described below as an *extension* of the Zermelo–Fraenkel universe, one that enlarges the domain of study to include all those circular sets that the Axiom of Foundation normally excludes from consideration.

In this respect, what we are doing is analogous to the extension procedure that takes you from the real numbers to the complex numbers. New 'numbers' are introduced to enlarge the real number system to a richer structure in which more equations have solutions, etc. No properties of the real numbers are violated by this extension. More things become possible at no cost in terms of existing theory.

So too in our introduction of a 'non-well-founded set theory', as I shall refer to any theory of sets that violates the Axiom of Foundation. Indeed, the analogy with the complex numbers is an even better one. Just as the complex numbers may be *defined* in terms of the real numbers, so too the non-well-founded (or circular) sets of our new theory may be defined in terms of the more familiar, *well-founded* (i.e. noncircular) sets of the

Figure 7.1: Graphical representation of two simple sets.

Zermelo–Fraenkel theory. And just as the 'new' complex number system shares many of the fundamental properties of the 'old' real numbers—for instance, both systems are *fields*—so too the universe of non-well-founded sets will satisfy many of the axioms of the well-founded Zermelo–Fraenkel universe of sets. Indeed, it satisfies *all* axioms except for Foundation.[1]

It should perhaps be pointed out that in the case of an *analytic* approach to set theory, it is quite natural to allow for atomic (i.e. non-set) elements, or *urelements*, entities that may be used in order to construct sets, but which are not themselves analyzed in a set-theoretic fashion. Traditionally, Zermelo–Fraenkel set theory does not allow for the existence of atoms, though it is easy to amend the axioms to do so. I shall denote by ZFCA the theory ZFC amended to allow for atoms.

An excellent illustration of the application of non-well-founded set theory is provided by Barwise and Etchemendy in their book *The Liar* [2], in which they provide a set-theoretic account of the classical Liar Paradox and some other logical paradoxes.

7.1 Set-Membership Diagrams

Consider then, some very simple, circular sets of the kind that might easily arise in a discussion of information storage, say

$$a = \{3, 5\} \quad \text{and} \quad b = \{\text{Zermelo, Fraenkel}\}.$$

We may picture these sets by means of simple diagrams as in Figure 7.1.

The idea in the case of such diagrams is to represent set membership by means of directed line segments. Thus, referring to Figure 7.1, the

[1]Though, as we shall see, the Axiom of Extensionality does not always serve to distinguish non-well-founded sets as it does for well-founded sets, and another axiom will be required in order to overcome this problem.

arrows pointing from the set a to each of the two numbers 3 and 5 indicate that the set a has precisely the two elements 3 and 5, and likewise the arrows pointing from b to the two objects (atoms) 'Zermelo' and 'Fraenkel' represent the fact that the set b consists of precisely these two objects (and is thus a set consisting of two particular people). Thus Figure 7.1 provides an alternative means of indicating the set-theoretic structure of the sets a and b, other than the more familiar notation used above to introduce these sets.

Both notations show what it is that the two sets a and b have in common, as well as the way in which they differ. Any set is, of course, a purely abstract construct. In the case of set a, the elements of this set are themselves also abstract entities. Set b, on the other hand, is an abstract construct built out of two real objects in the world (or rather two objects that at one time did exist in the world). But in both cases, the set-theoretic structure itself is the same: each consists of two objects that are (conceptually) collected together to form a single (abstract) entity. With traditional set notation, this common structure is reflected in the fact that in each case precisely two objects occur between the braces { and }; in Figure 7.1, the obvious isomorphism between the two diagrams indicates the same common structure.

Now, in the case of simple sets like the two above, there seems to be little to choose between the two notations, the traditional and the diagramatic, but when it comes to indicating the hereditary (membership) structure of more complex sets, the diagramatic form can be much easier to understand, allowing as it does for the various membership paths to be traced along the connecting arrows. This is illustrated by Figure 7.2, which gives diagramatic representations of the first four ordinal numbers (under the familiar von Neumann definition used in this book, that takes any ordinal number to be just the set of its predecessors).

Both Figures 7.1 and 7.2 are examples of what are known as *graphs*. The points that occur in a graph, such as the points labeled $a, 3, 5$ in the first graph in Figure 7.1, are generally referred to as *nodes* of the graph, the lines (or arrows) connecting them as *edges*.[2]

In Figure 7.2, the ordinal 0, being the empty set, is depicted by a diagram consisting of a single node with no edges emanating from it. The graph for the ordinal 1, being the singleton set $\{\emptyset\}$, consists of two nodes, the top node depicting the ordinal (set) 1 itself, the node beneath it the single element, $\emptyset$, of that top node. And in the remaining two cases, the top node depicts the ordinal number concerned while the remainder of

[2]Strictly speaking, what we have here are *directed graphs* or *digraphs*, the adjective 'directed' indicating that the edges, being arrows, have a specified direction.

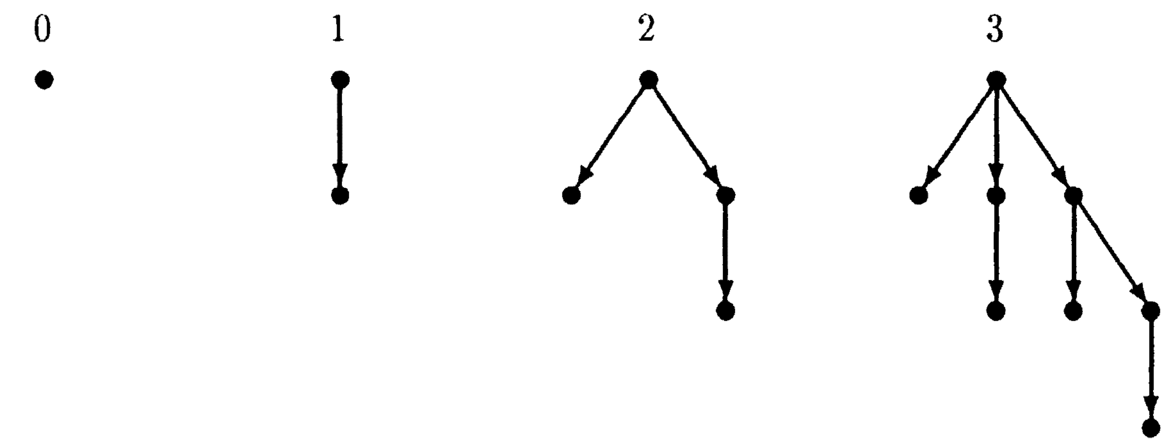


Figure 7.2: Graphical representation of the first four ordinal numbers.

the graph shows the set-theoretic structure of that ordinal number. An instructive exercise is to label each of the nodes in Figure 7.2 with the appropriate von Neumann ordinal.

One thing to notice concerning Figure 7.2 is that there was really no need to label the top nodes in each of the four cases. Since the only set depicted by a node from which no edges (arrows) emanate is the empty set, each of the bottom nodes in the four graphs must represent the empty set, so in each case we may work our way up the various paths through the graph in order to determine the exact nature of the set depicted.

This is quite unlike the situation in Figure 7.1. Here the bottom nodes all denote particular entities, as indicated by the labels attached to those nodes. In the case of the set a, if we regard the elements 3 and 5 as being sets under the von Neumann definition of an ordinal, then of course we may extend this particular graph to one without labels in the obvious way. But for the set b, such a procedure is clearly not possible, and the bottom nodes must be regarded as *atoms* or *atomic nodes* of the graph, depicting entities that either have no set-theoretic structure or whose set-theoretic structure is not pertinent.

In order to avoid confusion, I shall use hollow circles, rather than dots, to indicate atoms in graphs. Thus, the set {Zermelo, 1} will be represented graphically as in Figure 7.3.

If we allow infinite graphs in the case of infinite sets, then it is clear that any set may be represented by a membership graph in this fashion, providing a diagramatic representation of the entire hereditary structure of the set. Indeed, there is an obvious method for producing a graph that

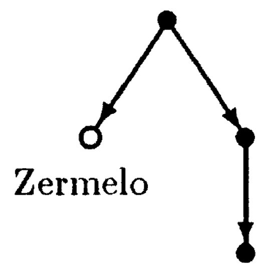


Figure 7.3: The set {Zermelo, 1}, where 1 = {0}.

Figure 7.4: Alternative graphical representations of the ordinals 2 and 3.

depicts a given set.[3] Namely, start with the set concerned as top node, and then enumerate all its elements beneath it, joining the top node to each of these by means of a downward pointing arrow. Then, for each of these nodes in turn enumerate all their members beneath them, and make the appropriate edge-connections. And so on.

Now, a particular set may be represented by more than one graph. For instance, referring back to Figure 7.2, in the graph depicting the ordinal number 2 there are two nodes denoting the ordinal number 0. If we identify these two nodes then we obtain the alternative graphical representation of the ordinal 2 shown on the left of Figure 7.4. Likewise, the graph depicting the ordinal 3 in Figure 7.2 has four nodes that correspond to 0 and two corresponding to the ordinal 1, and identification of the nodes in these two groupings leads to the graph shown on the right in Figure 7.4.

Again, it is an instructive exercise to label each of the nodes in Figure 7.4 with the appropriate ordinal number and to relate these two graphs with the corresponding graphs in Figure 7.2.

By allowing the appearance of loops within graphs it is possible to depict (some) non-well-founded sets by means of finite graphs. Indeed, this

[3]This procedure can only be actually carried out in the case of reasonably small finite graphs, but it is easy to see that it will work 'in principal' for any set.

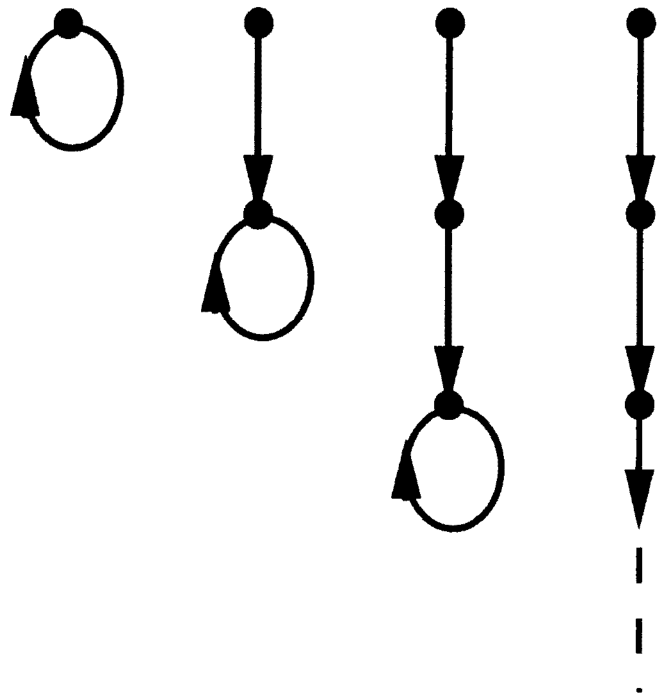


Figure 7.5: Different graphs depicting the set Ω.

is arguably the most appropriate means of depicting a circular set, since circularity is a 'looping' concept. Figure 7.5 illustrates this quite clearly, by giving a number of different graphs each of which represents the circular set

$$\Omega = \{\Omega\}.$$

Finally, consider the sets a, b, c defined as follows:

$$\begin{aligned} a &= \{b, c\}, \\ b &= \{\text{Zermelo, Fraenkel}, c\}, \\ c &= \{\text{Hilbert, Fraenkel}, b\}. \end{aligned}$$

Here we have both circularity and atoms. Figure 7.6 provides a graph depicting the set a.

Now, as things stand at the moment, all I appear to have done is exhibit a rather handy, though perhaps obvious, means of depicting sets—or rather the hereditary membership relation of sets—by means of graphs. Except, of course, that I have extended the discussion into what from the standpoint of classical (well-founded) set theory is the decidedly fanciful domain of 'sets' involving circularity. But, in fact, I have prepared the way for a significant payoff. All that needs to be done in order to collect that payoff is to recall the basic strategy of developing our theory of sets by an *analysis* of the constituency structure of the kinds of objects that arise in the real world.

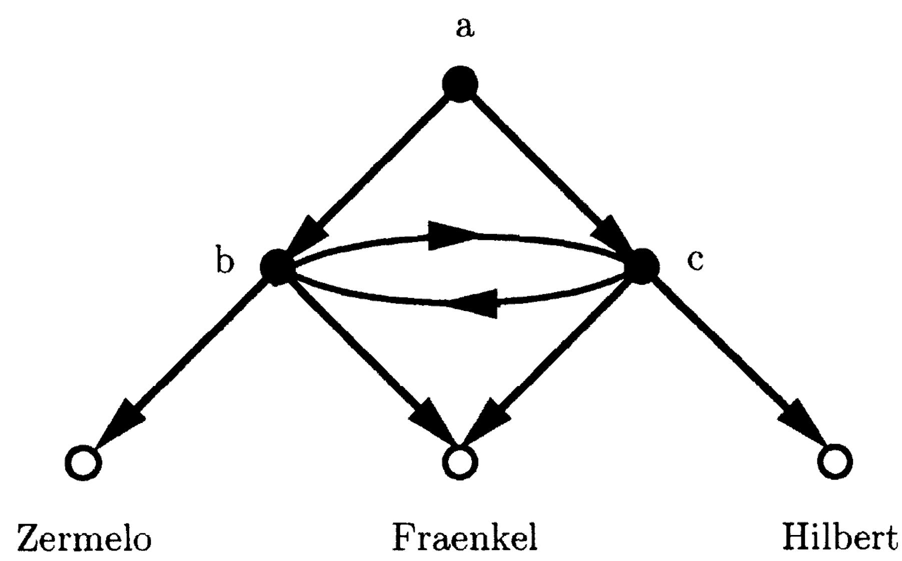


Figure 7.6: A circular set containing atoms.

According to that strategy, *graphs* (of the general forms of those discussed above) are, in a sense, *prior* to the sets they depict. Given some structured object a in the world, we may (in theory, at least) represent its hereditary constituency relation by means of a graph and thereby obtain a 'set-theoretic' model of a by moving from the graph to the set it depicts—namely, the set that corresponds to the top node of the graph.

In order for this process to work, what we need to know—and all that we need to know—is that to every graph $\mathcal{G}$ of the appropriate form (see momentarily) there is a set that $\mathcal{G}$ depicts (as its hereditary membership relation). And it is this concept of 'set from a graph' that I intend to work with.

Under this conception of 'set', all the 'usual' well-founded sets are available, since each is depicted by the graph of its hereditary membership relation, obtained as outlined above. In addition, any graph that has an infinite descending path or else contains a circuit (loop), as in Figures 7.5 and 7.6, will give rise to a non-well-founded (or circular) set. Thus non-well-founded sets arise quite naturally alongside the more familiar well-founded sets.

At this stage, I need to be precise as to just what kinds of graphs give rise to 'sets' in the above fashion.

First of all, we are restricting our attention to *directed graphs*, that is to say, graphs for which every edge has a single, designated direction. Within classical set theory, such a graph, $\mathcal{G}$, is usually defined as consisting of a nonempty set G of *nodes* (or *vertices*) and a set E of (*directed*) *edges*, where each *edge* in E is an ordered pair (x, y) of nodes. If $(x, y) \in E$, we say x and y are *joined by the edge* (x, y).

When I draw a particular graph, I represent an edge by means of an arrowed line connecting the two nodes concerned (in the appropriate direction). Thus if $(x, y) \in E$, I write $x \longrightarrow y$. In such a case, I say x is a *parent* of y or that y is a *child* of x.

It does not matter what elements of the set-theoretic universe are taken to act as the nodes of any given graph. A canonical choice—and the one I shall officially adopt—is to use the ordinal numbers for this purpose. The important issue is the graph-theoretic structure exhibited by that graph.

A *path* in a graph is a finite or infinite sequence

$$n_0 \longrightarrow n_1 \longrightarrow n_2 \longrightarrow \ldots$$

of nodes, each of which (except the first) is a child of its predecessor.

If there is a path

$$n_1 \longrightarrow n_2 \longrightarrow \ \ldots \ \longrightarrow n_k$$

from a node n_1 to a node n_k, I say that n_1 is an *ancestor* of n_k or that n_k is a *descendant* of n_1.

A graph is said to be *pointed* if there is a unique, distinguished node n_0 (called the *point* or *top node*, or sometimes the *root*, of the graph) such that all other nodes are descendants of n_0. Diagrams of pointed graphs generally show the 'top node' at the top of the picture. In this book, I shall assume all graphs are pointed. Thus, from now on, the word 'graph' should be taken to mean 'pointed, directed graph'.

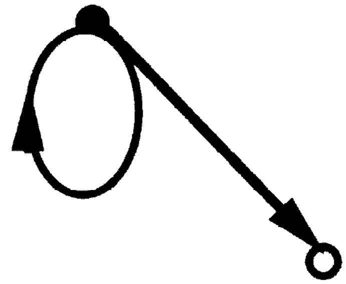
It is of course the top node of a graph that corresponds to the 'set' depicted by that graph.

7.2 The Anti-Foundation Axiom

Broadly speaking, the intuitions that lead to the axioms of Zermelo–Fraenkel set theory hold true in the present situation, except for the Axiom of Foundation. So, providing we can be assured that the resulting system is consistent (i.e. consistent relative to the Zermelo–Fraenkel system itself), it is sensible to combine our new conception of a 'set determined by an arbitrary graph' with the remaining axioms. But there is a problem. To see what it is, consider the two non-well-founded sets

$$a = \{\text{Zermelo}, a\}\ ,\ \ b = \{\text{Zermelo}, b\}.$$

Are the sets a and b equal or not? In the case of well-founded set theory, the answer to a question of this nature is readily obtained by applying the Axiom of Extensionality: two sets are equal if and only if they have the



Zermelo

Figure 7.7: Graph depicting the unique set a such that $a = \{a, \text{Zermelo}\}$.

same elements. But in the present case, this axiom simply leads to the conclusion

$$a = b \text{ if and only if } a = b.$$

So in order to resolve identity conditions where non-well-founded sets are concerned, we will have to look for some alternative principle. Given the motivation that lies behind out present theory of sets, it seems fairly clear where we should look—and indeed what the solution to our problem should be: any given graph should (presumably) depict only one set, or, to give an alternative formulation, two sets that are depicted by the same graph should be identical.

In the case of the above example, both sets give rise to the same hereditary membership graph, namely, the one shown in Figure 7.7. Consequently, these two sets are (i.e. should be) one and the same.

This consideration leads fairly rapidly to the formulation of the following additional axiom that ought to be assumed in order to obtain an intuitive and workable theory of sets that allows for the existence of circular sets.

Every graph depicts exactly one set.

Because this principle explicitly gives rise to the existence of non-well-founded sets, I shall follow Aczel[4] and refer to this principle as the *Anti-Foundation Axiom* (AFA).

Our task now is to develop our theory of sets in a rigorous manner to incorporate this extra principle.

Obviously, since our present conception of a set requires the notion of an arbitrary graph, we need to establish some form of basic set-theoretic framework before we can even *state* the axiom AFA introduced above. This means that we need to write down some initial collection of set-theoretic

[4]The present development of a non-well-founded set theory follows closely that of Peter Aczel [1].

Figure 7.8: Decorations of the graphs shown in Figure 7.4.

principles, principles that will not effect the issues addressed by AFA one way or the other.[5] Since the present aim is to remain as close to traditional set theory as possible, while remaining true to the modeling process we have in mind, I take for this initial framework the theory ZFCA (i.e. the Zermelo–Fraenkel axioms modified to allow for atoms), modified by dropping the Axiom of Foundation. I denote this theory by the acronym ZFCA$^-$. I denote the set of atoms by $\mathcal{A}$.

Let $\mathcal{G}$ be a graph with top node n_0. A *tagging* of $\mathcal{G}$ is an assignment to every childless node of $\mathcal{G}$ of either an atom (of the underlying set theory) or else the empty set, $\emptyset$. That is, a tagging is a function from the set of childless nodes of $\mathcal{G}$ into the collection $\mathcal{A} \cup \{\emptyset\}$.

Suppose now that $\mathcal{G}$ is *tagged*, that is, there is some tagging function, t, for $\mathcal{G}$. By a *decoration* of $\mathcal{G}$ (relative to t), I mean a function, d, defined on $\mathcal{G}$ such that:

(i) if n is a childless node, then $d(n) = t(n)$;

(ii) if n is not childless, then $d(n) = \{d(n') \mid n \longrightarrow n'\}$.

For example, the two graphs shown in Figure 7.4 have the decorations shown in Figure 7.8 (assuming the one childless node is tagged with the empty set in each case).[6]

A graph is said to be *well-founded* if it has no infinite path. The following fact concerning well-founded graphs is a slight reformulation of a standard result of classical set theory.

Theorem 7.2.1 [The Collapsing Lemma] Every well-founded tagged graph has a unique decoration.

[5] Recall that I took a similar course with Zermelo–Fraenkel set theory. Some initial axiomatic development of set theory is necessary in order to properly define the cumulative hierarchy that provides the underlying conception for the entire theory.

[6] A glance at this figure should indicate why I use the word 'decoration' for this concept.

Proof: A straightforward application of definition by recursion on the well-founded graph relation, giving d as the unique function satisfying the requirements (i) and (ii) above, for each node n of the graph. [Exercise: Fill in the details.] □

Given a set x, any tagged graph that has a decoration which assigns x to its top node is called a *picture* of x.

Thus, for example, Figure 7.2 gives pictures of the first four ordinal numbers, Figure 7.4 gives alternative pictures of the ordinals 2 and 3, Figure 7.5 gives a number of different pictures of the set Ω, and Figure 7.7 gives a picture of the unique set a such that

$$a = \{a, \text{Zermelo}\}.$$

[Exercise: Give two other pictures of this particular set, one a finite graph, the other infinite.]

As an immediate consequence of Theorem 7.2.1, we see that every well-founded graph is a picture of a unique set.

By simply regarding the hereditary membership relation of a given set as a graph (i.e. $n \longrightarrow n'$ if and only if $n' \in n$), we see that every set has at least one picture. In fact, we can say more. In graph-theoretic terminology, a *tree* (see Section 4.4) is a graph such that for any node n there is a *unique* path starting from the top node and terminating at n. Then we have

Lemma 7.2.2 Every set can be pictured by a tree.

Proof: Let $\mathcal{G}$ be a graph with top node n_0 that pictures the set x. Define a new graph $\mathcal{G}'$ as follows. The nodes of $\mathcal{G}'$ are the finite paths

$$n_0 \longrightarrow n_1 \longrightarrow \;\dots\; \longrightarrow n_k$$

starting from n_0, and the edges are the pairs

$$(n_0 \to \dots \to n_k \;,\; n_0 \to \dots \to n_k \to n_{k+1}).$$

It is easily seen that if d is a decoration of the graph $\mathcal{G}$, then d' is a decoration of $\mathcal{G}'$, where we define

$$d'(n_0 \to \dots \to n_k) = d(n_k).$$

(Taggings are likewise intimately related.)

Thus $\mathcal{G}'$ also pictures the set x. I refer to $\mathcal{G}'$ as the *unfolding* of $\mathcal{G}$. □

It should be noted that even when we restrict our attention to trees, pictures of sets will not be unique. For instance, the graphs shown in

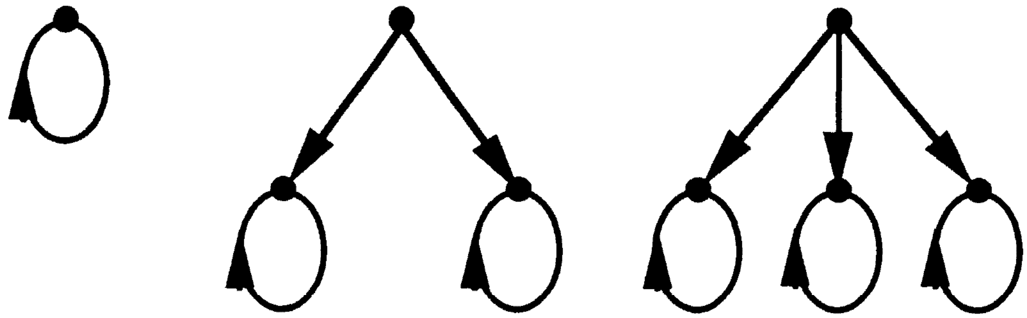


Figure 7.9: Nonisomorphic graphs for the set Ω.

Figure 7.9 all picture the set Ω, but they unfold to different (nonisomorphic) trees.

Using the newly introduced terminology, I may now state the axiom AFA:

> **The Anti-Foundation Axiom** (AFA): *Every tagged graph has a unique decoration.*

The existence part of AFA alone clearly violates the Axiom of Foundation. For instance, none of the graphs depicted in Figure 7.5 can be decorated using sets from the well-founded Zermelo–Fraenkel universe of sets.[7] On the other hand, each of these particular graphs can be decorated by assigning the non-well-founded set $\Omega = \{\Omega\}$ to each node.

By a *universe* for a theory T of sets we mean a collection V of sets that is a model of T. The following result is proved in Section 7.8.

Analogously to ZFCA, I denote by ZFC$^-$ the theory ZFC minus the Axiom of Foundation.

Theorem 7.2.3 If V is a universe for ZFC set theory (respectively, a universe for ZFCA set theory, where the atoms form a collection $\mathcal{A}$), then there is a universe V^* for ZFC$^-$ + AFA (respectively, ZFCA$^-$ + AFA with atoms from $\mathcal{A}$) such that $V \subset V^*$. □

Besides showing that the theory ZFCA$^-$ + AFA is consistent relative to ZF, the proof of this result shows how a given model of ZFC may be extended to a model of ZFC$^-$ + AFA (respectively, how a given model of ZFCA may be extended to a model of ZFCA$^-$ + AFA having the same collection of atoms).

[7] In fact the statement that no non-well-founded graph can be decorated is just a reformulation of the Axiom of Foundation.

7.3 The Solution Lemma

One of the most important consequences of AFA, as far as applications are concerned, is the way that it guarantees the existence of 'solutions' to systems of 'equations'.

The general problem is perhaps best introduced by way of a simple example.

Suppose $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ are set-indeterminates, and consider the system of equations

$$\begin{aligned}\mathbf{x} &= \{\text{Zermelo}, \mathbf{y}\}\\ \mathbf{y} &= \{\text{Fraenkel}, \mathbf{z}\}\\ \mathbf{z} &= \{3, 5\}\end{aligned}$$

(where 3 and 5 are the usual von Neumann ordinal numbers).

Then it is easy to 'solve' this system of equations for the unknowns $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$. The three sets concerned are

$$\begin{aligned}\mathbf{x} &= \{\text{Zermelo}, \{\text{Fraenkel}, \{3, 5\}\}\}\\ \mathbf{y} &= \{\text{Fraenkel}, \{3, 5\}\}\\ \mathbf{z} &= \{3, 5\}\end{aligned}$$

(where '3' and '5' here denote the corresponding von Neumann *sets*).

To obtain this solution, you simply observe that the last equation already gives a solution for $\mathbf{z}$, then substitute for $\mathbf{z}$ in the second equation to obtain the solution for $\mathbf{y}$, and finally substitute for $\mathbf{y}$ in the first equation to obtain the set corresponding to $\mathbf{x}$.

Now consider the amended system

$$\begin{aligned}\mathbf{x} &= \{\text{Zermelo}, \mathbf{y}\}\\ \mathbf{y} &= \{\text{Fraenkel}, \mathbf{z}\}\\ \mathbf{z} &= \{\mathbf{x}, \mathbf{y}\}\end{aligned}$$

where the sets 3 and 5 in the first system have been replaced by the indeterminates $\mathbf{x}$ and $\mathbf{y}$. Here the circularity in the system makes it impossible to derive a solution as for the first system. But, given the previous discussions, a natural approach is to investigate the graph that any solution would have to satisfy.

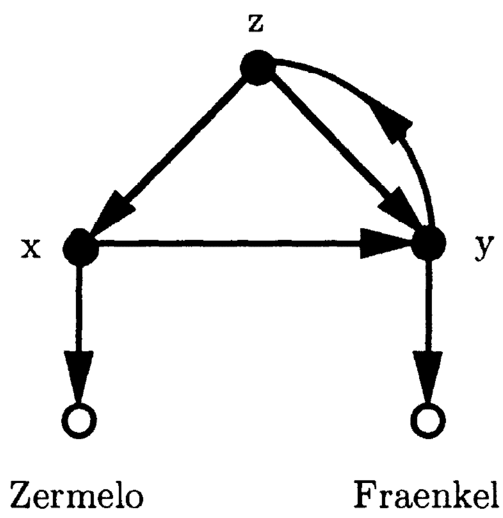


Figure 7.10: Solution of a system of equations using a graph.

A few moments analysis reveals that a graph as in Figure 7.10 provides a representation of the membership structure any solution must have. (Here I use the letters x, y, z to provide 'labels' for the nodes corresponding to the indeterminates **x**, **y**, **z**, respectively. For the sake of this informal, intuitive discussion, these labels should be regarded as nothing other than diagramatic markers that serve to distinguish the nodes until the application of AFA yields sets to which these nodes correspond.)

By AFA, the tagged graph in Figure 7.10 has a unique decoration, d. Then, if $d(x) = X, d(y) = Y, d(z) = Z$, the sets X, Y, Z clearly solve the system of equations (for **x**, **y**, **z**, respectively). That is to say, these three sets satisfy the identities

$$\begin{aligned} X &= \{\text{Zermelo}, Y\} \\ Y &= \{\text{Fraenkel}, Z\} \\ Z &= \{X, Y\}. \end{aligned}$$

Now, intuitively, it seems clear that this approach using graphs and AFA should work for any such system of equations, involving any number of unknowns, with the set-theoretic constructions on the right-hand sides of the equations being arbitrarily complex, having as many nestings of sets as required. As long as each indeterminate appears, on its own, on the left-hand side of precisely one equation in the system, it should be possible to draw a graph depicting the membership structure that any solution will have to have, and thus, by AFA, to obtain a (presumably unique) solution to the system.

The Solution Lemma, proved using AFA, says that this is indeed the case. In order to state the lemma properly, I need to first set up the appropriate machinery.

I denote by $V_{\mathcal{A}}$ the 'universe' of all sets (of the theory ZFCA$^-$ + AFA) built on the collection $\mathcal{A}$ of atoms. Let $\mathcal{X}$ be a collection of set-indeterminates. I denote by $V_{\mathcal{A}}[\mathcal{X}]$ the collection of all set terms that can be built up using elements of $V_{\mathcal{A}}$ and the indeterminates in $\mathcal{X}$. That is, $V_{\mathcal{A}}[\mathcal{X}]$ will be an extension of $V_{\mathcal{A}}$ that contains objects such as

$$\{a, b, \mathbf{x}, \{\mathbf{y}, c\}\}$$

$$\{a, \{\mathbf{x}, \{b, \{\mathbf{z}\}\}\}\}$$

$$\{1, 2, \{\Omega, \mathbf{x}\}\}$$

where $a, b, c \in V_{\mathcal{A}}$ and $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z} \in \mathcal{X}$.

Formally, I regard the indeterminates in $\mathcal{X}$ as extra atoms and take

$$V_{\mathcal{A}}[\mathcal{X}] = V_{\mathcal{A} \cup \mathcal{X}}.$$

This construction is clearly analogous to the formation of the ring $\mathcal{F}[X]$ of polynomials in indeterminates from X over a field $\mathcal{F}$. And just as the members of $\mathcal{F}[X]$ give rise to systems of polynomial equations to be solved in $\mathcal{F}$, so too the members of $V_{\mathcal{A}}[\mathcal{X}]$ provide systems of set equations to be solved in $V_{\mathcal{A}}$.

By an *equation in* $\mathcal{X}$, I mean an expression of the form

$$\mathbf{x} = t$$

where $t \in V_{\mathcal{A}}[\mathcal{X}]$.

By a *system of equations in* $\mathcal{X}$, I mean a family of equations

$$\{\mathbf{x} = t_{\mathbf{x}} \mid \mathbf{x} \in \mathcal{X}\},$$

where there is exactly one equation for each indeterminate $\mathbf{x} \in \mathcal{X}$.

By a *solution* to an equation

$$\mathbf{x} = t$$

I mean an *assignment*

$$f : \mathcal{X} \to V_{\mathcal{A}}$$

of sets or atoms to indeterminates such that the equation yields a valid set-theoretic identity when each occurrence of each indeterminate in the equation is replaced by its image under f.

Thus, to use a suggestive notation familiar from formal logic, if t is an element of $V_{\mathcal{A}}[\mathcal{X}]$ that involves the indeterminates $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, ..., and I write $t = t(\mathbf{x}, \mathbf{y}, \mathbf{z}, \ldots)$ to indicate this fact, then the assignment

$$f(\mathbf{x}) = a\ ,\ f(\mathbf{y}) = b\ ,\ f(\mathbf{z}) = c, \ldots$$

will be a solution to the above equation if and only if

$$a = t(a, b, c, \ldots).$$

More generally, I say that an assignment f of sets to the indeterminates in $\mathcal{X}$ is a *solution* to a system of equations

$$\mathbf{x} = t_{\mathbf{x}} \quad (\mathbf{x} \in \mathcal{X})$$

if and only if f is a solution for every equation in the system.

To formalize the above notions within our theory of sets, the idea is to proceed as follows. First prove that any assignment $f : \mathcal{X} \to V_{\mathcal{A}}$ extends in a natural and unique fashion to a function

$$\hat{f} : V_{\mathcal{A}}[\mathcal{X}] \to V_{\mathcal{A}}.$$

Then say that the assignment $f : \mathcal{X} \to V_{\mathcal{A}}$ is a *solution* to the equation

$$\mathbf{x} = t$$

if and only if

$$f(\mathbf{x}) = \hat{f}(t).$$

This formal development is carried out in detail in Section 7.6, where I also prove the following key result:

Theorem 7.3.1 [The Solution Lemma] Every system of equations in a collection $\mathcal{X}$ of indeterminates, over the universe $V_{\mathcal{A}}$, has a unique solution in $V_{\mathcal{A}}$. □

The general idea for the proof of this result is to develop a formal, and more general, analogue of the method used above in order to solve our sample system of three equations (where we proceeded via the graph in Figure 7.10 and then applied AFA to obtain the required sets).

It is worth remarking that the Solution Lemma is logically equivalent to AFA (over the theory ZFCA^{-}).

7.4 Inductive Definitions Under AFA

Inductive definitions pervade set theory and logic. For instance, the class of ordinals can be defined inductively as the smallest class *Ord* such that:

(i) $\emptyset \in Ord$;

(ii) if $\alpha \in Ord$, then $\alpha \cup \{\alpha\} \in Ord$;

(iii) if $x \subseteq Ord$ and x is a set, then $\bigcup x \in Ord$.

In the absence of the Axiom of Foundation, this definition serves to define the class of *well-founded* ordinals.

To see why this definition is described as *inductive*, imagine trying to construct the ordinals one by one in the following 'inductive' fashion. Start out with $0 = \emptyset$. The successor ordinal to an ordinal α is defined as the set $\alpha \cup \{\alpha\}$. In the case of limit ordinals, take unions, so that a limit ordinal α is given as

$$\alpha = \bigcup\{\beta \mid \beta < \alpha\} = \bigcup\alpha.$$

Of course, this procedure to define the ordinals cannot be carried out as described, since it assumes that the ordinals are *already available* to index the definition (i.e. to provide the domain of the sequence of ordinals being defined). But the original definition of the class *Ord* given above serves to capture the class of ordinals, by taking minimal closure under the two constructive principles (successor and union) used in this attempted iterative construction.

As a first step toward obtaining a general framework that encompasses such minimal-closure, inductive definitions, consider the function γ from sets to sets defined by

$$\gamma(x) = \{\emptyset\} \cup \{\bigcup x\} \cup \{y \cup \{y\} \mid y \in x\}.$$

For any class X now, define

$$\Gamma(X) = \bigcup\{\gamma(x) \mid x \subseteq X \wedge x \text{ is a set}\}.$$

Then clearly, Γ is an operator taking classes to classes, that is *monotone*, in the sense that

$$X \subseteq Y \quad \text{implies} \quad \Gamma(X) \subseteq \Gamma(Y).$$

Moreover, Γ is *set-based*, which means that, for any set z,

$$\text{if } z \in \Gamma(X), \text{ then } z \in \Gamma(x) \text{ for some set } x \subseteq X.$$

Clearly, a straightforward translation of our definition of the class of ordinals now is that *Ord* is the smallest class X such that $\Gamma(X) = X$. (Since $X \subseteq \Gamma(X)$ for any class X, this is equivalent to *Ord* being the smallest class X such that $\Gamma(X) \subseteq X$.)

In general now, if Γ is any class operator that is *monotone* and *set-based*, as defined above, then, as I shall prove in Section 7.7, there will be a *least fixed-point* X for Γ, that is, a smallest class X such that $\Gamma(X) = X$. I then say that the operator Γ thereby provides an *inductive definition* of the class X.

I shall also prove that every monotone, set-based operator has a *greatest fixed-point.* If Y is the greatest fixed-point of Γ, I shall say that Γ provides a *co-inductive definition* of the class Y.

In the case of the particular operator Γ defined above, the greatest fixed-point is the class, V, the entire universe of sets (this is easily seen), so the co-inductive definition gives us nothing new. But for other examples the greatest fixed-point can be both nontrivial (i.e. not just V) and distinct from the least fixed-point. And in cases where the underlying set theory is ZFCA$^-$ + AFA rather than ZFCA, it is often the greatest fixed-point that is of more use than the least fixed-point. The example below is a case in point.

Assume for simplicity that the collection $\mathcal{A}$ of atoms is finite. Consider the operator Γ that assigns to any class X the class of all finite subsets of $X \cup \mathcal{A}$. In ZFCA, this operation has a unique fixed-point, the set HF of all hereditarily finite sets. But in ZFCA$^-$ + AFA, there are many distinct fixed-points. The smallest fixed-point, HF_0, can be characterized as the smallest set satisfying the condition

$$\text{if } a \subseteq HF_0 \cup \mathcal{A} \text{ and } a \text{ is finite, then } a \in HF_0$$

(i.e. $\Gamma(HF_0) \subseteq HF_0$.)

The greatest fixed-point, HF_1, can be characterized as the largest set satisfying

$$\text{if } a \in HF_1, \text{ then } a \subseteq HF_1 \cup \mathcal{A} \text{ and } a \text{ is finite}$$

(i.e. $HF_1 \subseteq \Gamma(HF_1)$.)

It is clear that $HF_0 \subseteq HF_1$, and in ZFCA these two sets coincide. But under AFA, the inclusion is proper. In particular, it is easily demonstrated that every member of HF_0 is well-founded, but HF_1 contains non-well-founded sets. For example, Ω is a member of HF_1. Indeed, HF_1 consists of all and only those sets that can be pictured by at least one finitely branching graph. Since this latter is obviously the correct notion of hereditarily finite set under our present conception of sets as determined by graphs, in this case the co-inductive definition provides the most appropriate definition.

The above example is typical of the situation in non-well-founded set theory. A pair of inductive and co-inductive definitions that characterize the same set or class in classical set theory often yield distinct classes under

AFA. The least fixed-point, specified by the inductive definition, usually consists of the well-founded members of the largest fixed-point, given by the co-inductive definition. For reasons outlined below, it is usually the latter that is required for applications (under AFA). (Though in the case of the class *Ord* considered above, it is the inductive definition that is by far the more important of the two. But this is for the special reason that the *well-foundedness* of the ordinals that is one of their most significant properties.)

It is largely because of the way the Solution Lemma operates that, when AFA is assumed, co-inductive definitions are often more useful than inductive definitions. The situation is best explained by starting with a simple example, namely, the co-inductively defined set HF_1 of all hereditarily finite sets in the AFA universe (with a finite set of atoms).

Suppose we have some finite system of equations of the form

$$\mathbf{x} = a_{\mathbf{x}}(\mathbf{x},\mathbf{y},\ldots)$$

where each $a_{\mathbf{x}}$ is in the collection $HF_1^{\mathcal{X}}$ of all hereditarily finite sets in the expanded universe $V_{\mathcal{A}}[\mathcal{X}]$ (which, you may recall, is formally the same as $V_{\mathcal{A}\cup\mathcal{X}}$). And suppose that we apply the Solution Lemma to obtain a solution f to this system of equations. Intuitively, the set-theoretic structure of each $V_{\mathcal{A}}[\mathcal{X}]$-set $a_{\mathbf{x}}$ is that of a hereditarily finite set, and consequently one might expect that the solution sets $f(\mathbf{x})$ are also hereditarily finite, that is, in the collection HF_1 as defined in the universe $V_{\mathcal{A}}$. That this is indeed the case is a special case of what is known as the Co-Inductive Closure Theorem, proved in Section 7.7. A nonrigorous argument for the present example is given below.

Recall that in my original motivation for the Solution Lemma, I showed how, in the case of a simple example at least, a system of equations may be 'unraveled' to produce a graph that any solution will have to satisfy, whence by AFA we can conclude that there is in fact a solution. As I mentioned at the time, the proof of the Solution Lemma consists of a formal analogue of this heuristic argument. The idea behind the proof of the Co-Inductive Closure Theorem is to trace through the proof of the Solution Lemma and check that closure is indeed achieved. (This requires that the class operator Γ concerned satisfies some fairly general additional requirements that will be made precise when I give the formal proof.) In the case of the present example, the following argument gives the desired result.

First of all, by introducing more indeterminates, we may assume that each equation is of one of the following simple forms:

- $\mathbf{x} = \emptyset$;
- $\mathbf{x} = a$, for some atom $a \in \mathcal{A}$;

- $\mathbf{x} = \{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$,

where $\mathbf{y}_1, \ldots, \mathbf{y}_n$ are other indeterminates with their own equations in the system.

Let f be the solution to this modified system. It is clear that the collection $HF_1 \cup \mathrm{ran}(f)$ satisfies the defining condition for HF_1. So, by the maximality of HF_1, $\mathrm{ran}(f) \subseteq HF_1$, as required.

The general statement of the Co-Inductive Closure Theorem runs roughly like this. Suppose Γ is some monotone, set-based class operator. Using Γ, we can co-inductively define a collection of objects from the universe $V_{\mathcal{A}}$ as the largest fixed point of Γ in $V_{\mathcal{A}}$. Call the objects in this collection *Γ-objects.* Likewise, we may use the same operator Γ in order to define an analogous collection in the universe $V_{\mathcal{A}}[\mathcal{X}]$. Call the objects in this collection *parametric Γ-objects.* What the Closure Theorem says is that, providing Γ satisfies some fairly general requirements, any system of equations involving only parametric Γ-objects will have only Γ-objects as solutions.

The combination of the Solution Lemma and the Co-Inductive Closure Theorem provides a powerful tool for handling non-well-founded sets under AFA and, in this respect, takes on the role played by the recursion principle in Zermelo–Fraenkel set theory.

7.5 Graphs and Systems

The notion of a *graph* has been precisely defined already. In order to obtain, in particular, a proof of the consistency of AFA, I require the following generalization to allow for a proper class of nodes.

By a *system* I mean a class M of *nodes* together with a class of (*directed*) *edges*, each edge being an ordered pair (n, n') of nodes. I write $n \longrightarrow n'$ if (n, n') is an edge of M. Any system is required to satisfy the requirement that, for each node n, the collection

$$\mathrm{ch}_M(n) = \{n' \in M \mid n \longrightarrow n'\}$$

of all *children* of n is a *set.*

Clearly, any graph is a system. For an example of a system that is not a graph (because the collection of nodes forms a proper class), take the collection of nodes to be the universe V of all pure (i.e. atomless) sets, with the edges given by $x \longrightarrow y$ if and only if $y \in x$.

Note that whereas graphs are assumed to have a unique *top node*, no such requirement is placed on systems.

Because of the different roles played by the two collections of atoms in our theory, *taggings* are defined as *partial* functions. Thus, a tagging of the system M is an assignment, t, to some or all of the childless nodes, a, of M, of an atom, $t(a)$ (i.e. a member of $\mathcal{A} \cup \mathcal{X}$). I denote such a tagged system by (M, t). (Note that t may be a 'function' only in the proper class sense.)

Notice that if t is the nowhere-defined tagging on M, then the tagged system (M, t) is essentially the same as the untagged system M. Accordingly, I shall henceforth use the terms 'system' and 'graph' to mean 'tagged system' and 'tagged graph', respectively.

In order to establish the Solution Lemma, I shall need to associate atoms ('indeterminates') with nodes, as well as be able to handle the assignment to each indeterminate of a set in $V_{\mathcal{A}}$ when the equational system is solved. The following definition supplies the appropriate machinery. Since it may be necessary to associate more than one indeterminate to a given node, the 'labeling' function defined below assigns not a single set/atom but a set of sets/atoms, to each node.

A *labeling* of a (tagged) system (M, t) is a function l (possibly a 'function' in the proper class sense) defined on $M - \mathrm{dom}\,(t)$ that assigns to each node n not in $\mathrm{dom}\,(t)$, a (possibly empty) set $l(n)$ of sets/atoms.

The elements of the set $l(n)$, for any node n, are the *labels* assigned to the node n by the labeling function.

A *labeled system* then is just a system, (M, t), together with a labeling function, l. I denote such a system by (M, t, l).

A *decoration* of a labeled system (M, t, l) is an assignment d of a set $d(n)$ to each node n such that:

(i) if $n \in \mathrm{dom}(t)$, then $d(n) = t(n)$;

(ii) if $n \notin \mathrm{dom}(t)$, then

$$d(n) = \{d(n') \mid n \longrightarrow n'\} \cup l(n).$$

By virtue of the above remark, this definition includes the special case of a decoration of an unlabeled system (M, t): if $l(n) = \emptyset$ for each parent node n of M, then $d(n) = t(n)$ for all tagged nodes and $d(n) = \{d(n') \mid n \longrightarrow n'\}$ for all untagged nodes. This simply extends to (tagged) *systems*, the definition of a decoration of a (tagged) *graph* given in Section 7.2.

Our starting point is the axiom AFA:

> **The Anti-Foundation Axiom (AFA)**: *Every (tagged) graph has a unique decoration.*

I shall prove that this formulation is already enough to prove the apparently stronger result that every labeled system has a unique decoration. The following theorem provides the first of two steps toward this goal, by showing that it is possible to go from decorations of unlabeled *graphs* to decorations of unlabeled *systems.*

Theorem 7.5.1 (Assuming AFA.) Every (tagged) system has a unique decoration.

Proof: Let (M, t) be a system. For each $n \in M$, we may define a graph M_n by taking the nodes of M_n to be all nodes of M that lie on some path of M starting from node n, and taking as edges all edges of M that connect two members of M_n. Since the collection of all children of any given node in M forms a set, it is easily seen that M_n is itself a set. Indeed, if we take $X_0 = \{n\}$ and, for each natural number i, define

$$X_{i+1} = \bigcup \{\mathrm{ch}_M(m) \mid m \in X_i\},$$

then each X_i is a set, and we have $M_n = \bigcup_{i=0}^{\infty} X_i$.

The restriction t_n of the tagging function t to M_n is obviously a tagging of the graph M_n for each n. By AFA, each (M_n, t_n) has a unique decoration d_n. Define d on M by

$$d(n) = d_n(n) \quad (\forall n \in M).$$

I show that d is the unique decoration of (M, t).

First note that if $n \in \mathrm{dom}(t)$, then n is the only node of M_n and

$$d(n) = d_n(n) = t_n(n) = t(n).$$

To handle the remaining nodes of M, we observe that if $n \longrightarrow m$ in M, then every node of M_m will be a node of M_n and the restriction of d_n to M_m will be a decoration of M_m and, hence, equal to d_m, the unique decoration of (M_m, t_m). Thus whenever $n \longrightarrow m$ in M, we have $d_n(m) = d_m(m) = d(m)$. Consequently, for each untagged node $n \in M$, we have

$$d(n) = d_n(n) = \{d_n(m) \mid n \longrightarrow m \text{ in } M_n\} = \{d(m) \mid n \longrightarrow m \text{ in } M\}.$$

Thus d is a decoration of (M, t).

To see that d is unique, simply notice that any decoration of (M, t) will restrict to a decoration of (M_n, t_n) for any node n, hence, must extend d_n, and, therefore, has to be equal to d. □

The following theorem completes our extension of AFA to cover *labeled* systems.

Theorem 7.5.2 (Assuming AFA.) Every labeled (tagged) system has a unique decoration.

Proof: Let (M, t, l) be a labeled system. Define a new, unlabeled, system (M', t') as follows. Let the nodes of M' be the members of the set

$$\{(1, n) \mid n \in M\} \cup \{(2, a) \mid a \in V_{\mathcal{A}}[\mathcal{X}]\}.$$

The edges of M' are:

- $(1, n) \longrightarrow (1, n')$, whenever $n \longrightarrow n'$ in M;
- $(1, n) \longrightarrow (2, a)$, whenever $n \in M$, $n \notin \operatorname{dom}(t)$, and $a \in l(n)$;
- $(2, a) \longrightarrow (2, b)$, whenever $b \in a$.

Define the tagging t' on M' by:

- $t'(1, n) = t(n)$, if $n \in \operatorname{dom}(t)$;
- $t'(2, a) = a$, if $a \in \mathcal{A} \cup \mathcal{X}$.

By Theorem 7.5.1, (M', t') has a unique decoration, d. Thus, for each node $n \in \operatorname{dom}(t)$,

$$d(1, n) = t'(1, n) = t(n),$$

and, for each $a \in \mathcal{A} \cup \mathcal{X}$,

$$d(2, a) = t'(2, a) = a.$$

Moreover, for each untagged (by t) node $n \in M$,

$$d(1, n) = \{d(1, n') \mid n \longrightarrow n' \text{ in } M\} \cup \{d(2, a) \mid a \in l(n)\},$$

and, for each nonatomic $a \in V_{\mathcal{A}}[\mathcal{X}]$,

$$d(2, a) = \{d(2, b) \mid b \in a\}.$$

Now, the assignment of the set $d(2, a)$ to each $a \in V_{\mathcal{A}}[\mathcal{X}]$ is a decoration of the system $V_{\mathcal{A}}[\mathcal{X}]$, tagged with the identity function on $\mathcal{A} \cup \mathcal{X}$. But the identity function on $V_{\mathcal{A}}[\mathcal{X}]$ is also a decoration of the same tagged system. So by Theorem 7.5.1, we must have $d(2, a) = a$ for all $a \in V_{\mathcal{A}}[\mathcal{X}]$.

Define e on M now by

$$e(n) = d(1, n).$$

Then if n is a tagged node of M,

$$e(n) = t(n),$$

and if n is an untagged node of M, then

$$\begin{aligned} e(n) &= \{e(n') \mid n \longrightarrow n' \text{ in } M\} \cup \{a \mid a \in l(n)\} \\ &= \{e(n') \mid n \longrightarrow n' \text{ in } M\} \cup l(n). \end{aligned}$$

So e is a decoration of (M, t, l).

To check uniqueness, suppose e' is also a decoration of (M, t, l). Then d' is a decoration of (M', t'), where we define

- $d'(1, n) = e'(n)$, for $n \in M$;
- $d'(2, a) = a$, for $a \in V_{\mathcal{A}}[\mathcal{X}]$.

By Theorem 7.5.1, we have $d' = d$. Hence for all $n \in M$, we have

$$e'(n) = d'(1, n) = d(1, n) = e(n),$$

so $e' = e$. □

In the future, I shall often simply refer to Theorem 7.5.2 above as AFA.

The following general result establishes the key facts I shall use in the proof of the Solution Lemma.

Theorem 7.5.3 (Assuming AFA.) Let (M, t, l) be a labeled system (in $V_{\mathcal{A}}[\mathcal{X}]$) such that $t(n) \in \mathcal{A}$ for all tagged nodes $n \in M$, and $l(n) \subseteq \mathcal{X}$ for all untagged nodes $n \in M$.

(i) Let $\pi : \mathcal{X} \to V_{\mathcal{A}}$. Then there is a unique map $\widehat{\pi} : M \to V_{\mathcal{A}}$ such that for each $n \in M$:

- if n is a tagged node of M, then $\widehat{\pi}(n) = t(n)$;
- if n is an untagged node of M, then

$$\widehat{\pi}(n) = \{\widehat{\pi}(n') \mid n \longrightarrow n' \text{ in } M\} \cup \{\pi(x) \mid x \in l(n)\}.$$

(ii) Suppose that to each $x \in \mathcal{X}$ there is assigned a node a_x of M. Then there is a unique map $\pi : \mathcal{X} \to V_{\mathcal{A}}$, such that for all $x \in \mathcal{X}$,

$$\pi(x) = \widehat{\pi}(a_x).$$

Proof: (i) Let $\pi : \mathcal{X} \to V_{\mathcal{A}}$ be given. Let l_π be a new labeling of (M, t), defined by setting

$$l_\pi(n) = \{\pi(x) \mid x \in l(n)\}$$

for all untagged nodes n of M.

Clearly, the unique decoration of the labeled system (M, t, l_π) is the desired map $\widehat{\pi}$.

(ii) Let M' be the system having the same nodes as M, and all the edges of M, together with the edges $n \longrightarrow a_x$ whenever $n \in M$ and $x \in l(n)$. By Theorem 7.5.1, the unlabeled system (M', t) has a unique decoration, d. Thus, for each tagged node $n \in M'$,

$$d(n) = t(n),$$

and, for each untagged node $n \in M'$,

$$d(n) = \{d(n') \mid n \longrightarrow n' \text{ in } M\} \cup \{d(a_x) \mid x \in l(n)\}.$$

Let $\pi(x) = d(a_x)$ for each $x \in \mathcal{X}$. Thus $\pi : \mathcal{X} \to V_{\mathcal{A}}$. Moreover, for each untagged node $n \in M$,

$$d(n) = \{d(n') \mid n \longrightarrow n' \text{ in } M\} \cup \{\pi(x) \mid x \in l(n)\}.$$

So by part (i) of the theorem, $d = \widehat{\pi}$. So, in particular, for all $x \in \mathcal{X}$, we have

$$\pi(x) = \widehat{\pi}(a_x).$$

To show that π is unique with this property, suppose that $\pi' : \mathcal{X} \to V_{\mathcal{A}}$ is such that $\pi'(x) = \widehat{\pi}'(a_x)$ for all $x \in \mathcal{X}$. Then clearly, $\widehat{\pi}'$ will be a decoration of (M', t). Thus by Theorem 7.5.1, $\widehat{\pi}' = d$. Hence for any $x \in \mathcal{X}$,

$$\pi'(x) = \widehat{\pi}'(a_x) = d(a_x) = \pi(x).$$

Thus $\pi' = \pi$. □

7.6 Proof of the Solution Lemma

I shall present the proof of the Solution Lemma in two parts. The first, which I shall call the Substitution Lemma, says that if you start with a collection, $\mathcal{C}$, of members of $V_{\mathcal{A}}[\mathcal{X}]$, and if you replace each indeterminate x that occurs (in the transitive closure of) some member of $\mathcal{C}$ by some member b_x of $V_{\mathcal{A}}$, then the result will be a family $\mathcal{C}'$ of well-defined members of $V_{\mathcal{A}}$.

Theorem 7.6.1 [Substitution Lemma] (Assuming AFA.) Let $\pi : \mathcal{X} \to V_{\mathcal{A}}$. Then there is a unique map $\widehat{\pi} : V_{\mathcal{A}}[\mathcal{X}] \to V_{\mathcal{A}}$ such that:

(i) $\widehat{\pi}(a) = a$, for all $a \in \mathcal{A}$;

(ii) $\widehat{\pi}(a) = \{\widehat{\pi}(b) \mid b \in V_{\mathcal{A}}[\mathcal{X}] \ \& \ b \in a\} \cup \{\pi(x) \mid x \in \mathcal{X} \ \& \ x \in a\}$, for all other a.

Proof: Let M be the system whose nodes are the members of $V_{\mathcal{A}}[\mathcal{X}]$ and whose edges are given by

$$a \longrightarrow b \quad \text{if and only if} \quad b \in a.$$

Let t be the identity function on $\mathcal{A}$. (So t is a tagging for M.) Define a labeling l of (M, t) by setting

$$l(a) = a \cap \mathcal{X}$$

for all $a \in V_{\mathcal{A}}[\mathcal{X}] - \mathcal{A}$. (Thus $l(a) \subseteq \mathcal{X}$ for all $a \in \text{dom}(l)$.)

Let $\widehat{\pi}$ be related to (M, t, l) and π as in Theorem 7.5.3(i). Clearly, $\widehat{\pi}$ is as required. □

Theorem 7.6.2 [Solution Lemma] (Assuming AFA.) Let a_x be a member of $V_{\mathcal{A}}[\mathcal{X}]$ for each indeterminate x. Then the system of equations

$$x = a_x \quad (x \in \mathcal{X})$$

has a unique solution. That is, there is an assignment $\pi : \mathcal{X} \to V_{\mathcal{A}}$ such that

$$\pi(x) = \widehat{\pi}(a_x)$$

for all $x \in \mathcal{X}$.

Proof: Let (M, t, l) be as in the proof of Theorem 7.5.3 and apply Theorem 7.6.1(ii). □

7.7 Co-Inductive Definitions

I indicated earlier that the Solution Lemma can often be combined with co-inductive definitions in order to obtain solution sets with particular properties. In this section I develop this idea formally.

I start off by recalling that a class operator Γ is said to be *monotone* if

$$X \subseteq Y \Rightarrow \Gamma(X) \subseteq \Gamma(Y),$$

and is *set-based* if

$$a \in \Gamma(X) \Rightarrow a \in \Gamma(x), \text{ for some set } x \subseteq X.$$

Taken together, these two conditions are equivalent to the following: for any class X,

$$\Gamma(X) = \bigcup\{\Gamma(x) \mid x \subseteq X \ \wedge x \text{ is a set}\}.$$

Operators that satisfy this requirement are usually said to be *set-continuous* (or, simply, *continuous*).

It is a standard fact of ZFC$^-$ set theory that every continuous operator, Γ, has both a least fixed-point and a greatest fixed-point. The least fixed-point of Γ is the unique smallest class I such that $\Gamma(I) \subseteq I$. The largest fixed-point is the unique largest class J such that $J \subseteq \Gamma(J)$. Our present interest is in the largest fixed-point, and accordingly I commence with a proof that such a largest class J exists.

Note that as an operator on *classes*, a class operator Γ should be thought of in terms of some defining formula, not as some form of extensional object. (The use of the word 'operator', as opposed to 'function', is intended to emphasize this point.)

Given Γ, define J by

$$J = \bigcup\{x \mid x \text{ is a set } \ \wedge x \subseteq \Gamma(x)\}.$$

Lemma 7.7.1 $J \subseteq \Gamma(J)$.

Proof: Let $a \in J$. Then by definition, $a \in x$ for some set x such that $x \subseteq \Gamma(x)$. Since $x \subseteq J$ and Γ is monotone, $\Gamma(x) \subseteq \Gamma(J)$. Thus $x \subseteq \Gamma(J)$. Hence $a \in \Gamma(J)$. □

Lemma 7.7.2 If $X \subseteq \Gamma(X)$, then $X \subseteq J$.

Proof: Assume $X \subseteq \Gamma(X)$, and let $a \in X$. I prove that $a \in J$.

I first show that for each set $x \subseteq X$, there is a set $x' \subseteq X$ such that $x \subseteq \Gamma(x')$. Let $x \subseteq X$. Then, by the assumption on X, $x \subseteq \Gamma(X)$. Hence as Γ is set-based,

$$(\forall y \in x)(\exists u)(y \in \Gamma(u) \wedge u \subseteq X).$$

By the Axiom of Replacement, there is a set A such that

$$(\forall y \in x)(\exists u \in A)(y \in \Gamma(u) \wedge u \subseteq X).$$

Set

$$x' = \bigcup\{u \in A \mid u \subseteq X\}.$$

Then x' is a subset of X. Moreover, as Γ is monotone, $\Gamma(u) \subseteq \Gamma(x')$ for all $u \in A$, so $x \subseteq \Gamma(x')$.

Using the above result, we can choose (using the Axiom of Choice) an infinite sequence $x_0, x_1, \ldots$ of subsets of X such that $x_0 = \{a\}$ and $x_n \subseteq \Gamma(x_{n+1})$ for all n. Set

$$x = \bigcup_{n=0}^{\infty} x_n.$$

Then x is a set. Moreover, if $y \in x$, then $y \in x_n$ for some n, so $y \in \Gamma(x_{n+1}) \subseteq \Gamma(x)$. Thus $x \subseteq \Gamma(x)$. Hence $x \subseteq J$. Since $a \in x_0 \subseteq x$, it follows that $a \in J$. □

Lemma 7.7.3 *J is the unique largest fixed-point of Γ.*

Proof: By Lemma 7.7.1 and the monotonicity of Γ,

$$\Gamma(J) \subseteq \Gamma(\Gamma(J)).$$

So by Lemma 7.7.2, $\Gamma(J) \subseteq J$. Thus by Lemma 7.7.1 again, $\Gamma(J) = J$, and so J is a fixed-point of Γ. By Lemma 7.7.2 again, J is the largest fixed-point of Γ. □

The task now is to establish a general result that will enable us to show that under certain conditions, the solution sets to a system of equations all satisfy a given co-inductive definition (where, you may recall, a co-inductive definition of a class is one that determines the class as the largest fixed-point of some continuous operator). The development should (continue to) be thought of as taking place in the set-theoretic universe $V_{\mathcal{A}}[\mathcal{X}]$.

Let Γ be a continuous operator. Assume Γ has the following 'absoluteness' property: for any set x, $\Gamma(x \cap V_{\mathcal{A}}) = \Gamma(x) \cap V_{\mathcal{A}}$. Let $J^{\mathcal{X}}$ be the largest fixed-point of Γ as defined in $V_{\mathcal{A}}[\mathcal{X}]$, and let J be the largest fixed-point as defined in $V_{\mathcal{A}}$. Notice that by virtue of the above absoluteness assumption on Γ, $J = J^{\mathcal{X}} \cap V_{\mathcal{A}}$. (This is easily proved.)

Let

$$x = a_x \quad (x \in \mathcal{X})$$

be a system of equations such that $a_x \in J^{\mathcal{X}}$, for all $x \in \mathcal{X}$.

The basic question to ask now is this. Given a solution

$$\pi(x) = b_x \quad (x \in \mathcal{X})$$

to this system, by sets b_x in $V_{\mathcal{A}}$, under what conditions may we conclude that each set b_x is in fact a member of J, the largest fixed-point of Γ as defined in $V_{\mathcal{A}}$? The answer, though not particularly pretty, is generally quite easy to apply in specific cases. It depends on the following definition.

Call a map $\tau : V_{\mathcal{A}}[\mathcal{X}] \to V_{\mathcal{A}}$ *faithful* (for the given system of equations) if $\tau(a) = a$ for all $a \in \mathcal{A}$, and for all other $a \in V_{\mathcal{A}}[\mathcal{X}]$,

$$\tau(a) = \{\tau(b) \mid b \in a\} \cup \{\tau(a_x) \mid x \in a \cap \mathcal{X}\}.$$

Theorem 7.7.4 [Co-Inductive Closure Theorem] (Assuming AFA.) Let Γ, $J^{\mathcal{X}}$, J, a_x ($x \in \mathcal{X}$) be as above. Suppose that for any faithful map $\tau : V_{\mathcal{A}}[\mathcal{X}] \to V_{\mathcal{A}}$, it is the case that

$$(*) \qquad a \in J^{\mathcal{X}} \Rightarrow \tau(a) \in \Gamma(K),$$

where K is the range of τ on $J^{\mathcal{X}}$.

Then the unique solution to the system of equations consists entirely of sets in J.

Proof: The Solution Lemma (Theorem 7.6.2) tells us that there is a unique map $\pi : \mathcal{X} \to V_{\mathcal{A}}$ such that

$$\pi(x) = \widehat{\pi}(a_x)$$

for all $x \in \mathcal{X}$, where $\widehat{\pi} : V_{\mathcal{A}}[\mathcal{X}] \to V_{\mathcal{A}}$ is such that $\widehat{\pi}(a) = a$ if $a \in \mathcal{A}$, and

$$\widehat{\pi}(a) = \{\widehat{\pi}(b) \mid b \in a\} \cup \{\pi(x) \mid x \in a \cap \mathcal{X}\}$$

if $a \notin \mathcal{A}$.

Since $\pi(x) = \widehat{\pi}(a_x)$ for all x, $\widehat{\pi}$ is faithful. Thus, by assumption, $\widehat{\pi}$ must satisfy condition $(*)$. So, if K is the range of $\widehat{\pi}$ on $J^{\mathcal{X}}$, we have

$$(**) \qquad a \in J^{\mathcal{X}} \Rightarrow \widehat{\pi}(a) \in \Gamma(K).$$

Now, if $b \in K$, then $b = \widehat{\pi}(a)$ for some $a \in J^{\mathcal{X}}$, so by $(**)$, $b \in \Gamma(K)$. Hence $K \subseteq \Gamma(K)$. So by the maximality of $J^{\mathcal{X}}$, $K \subseteq J^{\mathcal{X}}$. But $K \subseteq V_{\mathcal{A}}$. Hence, as $J = J^{\mathcal{X}} \cap V_{\mathcal{A}}$, $K \subseteq J$, and it follows that $\widehat{\pi}(a) \in J$. In particular, $\pi(x) = \widehat{\pi}(a_x) \in J$ for all $x \in \mathcal{X}$, as required. □

As an illustration of the use of the above result, take the example of the hereditarily finite sets discussed informally at the end of the previous chapter. The co-inductively defined collection *HF* of all hereditarily finite sets is the largest fixed point of the continuous operator

$$\Gamma(X) = \{a \mid a \subseteq X \cup \mathcal{A} \;\&\; a \text{ is finite}\}.$$

(As before, I assume that the collection $\mathcal{A}$ of atoms of $V_{\mathcal{A}}$ is finite here.) Notice that Γ satisfies the absoluteness requirement stipulated above for operators to which the Co-Inductive Closure Theorem may be applied. Suppose

$$x = a_x \quad (x \in \mathcal{X})$$

is a system of equations such that $a_x \in HF^{\mathcal{X}}$ for all $x \in \mathcal{X}$. Let $\tau : V_{\mathcal{A}}[\mathcal{X}] \to V_{\mathcal{A}}$ be a faithful map. I show that $(*)$ is satisfied.

Let $a \in HF^{\mathcal{X}}$. We must prove that $\tau(a) \in \Gamma(K)$, where K is the range of τ on $HF^{\mathcal{X}}$. If $a \in \mathcal{A}$ this is trivial. For the remaining cases,

$$\tau(a) = \{\tau(b) \mid b \in a\} \cup \{\tau(a_x) \mid x \in a \cap \mathcal{X}\}.$$

So, $\tau(a) \subseteq K$, and since a is finite, so too is $\tau(a)$. Thus $\tau(a) \in \Gamma(K)$, as required.

Hence, by the Co-Inductive Closure Theorem, the unique solution to the system consists of hereditarily finite sets in the sense of $V_{\mathcal{A}}$.

7.8 A Model of ZF$^-$ +AFA

This final section is fairly technical and assumes a sound knowledge of basic model theory. It is included for completeness only, since the material presented is not, at the present time, widely available.

The relative consistency result for AFA, Theorem 7.2.3, depends on an investigation of the dual questions:

- When are two sets pictured by the same graph?
- When do two graphs picture the same set?

This is the task I turn to in this section. Unless otherwise indicated, the assumed underlying set theory is ZFC$^-$; that is, Zermelo–Fraenkel set theory without the Axiom of Foundation. (I shall therefore ignore the possibility of atoms from now on. They would play no role in our development and would only be an unnecessary encumberance.)

The fundamental graph-theoretic notion that underlies our answer to the first of the above two questions is that of a *bisimulation*.[8]

Let M be a system. A binary relation R on M is called a *bisimulation* on M if, whenever aRb, then

$$(\forall x \in \mathrm{ch}_M(a))(\exists y \in \mathrm{ch}_M(b))(xRy) \wedge (\forall y \in \mathrm{ch}_M(b))(\exists x \in \mathrm{ch}_M(a))(xRy).$$

In words, if a and b are related via R, then for every child, x, of a there is a child, y, of b that is related to x, and vice versa.

The following example of this notion is basic. For two sets a, b, write $a \equiv b$ if and only if there is a graph M that is a picture of both a and b. Then $\equiv$ is a binary relation on the system V (i.e. the class of all sets, with the edge relation $x \longrightarrow y$ if and only if $y \in x$).

[8]The name comes from earlier uses of this notion in Computer Science, where it is related to a pair of processes each of which could 'simulate' the behavior of the other.

Lemma 7.8.1 The relation $\equiv$ is a bisimulation on V.

Proof: Suppose $a \equiv b$. Then there is a graph M, with top node m, and decorations d_1, d_2 of M, such that $d_1(m) = a$ and $d_2(m) = b$. Let $x \in a$. Then, as d_1 is a decoration

$$x \in \{d_1(n) \mid m \longrightarrow n\},$$

so $x = d_1(n)$ for some $n \in \mathrm{ch}_M(m)$. Let $y = d_2(n)$. Thus $y \in b$. I claim that $x \equiv y$. (By symmetry, this will be enough to establish the lemma.) In fact, the graph that pictures both x and y is just M_n, the restriction of M to all nodes that lie on some path starting from n. (The decorations that produce both x and y from this graph are simply the restrictions of d_1 and d_2 to M_n, respectively.) □

In general, a system will have many bisimulations. But, as I show below, there is always a unique maximal bisimulation. (The relation $\equiv$ of the above lemma is the maximal bisimulation on the system V.) The definition of the maximal bisimulation on a given system is straightforward.

Call a relation R on a system M *small* if it is a set. Then define a relation $\equiv_M$ on M by

$$a \equiv_M b \quad \text{if and only if} \quad aRb \text{ for some small bisimulation } R \text{ on } M.$$

As I show below, the relation $\equiv_M$ is the maximal bisimulation on M.

The following auxiliary notion will be helpful in our proof. If R is a binary relation on a system M, define the binary relation R^+ on M by aR^+b if and only if

$$(\forall x \in \mathrm{ch}_M(a))(\exists y \in \mathrm{ch}_M(b))(xRy) \wedge (\forall y \in \mathrm{ch}_M(b))(\exists x \in \mathrm{ch}_M(a))(xRy).$$

Then a relation R will be a bisimulation on M if and only if $R \subseteq R^+$, i.e. if and only if

$$aRb \;\Rightarrow\; aR^+b.$$

Note that the operator $(\;)^+$ is monotone; that is, if $R_1 \subseteq R_2$, then $R_1^+ \subseteq R_2^+$.

Lemma 7.8.2 Let M be any system. Then the relation $\equiv_M$ is the unique maximal bisimulation on M. That is:

(i) $\equiv_M$ is a bisimulation on M; and

(ii) if R is any bisimulation on M, then for any $a, b \in M$,

$$aRb \;\Rightarrow\; a \equiv_M b.$$

Proof: (i) Let $a \equiv_M b$. Thus aRb for some small bisimulation R on M. By definition of $\equiv_M$,

$$xRy \Rightarrow x \equiv_M y \quad (\forall x, y \in M).$$

So as $(\ )^+$ is monotone

$$xR^+y \Rightarrow x \equiv_M^+ y \quad (\forall x, y \in M).$$

But R is a bisimulation, so $R \subseteq R^+$. So, in particular, aR^+b, and hence $a \equiv_M^+ b$. This shows that $\equiv_M \subseteq \equiv_M^+$, which proves (i).

(ii) Let R be a given bisimulation on M, and let aRb. I show that $a \equiv_M b$. Let

$$R_0 = R \cap (M_a \times M_b).$$

It is routine to check that R_0 is a bisimulation on M such that aR_0b. But R_0 is small. Hence by definition of $\equiv_M$, $a \equiv_M b$. □

I am now in a position to show that the relation $\equiv$ on V is the maximal bisimulation on V.

Theorem 7.8.3 For all sets a, b

$$a \equiv b \Leftrightarrow a \equiv_V b.$$

Proof: By the maximality of $\equiv_V$, we know that

$$a \equiv b \Rightarrow a \equiv_V b.$$

Conversely, assume $a \equiv_V b$. Thus for some small bisimulation R on V, aRb. Define a new system M as follows. The nodes of M are the elements of R, that is, the ordered pairs (x, y) such that xRy. The edges of M are

$$(x, y) \longrightarrow (u, v) \quad \text{if and only if} \quad u \in x \;\&\; v \in y.$$

Now, if we define d_1 and d_2 on M by

$$d_1(x, y) = x, \ d_2(x, y) = y,$$

then it is easily seen that d_1 and d_2 are both decorations of M. But $(a, b) \in M$, so $M_{(a,b)}$ is a picture of both a and b. Thus by definition, $a \equiv b$. □

In general, bisimulation relations are not equivalence relations. But as the notation suggests, maximal bisimulations *are* equivalence relations.

Lemma 7.8.4 For any system M, the relation $\equiv_M$ is an equivalence relation on M.

Proof: *Reflexivity.* Since the identity relation on M is clearly a bisimulation relation, $\equiv_M$ is reflexive.

Symmetry. Suppose $a \equiv_M b$. Thus for some small bisimulation R, aRb. Let S be the reversal of R, i.e.

$$ySx \Leftrightarrow xRy.$$

It is easily seen that S is a bisimulation. Since bSa, it follows that $b \equiv_M a$.

Transitivity. Suppose $a \equiv_M b$ and $b \equiv_M c$. Let R, S be small bisimulations such that aRb and bRc. Define a relation T on M by

$$xTz \Leftrightarrow \exists y(xRy \wedge ySz).$$

It is routine to verify that T is a bisimulation on M. Since aTc, it follows that $a \equiv_M c$. □

The following simple lemma provides two conditions that imply $a \equiv_M b$.

Lemma 7.8.5 Let M be any system. Then for all $a, b \in M$:

(i) $\mathrm{ch}_M(a) = \mathrm{ch}_M(b) \;\Rightarrow\; a \equiv_M b$;

(ii) $M_a \cong M_b \;\Rightarrow\; a \equiv_M b$.

Proof: (i) Define R on M by

$$R = \{(a, b)\} \cup \{(x, x) \mid x \in M_a\}.$$

It is easily seen that R is a bisimulation on M such that aRb. Hence $a \equiv_M b$.

(ii) Let $\theta : M_a \cong M_b$, and define R on M by

$$xRy \;\Leftrightarrow\; x \in M_a \wedge y \in M_b \wedge \theta(x) = y.$$

Again it is routine to check that R is a bisimulation on M, so as aRb we again conclude that $a \equiv_M b$. □

A system M is said to be *extensional*[9] if, for all $a, b \in M$,

$$a \equiv_M b \;\Rightarrow\; a = b.$$

Theorem 7.8.6 The following are equivalent.

[9] In [1], Aczel uses the phrase 'strongly extensional' for this notion. In my development, I have no need for the weaker notion that Aczel refers to as 'extensional'.

(i) Every graph has at most one decoration.

(ii) V is extensional.

Proof: Assume (i). Let $a \equiv_V b$. Then by Theorem 7.8.3, $a \equiv b$, so there is a graph G with top node n, and decorations d_1 and d_2 of G, such that $d_1(n) = a$ and $d_2(n) = b$. By (i), $d_1 = d_2$. Hence $a = b$. This proves (ii).

Assume (ii). Let d_1 and d_2 be decorations of a graph G. If $x \in G$, then G_x is a picture of both $d_1(x)$ and $d_2(x)$, so $d_1(x) \equiv d_2(x)$. Hence by Theorem 7.8.3, $d_1(x) \equiv_V d_2(x)$. So by (ii), $d_1(x) = d_2(x)$. Hence $d_1 = d_2$. This proves (i). □

A *system map* from a system M to a system M' is a map $\pi : M \to M'$ such that for all $a \in M$, π maps the children of a in M onto the children of $\pi(a)$ in M'; i.e. for all $a \in M$,

$$\mathrm{ch}_{M'}(\pi(a)) = \{\pi(b) \mid b \in \mathrm{ch}_M(a)\}.$$

For example, any system map from a graph G into V is just a decoration of G.

The following result, which indicates how system maps preserve bisimulations, will be of use later.

Lemma 7.8.7 Let $\pi_1, \pi_2 : M \to M'$ be system maps.

(i) If R is a bisimulation on M, then $R' = (\pi_1 \times \pi_2)R$ is a bisimulation on M', where we define

$$(\pi_1 \times \pi_2)R = \{(\pi_1(a_1), \pi_2(a_2)) \mid a_1 R a_2\}.$$

(ii) If S' is a bisimulation on M', then $S = (\pi_1 \times \pi_2)^{-1}S'$ is a bisimulation on M, where we define

$$(\pi_1 \times \pi_2)^{-1}S' = \{(a_1, a_2) \in M \times M \mid (\pi_1(a_1))S'(\pi_2(a_2))\}.$$

Proof: (i) Let $b_1 R' b_2$ and suppose $b_1' \in \mathrm{ch}_{M'}(b_1)$. I show that there is a $b_2' \in \mathrm{ch}_{M'}(b_2)$ such that $b_1' R' b_2'$. Let a_1, a_2 be such that $b_1 = \pi(a_1), b_2 = \pi(a_2), a_1 R a_2$. Since $b_1' \in \mathrm{ch}_{M'}(b_1)$, there is an $a_1' \in \mathrm{ch}_M(a_1)$ such that $b_1' = \pi(a_1')$. Since R is a bisimulation, there is an $a_2' \in \mathrm{ch}_M(a_2)$ such that $a_1' R a_2'$. Let $b_2' = \pi(a_2')$. Then b_2' is as required.

Likewise, if $b_1 R' b_2$ and $b_2' \in \mathrm{ch}_{M'}(b_2)$, then there is a $b_1' \in \mathrm{ch}_{M'}(b_1)$ such that $b_1' R' b_2'$. Thus R' is a bisimulation on M'.

(ii) This is entirely analogous to the proof of part (i). □

Suppose now we have a system M and a bisimulation R on M that is also an equivalence relation on M. A system M' is said to be a *quotient* of M by R if and only if there is a surjective map $\pi : M \to M'$ such that for all $a, b \in M$,

$$aRb \Leftrightarrow \pi(a) = \pi(b).$$

Our main interest in quotients here concerns the extensional ones. The following lemma supplies some information about this.

Lemma 7.8.8 Let R be a bisimulation equivalence relation on a system M, and let $\pi : M \to M'$ be the corresponding quotient of M. Then M' is extensional if and only if R is the relation $\equiv_M$.

Proof: Suppose R is the relation $\equiv_M$. Let $\pi(a) \equiv_M \pi(b)$. I show that $\pi(a) = \pi(b)$. By Lemma 7.8.7(ii), $R' = (\pi \times \pi)^{-1}R$ is a bisimulation on M such that $aR'b$. Thus $a \equiv_M b$. But $\pi : M \to M'$ is the quotient of M by $\equiv_M$ (since this is R), so this implies that $\pi(a) = \pi(b)$.

Conversely, suppose that M' is extensional. I show that if S is any small bisimulation on M, and if aSb, then aRb, which at once implies that R is $\equiv_M$. By Lemma 7.8.7(i), $S' = (\pi \times \pi)S$ is a bisimulation on M' such that $\pi(a)S'\pi(b)$. Thus $\pi(a) \equiv_M \pi(b)$. Hence as M' is extensional, $\pi(a) = \pi(b)$. Thus aRb, as required. □

Using the above lemma, I can prove that every system, M, has an extensional quotient. The overall approach is as follows: take the bisimulation equivalence relation $\equiv_M$ on M, and construct a map π with domain M such that for all $a, b \in M$,

$$(*) \qquad \pi(a) = \pi(b) \Leftrightarrow a \equiv_M b.$$

In the case where M is a set, there is no difficulty in carrying out such a construction—it is all quite standard. The elements of the new system M' are taken to be the equivalence classes of M under the equivalence relation $\equiv_M$, and π maps each element of M to its equivalence class.

But in the case where M is a proper class, problems arise if any of the equivalence classes is a proper class. To circumvent this difficulty, the usual trick when working in well-founded Zermelo–Fraenkel set theory is to define 'equivalence classes' as being subsets of the least level of the cumulative hierarchy (of sets) at which they are nonempty. That is, given any $a \in M$, take the 'equivalence class' of a modulo $\equiv_M$ to be the *set*

$$\{b \in V_\alpha \mid b \in M \ \&\ a \equiv_M b\},$$

where α is minimal such that this collection is nonempty.

But in the absence of Foundation, this approach will not work. Instead, we adopt the following alternative.

For each $a \in M$, the set M_a is (by the Axiom of Choice) in one-one correspondence with some ordinal number, and this induces an isomorphism between the graph M_a and a corresponding graph whose domain is an ordinal. Let T_a be the class of all graphs with domain an ordinal, that are isomorphic to M_b for some $b \in M$ such that $a \equiv_M b$. Let

$$\pi(a) = \{G \in V_\alpha \mid G \in T_a\}$$

where α is the least ordinal such that this set is nonempty. I show that this definition satisfies $(*)$, as required.

If $a_1 \equiv_M a_2$, then $T_{a_1} = T_{a_2}$, so $\pi(a_1) = \pi(a_2)$. Conversely, if $a_1, a_2 \in M$ are such that $\pi(a_1) = \pi(a_2)$, then there is a graph G such that $G \in T_{a_1}$ and $G \in T_{a_2}$. Since $G \in T_{a_1}$, there is an $a_1' \in M$ such that $a_1 \equiv_M a_1'$ and $G \cong M_{a_1'}$. Likewise, as $G \in T_{a_2}$, there is an $a_2' \in M$ such that $a_2 \equiv_M a_2'$ and $G \cong M_{a_2'}$. Then $M_{a_1'} \cong M_{a_2'}$, so by Lemma 7.8.5(ii), $a_1' \equiv_M a_2'$. Thus $a_1 \equiv_M a_2$.

Theorem 7.8.9 Let M be any system. The following are equivalent:

(i) M is extensional;

(ii) for each (small) system M_0 there is at most one system map

$$\pi : M_0 \to M;$$

(iii) for each system M', every system map $\pi : M \to M'$ is one-one.

Proof: (i) $\Rightarrow$ (ii). Let $\pi_1, \pi_2 : M_0 \to M$ be system maps. By Lemma 7.8.7(i), $R = (\pi_1 \times \pi_2)(=_{M_0})$ is a bisimulation on M, where $=_{M_0}$ is the identity relation on M_0. Now, if $m \in M_0$, then $(\pi_1(m))R(\pi_2(m))$, so $\pi_1(m) \equiv_M \pi_2(m)$, and hence by (i), $\pi_1(m) = \pi_2(m)$. Thus $\pi_1 = \pi_2$, proving (ii).

(ii) $\Rightarrow$ (i). (For arbitrary systems M_0.) Let M_0 be the system whose nodes are the pairs (a, b) such that $a \equiv_M b$, and whose edges are all $(a, b) \longrightarrow (a', b')$ where $a \longrightarrow a'$ and $b \longrightarrow b'$ in M. Define $\pi_1, \pi_2 : M_0 \to M$ by $\pi_1(a, b) = a$, $\pi_2(a, b) = b$. It is routine to verify that π_1 and π_2 are system maps. Thus by (ii), $\pi_1 = \pi_2$, and hence $a = b$ whenever $a \equiv_M b$, proving (i).

(For small systems M_0.) It suffices to show that (ii) for small systems implies the unrestricted form of (ii). Let M_0 be a system, and let $\pi_1, \pi_2 : M_0 \to M$ be system maps. Let $a \in M_0$. Then $(M_0)_a$ is a small system, and $\pi_1 \quad (M_0)_a = \pi_2 \quad (M_0)_a$. In particular, $\pi_1(a) = \pi_2(a)$. But $a \in M_0$ was arbitrary. Hence $\pi_1 = \pi_2$.

(i) $\Rightarrow$ (iii). Let $\pi : M \to M'$ be a system map. By Lemma 7.8.7(ii), $R = (\pi \times \pi)^{-1}(=_{M'})$ is a bisimulation on M (where $=_{M'}$ is the identity relation on M'). So, if $\pi(a) = \pi(b)$, then aRb, so $a \equiv_M b$, whence by (i), $a = b$. Thus π is one-one, as required.

(iii) $\Rightarrow$ (i). Let $\pi : M \to M'$ be an extensional quotient of M. By (iii), π is one-one. Hence $\pi : M \cong M'$. So, as M' is extensional, so too is M. □

I am now ready to give the construction of a model of the theory ZFC$^-$ + AFA.

Given a system M, an *M-decoration* of a graph G is just a system map $\pi : G \to M$.

Thus, in particular, a V-decoration of G is simply a decoration of G.

I call a system M *complete* if every graph has a unique M-decoration. (AFA says that V is a complete system.)

By Theorem 7.8.9, every complete system is extensional.

Let V_0 be the class of all graphs. Notice that every member of V_0 is of the form G_a, where G is a graph and a is a node of G. Using this observation, we make V_0 into a system by introducing the edges $G_a \longrightarrow G_b$ whenever G is a graph and $a \longrightarrow b$ in G.

Let $\pi_c : V_0 \to V_c$ be the extensional quotient of V_0.

Lemma 7.8.10 For each system M, there is a unique system map

$$\pi : M \to V_c.$$

Proof: If $a \in M$, then $M_a \in V_0$. Define $\pi : M \to V_0$ by $\pi(a) = M_a$. Clearly, π is a system map. Then $\pi_c \circ \pi : M \to V_c$ is a system map, which is unique by virtue of Theorem 7.8.9. □

Corollary 7.8.11 V_c is complete.

Proof: Immediate. □

Given any system M, we may obtain an interpretation of the language of set theory by letting the variables range over the nodes of M, and interpreting the predicate symbol '$\in$' by the relation $\in_M$ defined on M by

$$a \in_M b \text{ if and only if } b \longrightarrow a \text{ in } M$$

for all $a, b \in M$.

By virtue of the above corollary, the following result, which will be proved in just a moment, establishes the consistency (relative to that of the theory ZF$^-$) of the theory ZFC$^-$ + AFA.

Theorem 7.8.12 Every complete system is, under the interpretation described above, a model of ZFC$^-$ + AFA. □

Combining this theorem with Corollary 7.8.11, we see that V_c is a model of ZFC$^-$ + AFA. In fact, by virtue of Lemma 7.8.10, there is a unique system map $\pi : V \rightarrow V_c$, so V_c is a model of ZFC$^-$ + AFA that canonically embeds V. Thus we may regard our construction of the model V_c as providing an *extension* of the universe V. This gives the result stated as Theorem 7.2.3.

Call a system M *full* if for every set $u \subseteq M$, there is a unique element $a \in M$ such that $u = \mathrm{ch}_M(a)$.

For example, V is a full system, as is W, the class of all well-founded sets.

Lemma 7.8.13 Every complete system is full.

Proof: Let M be a complete system. Let $u \subseteq M$ be a set. Let G_0 be the graph consisting of all nodes and edges of M that lie on paths starting from a node in u. Obtain G from G_0 by adding one more node, t, together with edges $t \longrightarrow x$ for all $x \in u$.

Since M is complete, G has a unique M-decoration, d. Let $d_0 = d \quad G_0$. Then d_0 is an M-decoration of G_0. But the identity map is clearly the unique M-decoration of G_0. Hence $d_0(x) = x$ for all $x \in G_0$. So if we set $a = d(t)$, then $a \in M$ and

$$\begin{aligned} \mathrm{ch}_M(a) &= \{d(x) \mid t \longrightarrow x \text{ in } G\} \\ &= \{x \mid t \longrightarrow x \text{ in } G\} \\ &= u \end{aligned}$$

For uniqueness, suppose $a' \in M$ is also such that $\mathrm{ch}_M(a') = u$. Then we may define an M-decoration d' of G by setting $d'(t) = a'$, and $d'(x) = x$ for all $x \in G_0$. So by the uniqueness of d, $d' = d$. Hence, in particular,

$$a' = d'(t) = d(t) = a.$$

The proof is complete. □

Theorem 7.8.14 Every full system is a model of ZFC$^-$.

Proof: Let M be a full system. Fullness tells us that for each set $u \subseteq M$ there is a unique $a \in M$ such that $u = \mathrm{ch}_M(a)$. We shall denote this unique a by u^M. Using this notation, we check each of the axioms of ZFC$^-$ in turn.

Extensionality. Let $a, b \in M$ be such that

$$M \models (\forall x)(x \in a \leftrightarrow x \in b).$$

Then $\mathrm{ch}_M(a) = \mathrm{ch}_M(b)$. But $a = (\mathrm{ch}_M(a))^M$ and $b = (\mathrm{ch}_M(b))^M$. Hence $M \models a = b$.

Pairing. Let $a, b \in M$. Then $\{a, b\} \subseteq M$, so let $c = \{a, b\}^M$. Clearly,

$$M \models [a \in c \wedge b \in c].$$

Union. Let $a \in M$. Then $x = \bigcup\{\mathrm{ch}_M(y) \mid y \in \mathrm{ch}_M(a)\}$ is a subset of M, so let $c = x^M$. Then

$$M \models (\forall y \in a)(\forall z \in y)(z \in c).$$

Power set. Let $a \in M$. Then $x = \{y^M \mid y \subseteq \mathrm{ch}_M(a)\}$ is a subset of M, so let $c = x^M$. Then

$$M \models \forall x[(\forall z \in x)(z \in x) \to (x \in c)].$$

Infinity. Let

$$\begin{aligned} \theta_0 &= \emptyset^M, \\ \theta_{n+1} &= (\mathrm{ch}_M(\theta_n) \cup \{\theta_n\})^M, \quad \text{for } n = 0, 1, 2, \ldots . \end{aligned}$$

Then $\theta_n \in M$ for all n, so

$$\theta = \{\theta_n \mid n = 0, 1, 2, \ldots\}^M \in M.$$

Clearly,

$$M \models [\theta_0 \in \theta \wedge (\forall x \in \theta)(\exists y \in \theta)(x \in y)].$$

Separation. Let $a \in M$, and let $\phi(x)$ be a formula, possibly containing constants for elements of M, with at most the variable x free, and set

$$c = \{b \in \mathrm{ch}_M(a) \mid M \models \phi(b)\}^M.$$

Then

$$M \models \forall x(x \in c \;\leftrightarrow\; x \in a \wedge \phi(x)).$$

Collection. Let $a \in M$, and let $\phi(x, y)$ be a formula, possibly containing constants for elements of M, with at most the variables x and y free, and suppose that

$$M \models (\forall x \in a)(\exists y)\phi(x, y).$$

Then

$$(\forall x \in \mathrm{ch}_M(a))(\exists y)[y \in M \;\&\; M \models \phi(x, y)].$$

By the Collection Schema, there is a set b such that

$$(\forall x \in \mathrm{ch}_M(a))(\exists y \in b)[y \in M \;\&\; M \models \phi(x, y)].$$

Let $c = (b \cap M)^M$. Then

$$M \models (\forall x \in a)(\exists y \in c)\phi(x, y).$$

Choice. Let $a \in M$ be such that

$$M \models (\forall x \in a)(\exists y)(y \in x)$$

and

$$M \models (\forall x_1, x_2 \in a)[\exists y(y \in x_1 \wedge y \in x_2) \rightarrow (x_1 = x_2)].$$

Then

$$(\forall x \in \mathrm{ch}_M(a))(\mathrm{ch}_M(x) \neq \emptyset),$$

and, for all $x_1, x_2 \in \mathrm{ch}_M(a)$,

$$\mathrm{ch}_M(x_1) \cap \mathrm{ch}_M(x_2) \neq \emptyset \;\Rightarrow\; x_1 = x_2.$$

Thus $\{\mathrm{ch}_M(x) \mid x \in \mathrm{ch}_M(a)\}$ is a set of nonempty, pairwise-disjoint sets. So by the Axiom of Choice there is a set b such that for each $x \in \mathrm{ch}_M(a)$, the set $b \cap \mathrm{ch}_M(x)$ has a unique element $c_x \in M$. Then $c = \{c_x \mid x \in \mathrm{ch}_M(a)\}^M$ is such that

$$M \models (\forall x \in a)(\exists y \in x)(\forall u \in x)[u \in c \leftrightarrow u = y].$$

The proof is complete. □

By virtue of Lemma 7.8.13, the above result tells us that every complete system M is a model of ZFC$^-$. Thus the following completes our proof of Theorem 7.8.12.

Theorem 7.8.15 Every complete system is a model of AFA.

Proof: Let M be a complete system. For $a, b \in M$, define the "M-ordered pair" $(a,b)_M$ of a, b by

$$(a,b)_M = \{\{a\}^M, \{a,b\}^M\}^M.$$

(Thus, within M, $(a,b)_M$ has the standard set-theoretic structure of the usual ordered pair of a, b.)

Now, a *graph* is, officially, an ordered pair consisting of a set and a binary relation on that set. Thus for $c \in M$,

$$M \models \text{“}c \text{ is a graph”}$$

if and only if there are $a, b \in M$ such that $c = (a,b)_M$ and

$$M \models \text{“}b \text{ is a binary relation on } a\text{”}.$$

This last requirement reduces to

$$\mathrm{ch}_M(b) \subseteq \{(x,y)_M \mid x, y \in \mathrm{ch}_M(a)\}.$$

Hence, if $c \in M$ is such that $M \models$ "c is a graph", we may define a genuine graph G by taking a, b as above and letting the elements of $\mathrm{ch}_M(a)$ be the nodes of G and the pairs (x, y) such that $(x,y)_M \in \mathrm{ch}_M(b)$ the edges. Since M is complete, G has a unique M-decoration, d. Then $d : \mathrm{ch}_M(a) \to M$, and for all $x \in \mathrm{ch}_M(a)$,

$$d(x) = \{d(y) \mid (x,y)_M \in \mathrm{ch}_M(b)\}.$$

Set

$$f = \{(x, d(x))_M \mid x \in \mathrm{ch}_M(a)\}^M.$$

Then $f \in M$, and it is routine to verify that

$$M \models \text{“}f \text{ is the unique decoration of the graph } G\text{”}.$$

The proof is complete. □

Bibliography

[1] P. Aczel, *Non-Well-Founded Sets*, CSLI Lecture Notes, Vol.14, CSLI Publications, Stanford, California, 1988.

[2] J. Barwise and J. Etchemendy, *The Liar*, Oxford University Press, London, 1987.

[3] J. L. Bell, *Boolean-Valued Models and Independence Proofs in Set Theory*, Oxford University Press, London, 1977.

[4] K. J. Devlin, *Constructibility*, Springer-Verlag, New York, 1983.

[5] P. R. Halmos, *Naive Set Theory*, Van Nostrand, Princeton, N.J., 1960.

[6] P. R. Halmos, *Lectures on Boolean Algebras*, Van Nostrand, Princeton, N.J., 1963.

[7] J. D. Monk, *Introduction to Set Theory*, McGraw-Hill, New York, 1969.