Coalgebra and Circularity

Lawrence S. Moss

June 16, 2008

Abstract

These notes lecture notes for my 2008 MATHLOGAPS course. The sources for them are

1. My article on *Set theory and circularity* for the Stanford Encyclopedia of Philosophy.
2. Lecture notes from a course on coalgebra for mathematics students at Indiana University.

What appears here is more than I will be able to cover in the MATHLOGAPS course, but rather less than the semester-long coalgebra. My plans for the course are also to emphasize the elementary category theory instead of the basic set theory.

Contents

1 Circular Phenomena in Set Theory

It is difficult to say in a general way what makes a definition circular. In this course we are concerned exclusively with mathematical definitions of various sorts. Consider the equation $x = \frac{1}{2}x+1$. Is this a *circular definition* of the number 2? In a sense, it is just that: a number has been defined in terms of itself. But there is nothing problematic about this equation, and so one may wonder why this is in the same class of equations as $x = x + 1$ or $x = x$. In the set theoretic setting, we often employ circular definitions and characterizations of sets and classes. For example, the collection *HF* of *von Neumann natural numbers* may be characterized by

$$HF \text{ is the set of all } x \text{ such that } x \text{ is a finite subset of } HF. \tag{1}$$

With a bit of work, it can be shown that (1) defines a unique set in standard set theory *ZFC*. (1) is more of a *characterization* than a textbook definition, however. In other words, if one were presented with (1) as a putative definition, then the first step in understanding it would be to "straighten out" the circularity by providing a different definition D of a set, then to check that every set satisfying D satisfies the property defining *HF*, and vice-versa.

It is easier to think about circular *objects* than circular *definitions*. Even so, it will be useful in reading this course to keep circular definitions in mind. The most conspicuous form of object circularity would be a set having itself as an element; even worse would be a set x such that $x = \{x\}$. For those with a background in standard set theory, such sets are ruled out by the axioms in the first place, and in the second it is not clear why one would want to change the axioms in order to admit them. And if one does take the drastic step of altering the axioms of a well-established theory, what changes? This course is an extended discussion of this matter, and related ones.

1.1 Streams

Many of the ideas in this article may be illustrated using *streams*. A *stream of numbers* is an ordered pair whose first coordinate is a number and whose second coordinate is again a stream of numbers. The first coordinate is called the *head*, and the second the *tail*. The tail of a given stream might be different from it, but again, it might be the very same stream. For example, consider the stream s whose head is 0 and whose tail is s again. Thus the tail of the tail of s is s itself. We have $s = \langle 0, s\rangle$, $s = \langle 0, \langle 0, s\rangle\rangle$, etc. This stream s exhibits object circularity. It is natural to "unravel" its definition as

$$(0, 0, \ldots, 0, \ldots).$$

We are purposely using notation for the unraveled version of s which differs from the pairing notation in order to emphasize the conceptual difference. The best way to understand the unraveled form is as an *infinite sequence*; standardly, infinite sequences are taken to be functions whose domain is the set N of natural numbers. So we can take the unraveled form to be the constant function with value 0. Whether we want to take the stream s described above to *be* this function is an issue we want to explore in a general way. Notice that since we defined s to be an ordered pair, it follows from the way pairs are constructed in ordinary mathematics that s will not itself *be* the constant sequence 0.

One way to define streams is with *systems of equations* for them. For example, here is such a system:

$$\begin{aligned} x &\approx \langle 0, y\rangle \\ y &\approx \langle 1, z\rangle \\ z &\approx \langle 2, x\rangle \end{aligned} \tag{2}$$

We should comment on the $\approx$ notation here. We are concerned with modeling various types of ordinary mathematical objects in set theory, and one kind of object that we want to model will be that of a *system of equations*. This is an unusual thing to do. In anticipation of things to come, we use the $\approx$ sign for *equations we would like to solve*. So in our discussion of $x = \frac{1}{2}x + 1$ above, we would prefer to write $x \approx \frac{1}{2}x + 1$. The point is that '$x$' here is a symbol, but whatever we take symbols to be, it will almost never be the case that the symbol x is identical to the expression '$\frac{1}{2}x + 1$' or to anything related to it. For the solution to an equation or a system of them, we will use a "dagger" to refer to the solution. Thus for this equation, $x^\dagger = 2$; the reason that 2 satisfies the equation is that $x^\dagger = \frac{1}{2}x^\dagger + 1$ (and here we use $=$ rather than $\approx$).

Returning to (2), we take it to define streams $x^\dagger$, $y^\dagger$, and $z^\dagger$. These satisfy equations: $x^\dagger = \langle 0, y^\dagger\rangle$, $y^\dagger = \langle 1, z^\dagger\rangle$, and $z^\dagger = \langle 2, x^\dagger\rangle$. These streams then have unraveled forms. For example, the unraveled form of $y^\dagger$ is $(1, 2, 0, 1, 2, 0, \ldots)$.

There is a natural operation of "zipping" two streams. Also called "merging", it is defined by

$$\mathsf{zip}(s, t) \quad = \quad \langle \mathsf{head}(s), \mathsf{zip}(t, \mathsf{tail}(s))\rangle \tag{3}$$

So to zip two streams s and t one starts with the head of s, and then begins the same process of zipping all over again, but this time with t first and the tail of s second. For example, if $x^\dagger$, $y^\dagger$, and $z^\dagger$ are the solutions to the system in (2) above, then we have $\mathsf{zip}(x^\dagger, y^\dagger)$. In unraveled form, this is

$$(0, 1, 1, 2, 2, 0, 0, 1, 1, 2, 2, 0, \ldots).$$

But please note that our definition of zip does not work by recursion as one might expect; for on thing, there are no "base cases" of streams.

We can even ask about solving systems of equations written in terms of zip. It is easy to see that an equation like $x = \mathsf{zip}(x, x)$ is satisfied by all and only the constant streams. One like

$$x = \mathsf{zip}(\mathsf{head}(x) + 1, x)$$

has no solutions whatsoever. But if we do things right, we can define very interesting streams. For example, consider

$$\begin{aligned} x &\approx \langle 1, \mathsf{zip}(x, y)\rangle \\ y &\approx \langle 0, \mathsf{zip}(y, x)\rangle \end{aligned} \tag{4}$$

The system has a unique solution. The unraveled form of $x^\dagger$ begins as

$$(1,1,0,1,0,0,1,1,0,0,1,0,1,1,0,1,0,\ldots)$$

that of $y^\dagger$ begins $(0,0,1,0,1,1,0,0,1,1,0,1,0,0,1,0,1,\ldots)$. The first of these is a famous sequence, the *Thue-Morse* sequence t (actually it is t without its first entry, 0).[1]

The reduction of streams to functions We have been careful to emphasize the difference between streams as we originally spoke of them and their unraveled form as functions on the natural numbers. At this point we want to look at this matter more closely.

Before we turn to the details, let us consider the parallel matter of *sequences* construed as functions on the natural numbers. Anyone who teaches about (infinite) sequences of some sort, say sequences of integers or real numbers, may at some point need to say what a sequence actually is. Surely this is not done very often in elementary presentations: usually one would give examples instead of a formal definition, or illustrate what sequences are for by using them in some way or other. In any case, it happens that in the usual set-theoretic modeling of mathematics, sequences of real numbers would be taken to *be* functions from the set of natural numbers to the set of real numbers. So we have a *reduction* of one kind of object, sequences, to another, functions. Of course, functions are then reduced to sets of ordered pairs, ordered pairs to sets of a certain form, natural numbers to sets of yet another form, and real numbers in their own way. Concerning this kind of reduction, we should always ask whether it is necessary or silly, and whether it is useful to those using the mathematical objects in the first place. All of this is worth keeping in mind as we turn back to the sequences.

Let N^∞ be the set of streams of natural numbers, and let ${}^N N$ be the set of functions from N to N. The reduction employs two functions

$$\varphi : N^\infty \to {}^N N \qquad \psi : {}^N N \to N^\infty$$

defined as follows: For φ, we first take a stream s to a function $f_s : N \to N^\infty$. This time we use recursion:

$$\begin{array}{lcl} f_s(0) & = & s \\ f_s(n+1) & = & \mathsf{tail}(f_s(n)) \end{array}$$

Then from f we get a function $\varphi(s) : N \to N$ by $g(n) = \mathsf{head}(f_s(n))$. This defines φ, the precise definition of what we spoke of earlier by the name *unraveling*. In the other direction, we need *infinite* systems of equations. Given a function $f : N \to N$, consider

$$\begin{array}{lcl} x_0 & = & \langle f(0), x_1\rangle \\ x_1 & = & \langle f(1), x_2\rangle \\ & \cdots & \\ x_n & = & \langle f(n), x_{n+1}\rangle \\ & \cdots & \end{array} \tag{5}$$

[1]The interesting feature of the Thue-Morse sequence is that is has no three-in-a-row repeats: for every finite non-empty sequence w of 0's and 1's, the infinite sequence t does not contain www anywhere.

Then this system has a solution, and we take $\psi(f) = x_0^\dagger$. It is then possible to show that $\psi \circ \varphi$ is the identity on N^∞ and $\varphi \circ \psi$ is the identity on ${}^N N$. In plainer terms, we can pass back and forth from streams to functions on numbers.

At this point, we can ask whether questions about the reduction. The first question that come to might concern the ontological status of the entities:

> Let A be a collection of abstract objects (say functions from natural numbers to natural numbers) exists, and suppose that A one believes that the objects in A exist. Let B be a different collection of abstract objects. Suppose that A and B correspond in a natural way, and that everything one says about objects in B could well be said about their correspondents in A, perhaps using different language. Should one believe that the objects in B also exist?

Asking this about streams and functions on N is no different than asking it for any other kind of reduction of mathematical objects. Any discussion of it would take us to issues in the philosophy of mathematics that go beyond our goals in this course. However, there are two additional points to be made on this matter.

First, the standard modeling of pairs in set theory[2] would have us believe that from the beginning of this section onwards, we have been talking about things which do not exist: as we have literally defined them, there are no streams of numbers whatsoever! We discuss this at length in Section 4.2, when we talk about the Foundation Axiom of set theory. The point is is that this axiom forbids object-level circularity in a way that precludes streams in the exact form that we have them. Thus if one wants to model the intuitive notion of a stream as we have introduced it, one would need to say something like: "By a stream, we mean a function on numbers. We adopt special notation to make it look like streams are pairs of a certain sort, but deep down they are just functions on numbers."

Continuing with questions about the reduction of streams to functions, we can ask whether there is any conceptual difference using streams as opposed to functions. Certainly these represent different points of view, and for this reason it should be useful to have both available. To see the difference, let us return to the matter of zipping streams. Done in terms of functions $f, g : N \to N$, the zipped version would be

$$\mathsf{zip}(f,g)(n) \quad = \quad \begin{cases} f(\frac{n}{2}) & \text{if } n \text{ is even} \\ g(\frac{n-1}{2}) & \text{if } n \text{ is odd} \end{cases}$$

It would be harder to use this to turn equation (4) into the definition of two sequences by recursion.[3] The upshot is that we can start to see some kind of conceptual difference when

[2]The main move in the theory of hypersets is to keep the standard modeling of pairs but to change the underlying set theory. Formulating matters this way suggests that it might be possible to keep the Foundation Axiom but alter the treatment of pairs. This point is made by Forster [13] and Paulson [23]. We discuss this point further at the end of Section 2.10.

[3]Incidentally, the most standard definition of the solution $x^\dagger$ from equation (4) uses a different vocabulary altogether: the nth term of the sequence corresponding to $x^\dagger$ is 0 or 1 depending on whether there are an even or odd number of 1's in the binary representation of $n+1$.

we use one kind of representation instead of another. And this brings us to our second point on the reduction of streams to functions: conceptual differences worth exploring may be hidden under the surface of such a reduction.

At this point, we are done with our discussion of streams. Of course we shall revisit them in later sections to illustrate various points. We also broadly foreshadow the main points of this course:

- By changing the usual axioms of set theory, one can model circularly defined streams and other objects in a way which is closer to the intuition that one has about them. In particular, it is possible to work with object-level circularity in a relatively consistent set theory, and there may be reasons why one would want to do so.
- In the changed theory, we also find different results on collections defined in terms of themselves. We have already seen such a collection, *HF* from (1). The status of circular definitions changes when one alters the set theory, and this leads to a broader examination of several issues.
- There is also a deeper conceptual issue going far beyond set theory related to *top-down* vs. *bottom-up* treatments of various phenomena.

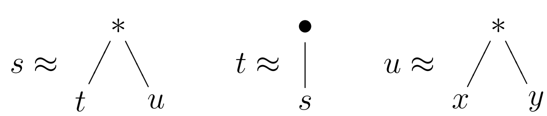
1.2 Infinite trees

We want to move from streams to a more complicated example, infinite trees. Some of the points that we make will be closely related to what we have seen for trees, and some will raise new issues.

Here is a class of objects which we shall call *trees*[4]:

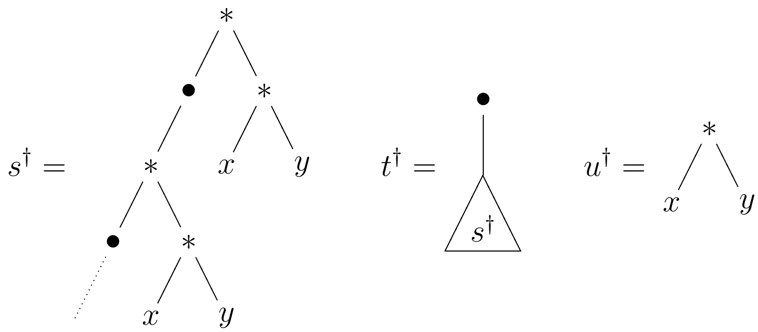
1. The variables x and y alone are trees.
2. If t is a tree, then adding a single node labeled $\bullet$ as a new root with t as its only subtree gives a tree.
3. If s and t are trees, then adding a single node labeled $*$ as a new root with s as the left subtree and t as the right subtree again gives a tree.
4. Trees may go on forever.

Trees may be specified by *tree systems* (of equations). Here is one such system:



[4]This is really but one species of tree. It would be more proper to use a fuller name for what we are studying, such as *finite or infinite trees whose leaves are labeled with* x *or* y *and whose interior nodes are either unary nodes labeled* $\bullet$ *or binary nodes labeled* $*$.

Again, we use the $\approx$ notation in variables for which we want to solve, and we superscript variables with a dagger in the solution. In this case, the one and only solution of this system might be pictured as



It will be useful to recast the definition of our trees in terms of pairs and triples:

1. The symbols x and y alone are trees.
2. If t is a tree, then $\langle \bullet, t \rangle$ is a tree.
3. If s and t are trees, then $\langle *, s, t \rangle$ is a tree.
4. Trees may be "infinitely deep".

Then our system above is

$$\begin{array}{lcl} s & \approx & \langle *, t, u \rangle \\ t & \approx & \langle \bullet, s \rangle \\ u & \approx & \langle *, x, y \rangle \end{array} \tag{6}$$

So now we have something that looks more like what we have seen with streams. But with streams we had an unraveled form, and so we might wonder what the unraveled form of trees is. To some extent, it would be the pictures that we have already seen. In particular, one could take a tree as we have defined them and give a description of how one would contruct the picture. (The full contruction would take forever, of course, but the same is true of our work on streams.) Conversely, given a picture, one could set down a tree system for it, where a "tree system" is a system of equations as in equation (6). (In general, the tree system would be infinite, but if you find a *regular structure* in the picture, then the system could be finite.)

On the other hand, pictures are not entirely respectable as standard mathematical objects, despite the work that has gone on and continues to go on to rehabilitate them. For work on trees, one would need a more complicated set of definitions. We are not going to present any of this.

More 'cheating' Let Tr be the set of trees that we have been discussing. Then our definition in terms of Tr would have

$$Tr \quad = \quad \{x, y\} \cup (\{\bullet\} \times Tr) \cup (\{*\} \times Tr \times Tr). \tag{7}$$

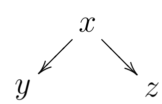
Now again the standard modeling in set theory gives us a problem: one can prove in *ZF* set theory that $Tr = \emptyset$, and this runs afoul of our pictures and intuition. The standard way out is to change the equals sign $=$ in (7) to something else. For most mathematical work this is perfectly fine, but it is the kind of move we explore.

1.3 Hypersets

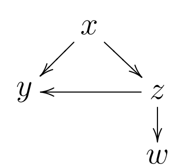
Let us turn from streams and trees to sets. Before presenting some analogs to what we have just seen, at *pictures of sets*. To make the discussion concrete, consider the set

$$x = \{\emptyset, \{\{\emptyset\}, \emptyset\}\}$$

Let us call this set x. We want to draw a picture of this set, so we start with a point which we think of as x itself. Since x has two elements, we draw add two children:

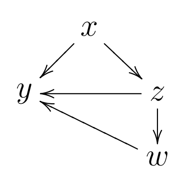


Again, we draw arrows on behalf of the members. We take y to be $\emptyset$ and z to be $\{\{\emptyset\}, \emptyset\}$. We do not add any children of y because it is empty. But we want to add two children to z, one for $w = \{\emptyset\}$ and one for $\emptyset$. So we have

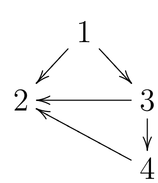


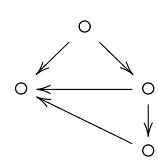
(8)

We conclude by putting an arrow from w to y, since $\emptyset \in \{\emptyset\}$.

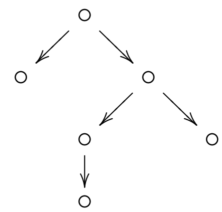


Now we want to forget the identity of the nodes. We could either trade in the four sets that we used for numbers (to mention just one way), or else finesse the issue entirely. We would get one of the pictures below:





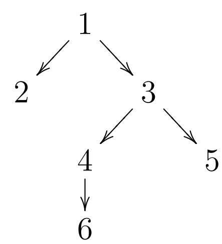
Incidentally, in building this graph, we allowed ourselves to share the node y both times we came to $\emptyset$. It would be possible to avoid doing this, using different nodes. The end result would be a tree:



A *graph* is a pair $(G, \rightarrow)$, where $\rightarrow$ is a relation on G (a set of ordred pairs from G). The idea is that we want to think of a graphs as *notations for sets*, just as systems of equations were notation for streams. This is explained by the concept of a *decoration*: A decoration d of a graph G is a function whose domain is G and with the property that

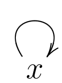
$$d(g) = \{d(h) : g \rightarrow h\}.$$

For example, let us introduce names for the nodes in the tree-like graph and then find its decoration:



Since 6 has no children, $d(6)$ must be $\emptyset$. Similarly, $d(5)$ and $d(2)$ are also $\emptyset$. $d(4) = \{d(6)\} = \{\emptyset\}$. $d(3) = \{d(4), d(5)\} = \{\{\emptyset\}, \emptyset\}$. And $d(1) = \{d(2), d(3)\} = \{\emptyset, \{\{\emptyset\}, \emptyset\}\}$. Note this is the set x with which we started. This is no accident, and you are encouraged to think about why this is true. A related point: for a graph like the one in equation (8), where we use the sets involved as the nodes of the graph, you should check that the identity function is a decoration.

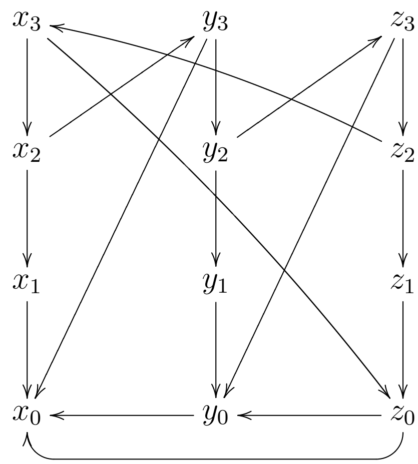
However, things get more interesting with an example like the loop graph



Let d be a decoration of this graph. Then we would have $d(x) = \{d(x)\}$. So writing Ω for $d(x)$, we have $\Omega = \{\Omega\}$. This set Ω is the most conspicuous example of object circularity: a set that is a member of itself. (Indeed, Ω is its own only member.)

Finally, we want to consider an example that harks back to the stream system (2) in

Section 1.1.



(9)

Let us try to understand what a decoration d of this graph would be. In order to follow the discussion below, you should remember from set theory that the standard rendering of the first few natural numbers is by

$$0 = \emptyset, \quad 1 = \{\emptyset\}, \quad 2 = \{0, 1\} = \{\emptyset, \{\emptyset\}\}$$

and also that the standard definition of the ordered pair $\langle x, y\rangle$ is as $\{\{x\}, \{x, y\}\}$.

Since x_0 has no children, $d(x_0)$ must be $\emptyset$. Then it follows that $d(y_0) = \{d(x_0)\} = \{\emptyset\} = 1$. And now $d(z_0) = \{d(x_0), d(y_0)\} = \{0, 1\} = 2$. Furthermore, $d(z_1) = \{2\}$. It follows now that

$$d(x_1) = \{0\}, \quad d(y_1) = \{1\}, \quad d(z_1) = \{2\}.$$

And then

$$\begin{array}{rcll} d(x_2) & = & \{d(y_3), d(x_1)\} & = \{\{0, d(y_2)\}, \{0\}\} & = & \langle 0, d(y_2)\rangle \\ d(y_2) & = & \{d(z_3), d(y_1)\} & = \{\{1, d(z_2)\}, \{1\}\} & = & \langle 1, d(z_2)\rangle \\ d(z_2) & = & \{d(x_3), d(z_1)\} & = \{\{2, d(x_2)\}, \{2\}\} & = & \langle 2, d(x_2)\rangle \end{array}$$

The upshot is that we can go back to our original stream system in equation (2) and then solve it by putting down our big graph and decorating it. The solution would be

$$x^\dagger = d(x_2) \qquad y^\dagger = d(y_2) \qquad z^\dagger = d(z_2).$$

A *hyperset* or *non-wellfounded set* is a set that is obtained by decorating an arbitrary graph.

Another way of thinking about hypersets is in terms of *systems of set equations*, as we have done it for streams and trees. By such a system we mean a set X which we think of as variables (any set will do), and then a function e from X to its power set $\mathcal{P}X$. That is, the value of e on each variable is again a set of variables. Set systems and related concepts correspond to ones for graphs in the following way:

the graph $(G, \to)$	the system of set equations (X, e)
the nodes of G	the set X of variables
the relation $\to$ on the nodes	the function $e : X \to \mathcal{P}X$
the children of x in G	the set $e(x) \in \mathcal{P}X$
a decoration of the graph	a solution of the system

Every graph corresponds to a system of set equations, and vice-versa. For example, corresponding to the picture in (9) we would take

$$
\begin{array}{llll}
X = \{x_0, y_0, z_0, x_1, y_1, z_1, x_2, y_2, z_2, x_3, y_3, z_3\} & & & \\
e(x_0) = \emptyset & e(x_1) = \{x_0\} & e(x_2) = \{x_1, y_3\} & e(x_3) = \{z_0, x_2\} \\
e(y_0) = \{x_0\} & e(y_1) = \{y_0\} & e(y_2) = \{y_1, z_3\} & e(y_3) = \{x_0, y_2\} \\
e(z_0) = \{x_0, y_0\} & e(z_1) = \{z_0\} & e(z_2) = \{z_1, x_3\} & e(z_3) = \{y_0, z_2\}
\end{array}
$$

So the way to go from the picture to the function is that each set $e(v)$ is the set of children of v. In terms of the kind of notation we have seen before, we prefer to write this system in a way that elides e:

$$
\begin{array}{llll}
x_0 \approx \emptyset & x_1 \approx \{x_0\} & x_2 \approx \{x_1, y_3\} & x_3 \approx \{z_0, x_2\} \\
y_0 \approx \{x_0\} & y_1 \approx \{y_0\} & y_2 \approx \{y_1, z_3\} & y_3 \approx \{x_0, y_2\} \\
z_0 \approx \{x_0, y_0\} & z_1 \approx \{z_0\} & z_2 \approx \{z_1, x_3\} & z_3 \approx \{y_0, z_2\}
\end{array}
$$

The study of non-wellfounded sets proposes to treat *every* graph as a picture of a unique set. In order to make this work, some kind of change is needed in set theory. The reason is that sets like $\Omega = \{\Omega\}$ do not exist in the most commonly-used set theory, *ZFC*. This is due to the Foundation Axiom (*FA*): we'll discuss this issue further in Section 4 below. For now, *FA* implies that the only graphs with decorations are those with no infinite sequence of points following the arrows. The change in set theory that we make is simply to replace this axiom *FA* with a different one called *AFA*. The content of *AFA* is that every graph has a unique decoration (alternatively, every system of set equations has a unique solution).

At the same time, there is a reduction of hypersets to ordinary sets. This means that one could regard all talk of hypersets as merely abbreviatory. This reduction is fairly complicated, and we shall present it in due course.

Adopting *AFA* not only helps with circularly defined sets, but it also helps with streams and trees. As we have mentioned, if one uses *FA*, there are no streams or trees according to our definitions. That is, N^∞ is literally the empty set with *FA*, as is Tr. But with *AFA* these sets are non-empty. Moreover, one can prove that under *AFA*, N^∞ and Tr have the properties that one would want them to have. (For example, one can prove that N^∞ corresponds to the function space ${}^N N$ in the way we have discussed.) Finally, working out the resulting theory gives tools that are useful in studying collections of circularly-defined objects such as streams and trees. The point is that this one axiom *AFA* gives us all of this, and more.

Terminology and history The Axiom *AFA* was first studied by Marco Forti and Furio Honsell in 1983. Their paper [14] studies a number of axioms which contradict the Foundation Axiom *FA*, continuing a much older line of work in set theory that deals with alternatives to *FA*. The one they call X_1 is equivalent to what has now come to be called *AFA*.

Peter Aczel's book [1] treats many axioms that contradict *FA*, but it pays most attention to *AFA*. It also proved many of the important results in the subject, including ones mentioned in these notes. Aczel's own entrance to the subject was an area of semantic modeling that he had been working on, concerning the calculus of communicating sysemts (CCS). He found it natural to propose a set theoretic semantics, and yet the most obvious modeling seemed to run into problems with Foundation. It is always a bold step to recommend changing the axioms of set theory in order to make an application of the subject. Usually it is a brash move. For the most part people resist the idea: when the proposal might well be cast in more standard forms (as can be done with work using *AFA*), people wonder why one wants to tamper with a standard theory; when it cannot be cast in a standard way, the reception is even worse.

Aczel's work became influential for two research areas. He visited Stanford in 1985, where Jon Barwise was director of the Center for the Study of Language and Information (and this author was a post-doc there). Barwise recognized the value of the work, partly because he had similar problems with Foundation in his own work on situation semantics, and partly because he saw in the work an appealing conception of set that was at odds with the iterative conception that had been received wisdom for him and practically everyone else raised in the mainstream tradition of mathematical logic.[5] He thought that *non-wellfounded sets* should be called by a name that reflected the change in conception, and he proposed calling them *hypersets* in parallel to the *hyperreal* numbers of non-standard analysis. This terminology has for the most part not stuck, but it is not completely outdated, either. In this article, we'll use both terms interchangeably.

Aczel's book was also immediately influential for people working on semantic questions in theoretical computer science. This was not so much because it raised questions about set theory, but rather because it showed the value of using the categorical notion of a *coalgebra*. The main use in the book is to organize certain concepts into an elegant subject. But it quickly became apparent that this notion of coalgebra could be studied on its own, that themes from the book had a field of application much wider than pure set theory.

These notes reflects the influence of all of these sources. To be sure, we shall see the main results on the set theory obtained using *AFA*. Also, we present enough of the theory that someone who needs to read papers that use it should be able to do start doing so. We also emphasize the conceptual underpinnings of the subject, and compare them to more standard foundational work. This is hardly ever done in technical papers on the subject, but should be of interest to people in several areas of philosophy. Finally, our work incorporates many ideas and results coming from the coalgebra research community in the years following the publication of Aczel [1].

[5] Both the applicability and the conceptual point are important: I am not aware of publications prior to Aczel's book that even hinted at applications. Also, the discussion of conceptual points is not developed at length there. Earlier papers such as Forti and Honsell [14] do not mention conceptual points at all, and indeed amid the multiplicity of axioms studied it is hard to see whether earlier workers were even thinking in these terms.

1.4 Universal Harsanyi type spaces

Type spaces are mathematical structures used in theoretical parts of economics and game theory. The are used to model settings where agents are described by their *types*, and these types give us “beliefs about the world”, “beliefs about each other’s beliefs about the world”, “beliefs about each other’s beliefs about each other’s beliefs about the world”, etc. That is, the formal concept of a type space is intended to capture in one structure an unfolding infinite hierarchy related to *interactive belief*.

In landmark papers published in 1967 and 1968, John C. Harsanyi showed how to convert a game with incomplete information into one with complete yet imperfect information. This matter is not relevant here, but three related points are noteworthy. First, the notion of types goes further than what we described above: an agent’s type involves (or induces) *beliefs about the types of the other agents*, their types involve beliefs about A’s type, etc. So the notion of a type is already circular. Second, despite this circularity, the informal concept of a *universal type space* (as a single type space in which all types may be found) is widespread in areas of non-cooperative game theory and economic theory. And finally, the *formalization* of type spaces was left an open area: Harsanyi did not really formalize type spaces in his original papers (he used the notion); this was left to later researchers.

Getting back to our very rough informal description above, what exactly are “beliefs”? And how can a structure contain types which give rise to beliefs about other types? What is the relation of this to the infinite hierarchy of beliefs about beliefs about $\cdots$ beliefs about the world? Can we characterize the space of all possible types?

Again, we are not concerned with conceptual matters concerning beliefs and games. Most of the important papers for our study are technical contributions dealing with the matter of *universal type spaces*. We are concerned with the conceptual matters related to our own topics, and with some of the technical matters connected with measure theory, probability, and the like.

Returning to type spaces, we recall that the usual modeling of belief in game theory is via probability. So we would expect that *type spaces should be probabilistic versions of Kripke models*. One should replace the functor $\mathcal{P}$ with something like Δ, where

$$\Delta(W) \quad = \quad \{\mu : \mu \text{ is a probability measure on } W\}. \tag{10}$$

Indeed, this is the case: most proposals in the literature do end up studying certain mappings from a space X to some variation of the functor Δ applied to X. This is our third clue of the connection. But note that (10) leaves a lot lacking: if W is just a set, how do we know that it has any probability measures? Does it matter which σ-algebra we use?

Let S be a fixed measurable space. A *type space over* S is a tuple (M, σ, N, τ), where M and N are measurable spaces, and

$$\begin{aligned} \sigma &: M \to \Delta(S \times N) \\ \tau &: N \to \Delta(S \times M) \end{aligned}$$

What we are describing would be better called a *two-player type space over* S, and the spaces M and N are the spaces of the two players. The idea is that S represents the possible “states

A *measurable space* is a pair $M = (M, \Sigma)$, where M is a set and Σ is a σ-algebra of subsets of M. The sets in Σ are called *measurable sets* or *events*. A *measure on M* is a function $\mu : \Sigma \to [0, \infty]$ which has the property that if S_0, S_1, ... is a countable collection of pairwise disjoint sets, then $\mu(\bigcup_n S_n) = \sum_n \mu(S_n)$. The measure μ is a *probability measure* if $\mu(M) = 1$.
For any space M, let be $\Delta(M)$ be the set of probability measures on M. For any measurable set E, we define

$$\beta^p(E) \quad = \quad \{\mu \in \Delta(M) : \mu(E) \geq p\}.$$

We want to specify a σ-algebra Σ^* on the set $\Delta(M)$. We take Σ^* to be the smallest σ-algebra containing all sets of the form $\beta^p(E)$ for $p \in [0, 1]$ and $E \in \Sigma$. So $(\Delta(M), \Sigma^*)$ is a measurable space.
Given two measurable spaces A and B, the product space $A \times B$ is the cartesian product of the sets A and B, endowed with the σ algebra generated by the sets of the form $E \times F$, where E is measurable in A and F is measurable in B. For a subset $E \subseteq A \times B$, the *sections* of E are the sets: $E_a = \{b : (a, b) \in E\}$ and $E^b = \{a : (a, b) \in E\}$. Each section of a measurable subset of the product is measurable.
If μ is a probability measure on A and ν a probability measure on B, we can define the probability measure $\mu \times \nu$ on $A \times B$ by $(\mu \times \nu)(E) = \int \mu(E^b) d\nu = \int \nu(E_a) d\mu$.
Going in the other direction, a probability measure μ on $A \times B$ induces via the projections, a measure on each of the factor spaces. These measures are called *marginals*, and defined and denoted by $mar_A\mu = \mu \circ \pi_A^{-1}$; $mar_B\mu = \mu \circ \pi_B^{-1}$.

Figure 1: Supplement on basic definitions concerning measureable spaces

of nature", each $m \in M$ is a possible "type" of the first player, and each $n \in N$ is a possible type for the second. Each player does not know the exact state of nature or which type of player he or she faces. But each does have some probabilistic "opinion" on this matter: each $\sigma(m)$ gives a measure on $S \times N$, and so the first player can measure subsets of both S and N (using marginals).

One assumption incorporated into our definition is that the players "know" their own types. This is why $\sigma(m)$ is a measure on $S \times N$ and not $S \times M \times N$. However, each $\mu \in \Delta(S \times N)$ and each $m \in M$ together define a measure $\hat{\mu}_m$ on the larger product $S \times M \times N$ by "concentrating μ on m:

$$\hat{\mu}_m(E) \quad = \quad \mu(m)(\{(s,n) : (s,m,n) \in E\}).$$

Everything we said for the first player goes through for the second, mutatis mutandis, and so we also have measures $\hat{\rho}_n \in \Delta(S \times M \times N)$ for each $\rho \in \Delta(S \times M)$ and each $n \in N$.

The basic modal language for a type space over S would have the following sentences:

- ⋆ An atomic sentence A for each measurable subset A of S.
- ⋆ If φ is a sentence and $p \in [0,1]$, then both $B^1_p\varphi$ and $B^2_p\varphi$ are sentences.
- ⋆ If φ and ψ are sentences, then so is $\varphi \wedge \psi$.

We read $B^1_p\varphi$ as saying that player 1 believes the probability of φ to be at least p.

We define a semantics for this language, interpreting each sentence by subset $[\![\varphi]\!] \subseteq S \times M \times N$ as follows:

$$\begin{array}{lcl}
[\![A]\!] & = & A \times M \times N \\
[\![B^1_p\varphi]\!] & = & \{(s,m,n) : \widehat{\sigma(m)}_m([\![\varphi]\!]) \geq p\} \\
[\![B^2_p\varphi]\!] & = & \{(s,m,n) : \widehat{\tau(n)}_n([\![\varphi]\!]) \geq p\} \\
[\![\varphi \wedge \psi]\!] & = & [\![\varphi]\!] \cap [\![\psi]\!]
\end{array}$$

One may also add negation or implication to the language, with the semantics from classical logic. A good first exercise would then be to show that $[\![B^1_p\varphi \to B^1_1 B^1_p\varphi]\!] = S \times M \times N$. This is an echo of the assumption agents know their types: they are able to introspect on their own beliefs, and to do so with certainty.

Reformulation of the problem of universal type spaces

1.5 Self-similar sets

Discussions of circularity rarely touch on *self-similar* sets in the usual sense, mostly because the mathematics involved seems so different. We want to raise the topic here because in recent years it has become clear that there really are commonalities.

First, here is a picture of a set in the plane which exhibits self-similarity. The idea is that the picture repeats itself at a smaller scale in various points. Clearly there is nothing problematic about this. It doesn't manifest object circularity because the picture is not a "part" of itself in the usual sense. On the other hand, there is a sense in which this does exhibit collection similarity. To clarify this, it makes sense to consider a simpler example, the *Cantor (middle-third) set*.

The Cantor set is a certain subset of the set of real numbers between 0 and 1. It has several equivalent definitions/characterizations:

1. Take the unit interval $[0, 1]$, then remove the open middle third $(\frac{1}{3}, \frac{2}{3})$, leaving two disconnected pieces. For each of those, remove the open middle third. Keep going for infinitely many steps. Then c is what remains "at the end".

2. c is the set of numbers possessing a ternary (base 3) decimal expansion with no 1's.

3. c is the unique non-empty compact subset of the unit interval $[0, 1]$ such that

$$c = \frac{1}{3}c \;\cup\; \left(\frac{2}{3} + \frac{1}{3}c\right), \tag{11}$$

   where $\frac{1}{3}c$ denotes the set $\{\,\frac{1}{3}x \mid x \in c\,\}$, and the second set is interpreted similarly, by also adding $\frac{2}{3}$ to each point.

The first characterization involves recursion in a certain way, though it is different from recursion on numbers.

The second point gives another interesting characterization, one which is not relevant to our discussion. We mention it only to underscore the point that important objects frequently have more than one description.

The last point is one that really does appear as collection circularity: it defines a set in terms of itself. It is possible to obtain a similar definition for the fractal picture at the beginning of this section. This kind of definition does not always define a set uniquely.

It will turn out in Section 6 that there are commonalities in the mathematics behind the diverse set of examples from this section. From streams to sets to fractals, there are similar principles at work.

2 Coalgebra

The purpose of this section is to compare *FA* and *AFA* in a technical way, using ideas from category theory. That is, the language of category theory and especially its built-in feature of *duality* are used to say something insightful about the relation between *FA* and *AFA*. Further, the dual statements about the axioms suggest a much more systematic and thoroughgoing duality about a host of other concepts. This deeper point is not a strictly mathematical result but rather more of a research program, and so the final subsection here will detail some of what is known about it.

As we said, our work here begins to use category theory. We realize that not all readers will be familiar with that subject at all. So we shall try to make this section as accessible as possible. In particular, we'll only present those notions from category theory that we actually need in our work of this section. We also illustrate all of the definitions on a few categories which will be of interest. And as we go on in future sections, we'll develop only the background that we need.[6]

Our use of category theory is mainly for the terminology and intuition. We know that there are philosophical issues connected with the use of category theory as a foundation for mathematics. These notes do not deal with any of these issues in a head-on way.

Initial and final objects We need a definition from category theory. Fix a category C. An object x is *initial* if for every object y there is exactly one morphism $f : x \to y$. Dually, an object x is *final* if for every object y there is exactly one morphism $f : y \to x$.

In Set, the empty set is an initial object; for every set y, the empty function is the only function from $\emptyset$ to y. In addition, the empty set is the only initial object.

As for final objects, every singleton set $\{x\}$ is a final object. For every set y, the constant function with value x is the only function from y to x. And the singletons are the only final objects in the category.

Proposition 2.1 *Let C be a category, and a and b be initial objects. Then a and b are* isomorphic *objects: there are $f : a \to b$ and $g : b \to a$ such that $g \circ f = id_a$ and $f \circ g = id_b$.*

Proof By initiality, we get (unique) morphisms f and g as in our statement. Note that $g \circ f$ is a morphism from a to itself. And since a is initial and id_a is also such a morphism, we see that $g \circ f = id_a$. Similarly for b. ⊣

We refer the reader to the any book on category theory for the definitions of category and functor.

We need to mention the objects and morphisms in the categories of sets and of classes, and also to spell out the functors of interest on them.

2.1 The category Set

We start with a description of the category Set of sets. We work in the standard set theory.

The objects of Set are the sets, and the morphisms are triples $\langle x, y, f\rangle$ where $f : x \to y$. That is, each triple $\langle x, y, f\rangle$ is a morphism from x to y. The identity morphism id_a for a set a is $\langle a, a, f\rangle$, where f is the identity function on a and the composition operation of morphisms is given by:

$$\langle y, z, g\rangle \circ \langle x, y, f\rangle = \langle x, z, g \circ f\rangle$$

[6] Since we have cut all the corners, the treatment here would be inadequate as an introduction to category theory.

[7] The rest of the category structure consists of the identity morphisms and the composition operation. These are all defined as expected, and the verification of the category properties is routine.

We continue with a discussion of some set-theoretic constructions related to category-theoretic notions.

Initial and final objects We need a definition from category theory. Fix a category C. An object x is *initial* if for every object y there is exactly one morphism $f : x \to y$. Dually, an object x is *final* if for every object y there is exactly one morphism $f : y \to x$.

In Set, the empty set is an initial object; for every set y, the empty function is the only function from $\emptyset$ to y. In addition, the empty set is the only initial object.

As for final objects, every singleton set $\{x\}$ is a final object. For every set y, the constant function with value x is the only function from y to x. And the singletons are the only final objects in the category.

Proposition 2.2 *Let C be a category, and a and b be initial objects. Then a and b are* isomorphic *objects: there are $f : a \to b$ and $g : b \to a$ such that $g \circ f = id_a$ and $f \circ g = id_b$.*

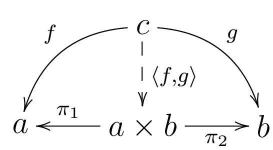
Proof By initiality, we get (unique) morphisms f and g as in our statement. Note that $g \circ f$ is a morphism from a to itself. And since a is initial and id_a is also such a morphism, we see that $g \circ f = id_a$. Similarly for b. ⊣

Products In a category C, a *product of objects a and b* is a triple $(a \times b, \pi_1, \pi_2)$ such that

a. $\pi_1 : a \times b \to a$.

b. $\pi_1 : a \times b \to a$.

c. If (c, f, g) is any triple with properties (a) and (b), then there is a unique $\langle f, g\rangle : c \to a \times b$ such that both $f = \pi_1 \circ \langle f, g\rangle$ and $g = \pi_2 \circ \langle f, g\rangle$.



[7] The reason why we do take the morphisms to simply be functions construed as sets of pairs is that doing so would not uniquely specify the codomains. For example, consider the set

$$\{\langle 0, 0\rangle, \langle 1, 0\rangle, \ldots, \langle n, 0\rangle, \ldots\}$$

We might in some settings wish to take this to be a morphism with codomain $\{0\}$, and in other settings to be a morphism with codomain N. Since a morphism can have only one codomain, we must include extra copies. Note that in contrast, the domains of functions may be read off from the set-theoretic versions. An alternative presentation of the category of sets would then take as morphism the pairs $\langle y, f\rangle$ with $y \supseteq \{a : (\exists b)\langle b, a\rangle \in f\}$.

Coproducts Let C be an arbitrary category. A *coproduct* of objects x and y, is an object $x+y$ with morphisms $\mathsf{inl} : x \to x+y$ and $\mathsf{inr} : y \to x+y$ meeting the following condition: if $f : x \to z$ and $g : y \to z$, then there is a unique $[f,g] : x+y \to z$ such that $[f,g] \circ \mathsf{inl} = f$ and $[f,g] \circ \mathsf{inr} = g$. Technically, the coproduct of x and y is the triple $(x+y, \mathsf{inl}, \mathsf{inr})$.

Usually there's no need to add the names of the objects to the coproduct maps inl and inr. But if we would need to, we could write

$$\mathsf{inl}_{a,a+b} : a \to a+b,$$

and similarly for $\mathsf{inr}_{b,a+b}$.

C has coproducts if every two objects have a coproduct.

A *category with coproducts* is a tuple $(C,+)$, where C is a category and $+$ is a coproduct operation on C, giving for each x and y the triple $(x+y, \mathsf{inl}_{x,x+y}, \mathsf{inr}_{y,x+y})$.

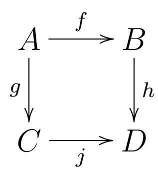
Set has coproducts in the following way: let

$$x+y \quad = \quad (x \times \{0\}) \cup (y \times \{1\}),$$

let $\mathsf{inl}(a) = (a,0)$ for $a \in x$, and let $\mathsf{inr}(b) = (b,1)$ for $b \in y$.

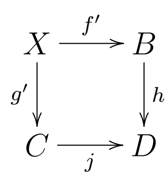
More generally, every set $\mathcal{X}$ of sets has a coproduct $\coprod \mathcal{X}$. There are injections $\mathsf{in}_x : x \to \mathcal{X}$ for each $x \in X$, and these have the expected universal property.

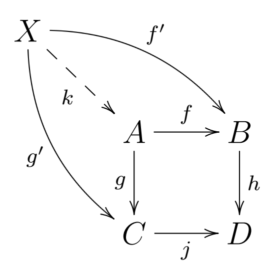
Pullbacks and weak pullbacks A *pullback square* in a category C is four objects and morphisms of C that look like



(12)

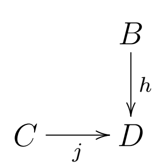
with the properties that (a) $h \circ f = j \circ g$, and (b) if $f' : X \to B$ and $g' : X \to C$ are any morphisms such that the diagram on the left





commutes, then there is a unique $k : X \to A$ such that the diagram on the right commutes. That is, $f' = f \circ k$ and $g' = g \circ k$.

A category C *has pullbacks* if every diagram in the category that looks like



(13)

can be completed to a pullback square. That is, for all B, C, h, and j as in (13) there are A, f and g which make a pullback square.

The category of sets has pullbacks. Specifically, given B, C, h, and j as in (13), let

$$A = \{(b, c) \in B \times C : h(b) = j(c)\}.$$

Let f and g be the evident projections. It is routine to check that we have a pullback square. Using this and the evident uniqueness of pullbacks, it follows that up to isomorphism, in every pullback square in Set we may assume that $A \subseteq B \times C$, $f = \pi_1$, and $g = \pi_2$.

In fact, it is good to know a few other pullbacks. First, if $f : A \to B$, then its kernel K is $\{(x, y) \in A \times A : f(x) = g(x)\}$. Together with its projections, we have a pullback square

$$\begin{array}{ccc} K & \xrightarrow{\pi_1} & A \\ {\scriptstyle \pi_1}\downarrow & & \downarrow{\scriptstyle f} \\ A & \xrightarrow[f]{} & B \end{array}$$

Second, if $A \subseteq C$ and $B \subseteq C$, then we have a pullback

$$\begin{array}{ccc} A \cap B & \xrightarrow{i(A \cap B, A)} & A \\ {\scriptstyle i(A \cap B, B)}\downarrow & & \downarrow{\scriptstyle i(A, C)} \\ B & \xrightarrow[i(B, C)]{} & C \end{array}$$

All of the morpisms are inclusions.

Third, if we have two relations $R \subseteq A \times R$ and $S \subseteq B \times C$, The pullback is known as

$$R \bowtie S \quad = \quad \{((a, b_1), (b_2, c)) \in R \times S : b_1 = b_2\}.$$

We then have a pullback square:

$$\begin{array}{ccc} R \bowtie S & \xrightarrow{\pi_1^{R \bowtie S}} & R \\ {\scriptstyle \pi_2^{R \bowtie S}}\downarrow & & \downarrow{\scriptstyle \pi_2^R} \\ S & \xrightarrow[\pi_1^S]{} & B \end{array} \tag{14}$$

Now $R \bowtie S$ itself is not quite the relational composition

$$R \circ S \quad = \quad \{(a, c) : (\exists b \in B) a \ R \ b \ S \ c\},$$

but it is close. To get $R \circ S$, note that

$$\begin{array}{l} \pi_1^R \circ \pi_1^X : X \to A \\ \pi_2^S \circ \pi_2^{R \bowtie S} : R \bowtie S \to C \end{array}$$

Then $R \circ S$ is the *joint image*

$$\{((\pi_1^R \circ \pi_1^{R \bowtie S})x, (\pi_2^S \circ \pi_2^{R \bowtie S})x) : x \in R \bowtie S\}.$$

Definition A *weak pullback* is again a square with the shape of (12) which commutes, and which has the weaker property that for every $f' : X \to B$ and $g' : X \to C$ such that $h \circ f' = j \circ g'$, there is a not-necessarily-unique $k : X \to A$ such that $f' = f \circ k$ and $g' = g \circ k$.

Here is a general picture of what weak pullbacks are like in Set. They are like pullbacks, but allowing many copies of the same elements. Suppose we have a commuting square as in (12). Let S be any set, and let $A \subseteq B \times C \times S$ be any set such that (i) if $fa = gb$, then for some $s \in S$, $(a, b, s) \in A$; and (ii) if $(a, b, s) \in A$, then $fa = gb$. We have projections $\pi_1 : S \to A$ and $\pi_2 : S \to B$. And (S, π_1, π_2) is a weak pullback of f and g, and all weak pullbacks are of this form.

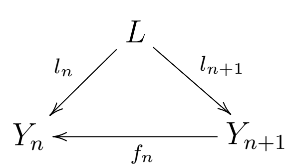
Limits of ω-chains Let

$$Y_0 \xleftarrow{f_1} Y_1 \xleftarrow{f_2} Y_2 \qquad \cdots \qquad Y_n \xleftarrow{f_n} Y_{n+1} \qquad \cdots \tag{15}$$

be any chain of sets and functions. Let

$$\begin{array}{rcl} L & = & \{g : g \text{ is a function with domain } \omega, \\ & & \text{for all } n,\ g(n) \in Y_n, \text{ and } g(n) = f_n(g(n+1))\} \end{array}$$

Then there are natural maps $l_n : L \to Y_n$ given by $l_n(g) = g(n)$. The set L together with the maps $l_n : L \to Y_n$ is the *limit cone* of the chain in (15). This means two things. First, for all n the diagram below commutes:



(We say that $(L, l_n : L \to Y_n)_{n \in \omega}$ is a *cone* over the chain (**??**).) And second, if $(Z, w_n : Z \to Y_n)_{n \in \omega}$ is any cone over (**??**) – that is, if $w_n = f_n \circ w_{n+1}$ for all n – then there is a unique $m : Z \to L$ such that for all n, $l_n \circ m = w_n$. This m is called the *mediating morphism* from the cone over Z to the cone over L.

To prove this second fact, let $m(z)$ be the function with domain ω defined by $m(z)(n) = w_n(z)$. To check that $m(z) \in L$, note that

$$m(z)(n) \quad = \quad w_n(z) \quad = \quad f_n(w_{n+1}(z)) \quad = \quad f_n(m(z)(n+1)).$$

So at this point we have $m : Z \to L$. Also, for all $z \in Z$ we have

$$(l_n \circ m)(z) \quad = \quad l_n(m(z)) \quad = \quad m(z)(n) \quad = \quad w_n(z).$$

This shows that $l_n \circ m = w_n$.

To wrap up this part, we show that m is unique. For suppose that $m' : Z \to L$ is such that for all n, $l_n \circ m' = w_n$. To show that $m = m'$, we check that for all $z \in Z$, $m(z) = m'(z)$. And since these are functions, it is enough to show that for all n, $m(z)(n) = m'(z)(n)$. But

$$m(z)(n) \quad = \quad l_n(m(z)) \quad = \quad w_n(z) \quad = \quad l_n(m'(z)) \quad = \quad m'(z)(n).$$

2.2 Functors on Set

We collect in this section some general facts about functors on sets, starting with examples of them.

As in any category, we have the identity functor 1 and the constant functors K_x for any set x. We also have a composition operation on functors, and we write the composition of F and G as either FG or $F \circ G$.

Examples of functors on Set We present here a few examples of functors that will appear throughout the course. Most of the verifications of the functorial properties are easy, and so we omit them.

We also have the *power set functor* $\mathcal{P}$ taking each set X to is power set

$$\mathcal{P}X \quad = \quad \{a : a \subseteq X\}.$$

If $f : X \to Y$, then $\mathcal{P}f : \mathcal{P}X \to \mathcal{P}Y$ is defined for $a \in \mathcal{P}X$ by taking setwise images. So

$$(\mathcal{P}f)(a) \quad = \quad f[a] \quad (= \{f(x) : x \in a\}).$$

For any set A, we have a functor $FX = X^A$, where X^A is the set of functions from A to X. If $f : X \to Y$, then $Ff : X^A \to Y^A$ is $g \mapsto f \circ g$.

We also mention a few functors which are different from the ones above. They lack many of the nice properties of the functors which we've seen, and so they will appear from time to time as sources of counterexamples.

The functor $C_{0,1}$ is defined by $C_{0,1}(0) = 0$, and $C_{0,1}x = 1$ for all other x. On morphisms, there are no choices, since 0 is initial and 1 is final.

Lemma 2.3 *Every functor F on* Set *preserves all surjective maps.*

Proof Suppose $g : X \to Y$ is surjective. Let $h : Y \to X$ be such that $g \circ h = id_Y$. Then $Fg \circ Fh = id_{FY}$, and so Fg must be surjective. $\dashv$

Proposition 2.4 *If $FX = \emptyset$ for some $X \neq \emptyset$, then F is the constant functor $\emptyset$.*

Proof Let Y be any set. Then there is $f : Y \to X$; f could be a constant, for example. And now $Ff : FY \to \emptyset$. So FY must be empty also, since there are no maps from a non-empty set to $\emptyset$. $\dashv$

2.3 Natural transformations

Let $F : C \to D$ be a functor between two categories, and let $G : C \to D$ also be a functor between the same two. Then a *natural transformation from F to G* is a family η of morphisms of D indexed by objects of C, in particular each η_x is a morphism in D from Fx to Gx. The requirement on η is that for each morphism in C of the form $f : x \to y$, the square below commutes:

$$\begin{array}{c} x \\ \downarrow{\scriptstyle f} \\ y \end{array} \qquad \begin{array}{ccc} Fx & \xrightarrow{\eta_x} & Gx \\ {\scriptstyle Ff}\downarrow & & \downarrow{\scriptstyle Gf} \\ Fy & \xrightarrow[\eta_y]{} & Gy \end{array} \tag{16}$$

In symbols, $Gf \circ \eta_x = \eta_y \circ Ff$.

One writes $\eta : F \to G$ to say that η is a natural transformation from F to G. Also, for each object x of the domain category, η_x is sometimes called the *component of η at x*.

Constructions on natural transformations There are several important ways to get new natural transformations from old ones. Suppose first that $\eta : F \to G$, and let $H : B \to C$ be another functor. Then $F \circ H : B \to D$ and $G \circ H : B \to D$. We get a natural transformation called ηH from $F \circ H$ to $G \circ H$ by

$$\eta_H b \quad = \quad \eta_{Hb}.$$

To check that this is indeed natural, let $f : x \to y$ be a morphism in B. Then for each x in B $(\eta H)_x : (F \circ H)x \to G \circ H)x$. And we have the diagram

$$\begin{array}{ccc} (F \circ H)x & \xrightarrow{(\eta H)_x} & (G \circ H)x \\ {\scriptstyle (F\circ H)f}\downarrow & & \downarrow{\scriptstyle (G\circ H)f} \\ (F \circ H)y & \xrightarrow[(\eta H)_y]{} & (G \circ H)y \end{array} \tag{17}$$

This is literally the same as

$$\begin{array}{ccc} F(Hx) & \xrightarrow{\eta_{Hx}} & G(Hx) \\ {\scriptstyle F(Hf)}\downarrow & & \downarrow{\scriptstyle G(Hf)} \\ F(Hy) & \xrightarrow[\eta_{Hy}]{} & G(Hy) \end{array}$$

This last diagram is just an instance of naturality of η. So it commutes. And thus the diagram in (17) commutes, verifying that indeed η_H is natural.

For our second construction, suppose again that $\eta : F \to G$, and this time let $H : D \to E$. So now $H \circ F$ and $H \circ G$ are functors from C to E. We get a natural transformation from $H \circ F$ to $H \circ G$, this time called $H\eta$, by

$$(H\eta)_x \quad = \quad H\eta_x.$$

That is, we apply the functor H to the morphism η_x. The verification of naturality is a little different: we appy H throughout (16).

If $\eta : F \to G$ and $\mu : F \to H$, then we get a natural transformation $\mu \circ \eta : F \to H$ by $(\mu \circ \eta)_x = \mu_x \circ \eta_x$. The verification of naturality is easy.

Finally, suppose that $F, G : D \to E$ and $H, K : C \to D$, and let $\eta : F \to G$ and $\mu : H \to K$. We get a natural transformation $\mu * \eta : F \circ H \to K \circ G$ by

$$\begin{array}{ccc} F \circ H & \xrightarrow{F\mu} & F \circ K \\ {\scriptstyle \eta H}\downarrow & \searrow^{\mu * \eta} & \downarrow{\scriptstyle \eta K} \\ G \circ H & \xrightarrow[G\mu]{} & G \circ K \end{array}$$

That is, we claim that the outside of the figure commutes, and then we define $\mu * \eta$ to be the composite in either direction; this will be a natural transformation by the three constructions which we have already seen. But for each object x of C the square above is a naturality square for η, applied to the morphsim $\mu_x : Hx \to Kx$.

Functor categories If C and D are categories, we define the *endofunctor category* $[C \to D]$ as follows: its objects are the endofunctors $F : C \to D$. The morphisms from F to G are the natural transformations. The identity on F is the natural transformation $1_F : F \to F$ given by $(1_F)_x = id_{Fx}$. The composition operation is the one we saw above. The verifications of the category properties are again easy.

2.4 Examples in Set

There are many important examples of natural transformations in our subject. Here is a first one. Let A and B be fixed sets. Recall that we have a functor $(\ )^A : \mathsf{Set} \to \mathsf{Set}$ given on sets by: X^A is all functions from A to X, and for $f : X \to Y$, $f^A : X^A \to Y^A$ is $k \mapsto f \circ k$.

Now let $FX = (X^A)^B$, and let $GX = X^{A \times B}$. We define $\eta : F \to G$ by

$$\eta_X(p)(a, b) \quad = \quad p(b)(a).$$

Here and below, $p \in (X^A)^B$, so $p(b) \in X^A$ and $p(b)(a) \in X$. To verify the naturality, let $f : X \to Y$. Then for all $(a, b) \in A \times B$,

$$\begin{array}{rcl} f^{A \times B}(\eta_X(p))(a, b) & = & f(\eta_X(p)(a, b)) \\ & = & f(p(b)(a)) \\ & = & (f^A(p(b)))(a) \\ & = & (f^A \circ p)(b))(a) \\ & = & (((f^A)^B(p))(b))(a) \\ & = & \eta_Y((f^A)^B)(p))(a, b) \end{array}$$

This for all p, a, and b shows the naturality of η.

Continuing with examples, here are some natural transformations related to the power set endofunctor $\mathcal{P}$. First, let 1 be the identity endofunctor on Set, and let $\eta : 1 \to \mathcal{P}$ be $\eta_A(x) = \{x\}$. That is, for all sets A, η_A takes each element of A to its singleton subset of A. To check that this is natural, let $f : A \to B$. Then $\mathcal{P}f \circ \eta_A : A \to \mathcal{P}B$ is $x \mapsto \{x\} \mapsto \{fx\}$. And $\eta_B f : A \to \mathcal{P}B$ is $x \mapsto fx \mapsto \{fx\}$.

Second, let $\bigcup : \mathcal{P}\mathcal{P} \to \mathcal{P}$ be the *union* natural transformation. For each set A and each $\mathcal{F} \in \mathcal{P}\mathcal{P}A$ (that is, for each family $\mathcal{F}$ of subsets of A),

$$\bigcup \mathcal{F} \quad = \quad \{a \in A : (\exists S \in \mathcal{F})(a \in S)\}.$$

(We should write $\bigcup_A \mathcal{F}$, but the subscript notation for unions conflicts with the subscript notation for natural transformations.) For the naturality here, let $f : A \to B$. Then for each $\mathcal{F} \in \mathcal{P}\mathcal{P}A$,

$$\begin{array}{rcll} \mathcal{P}f(\bigcup \mathcal{F}) & = & \mathcal{P}f(\{a \in A : (\exists S \in \mathcal{F})(a \in S)\}) & \\ & = & \{f(a) : (\exists a \in A)(\exists S \in \mathcal{F})(a \in S)\} & \\ & = & \{b \in B : (\exists T \in \mathcal{P}\mathcal{P}f(\mathcal{F}))(b \in T)\} & \text{(see below)} \\ & = & \bigcup \mathcal{P}\mathcal{P}f(\mathcal{F}) & \end{array}$$

For the marked line, note first that if $S \in \mathcal{F}$ is such that $a \in S$, then $fa \in \mathcal{P}\mathcal{P}f(S) = \{f(x) : x \in S\}$, and this set belongs to $\mathcal{P}\mathcal{P}f(\mathcal{F})$. In the other direction, if $T \in \mathcal{P}\mathcal{P}f(\mathcal{F})$, then T is of the form $\mathcal{P}f(S)$for some $S \in \mathcal{F}$. So if $b \in T$, then $b = f(a)$ for some S such that $S \in \mathcal{F}$.

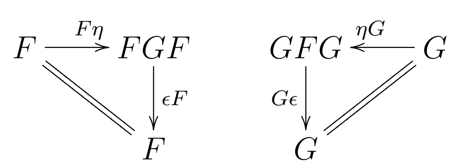
2.5 Adjoints and monads: the definitions

At this point, we wish to exhibit two category-theoretic definitions that come in terms of natural transformations. Although these notions will be important later, it is not important that you work with theory at this point. It is only important that you understand the notation involved.

Definition Let C and D be categories. An *adjunction* from C to D is a quadruple (F, G, η, ϵ) such that $F : C \to D$, $G : D \to C$, $\eta : 1 \to GF$, $\epsilon : FG \to 1$, and such that the two equations below are satisfied

$$\epsilon F \circ F\eta = 1_F \text{ and } G\epsilon \circ \eta G = 1_G.$$

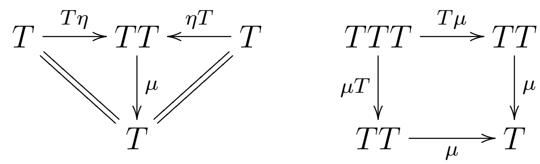
In symbols, the diagrams below commute:



Definition A *monad* on a category C is a triple (T, η, μ) such that $T : C \to C$ is a functor and $\eta : 1 \to T$ and $\mu : T^2 \to T$ are natural transformations, and the following laws hold:

identity $\mu \circ T\eta = 1_T = \mu \circ \eta T$.

associativity $\mu \circ \mu T = \mu \circ T\mu$.



Example 2.1 *We get monad on sets by* $(\mathcal{P}, \eta, \bigcup)$, *where (as above)* $\eta : 1 \to \mathcal{P}$ *is the singleton operation, and* $\bigcup$ *is the union.*

2.6 Standard functors

Proposition 2.5 *Concerning inclusions:*

1. $i_{b,c} \circ i_{a,b} = i_{a,c}$.
2. *Every inclusion is a one-to-one function, hence a monic morphism.*
3. *If* $f : a \to b$, $g : a \to c$, *and* $i \circ f = g$, *then* f *and* g *have the same images; i.e.,* $f[a] = g[a]$.
4. *If* $f : a \to b$ *has the property that* $i_{b,c} \circ f = i_{a,c}$, *then* $f = i_{a,b}$.
5. i_V, $i_{\mathcal{P}V}$, $\mathcal{P}i_V$, *and* $\mathcal{P}i_{\mathcal{P}V}$ *are all the same morphism, the identity on the universe.*
6. *if* a *is transitive, then* $a \subseteq \mathcal{P}a$. *Further* $i_{a,\mathcal{P}a} : a \to \mathcal{P}a$ *is a* $\mathcal{P}$*-coalgebra, and* $i_a : a \to V$ *is a* $\mathcal{P}$*-coalgebra morphism from it to* $(V, i_{\mathcal{P}V}^{-1})$.

Proof We only check part (4). For all $x \in a$, $f(x) = (i_{b,c} \circ f)x = i_{a,c}(x) = x$. ⊣

Definition For all sets a and b, if $a \subseteq b$ we have an *inclusion map* $i_{a,b} : a \to b$ given by $i(x) = x$ for all $x \in a$. When a and b are clear from context, we sometimes write this map as $i : a \hookrightarrow b$.

A functor on sets F is *monotone* if for all sets a and b, if $a \subseteq b$, then $Fa \subseteq Fb$. F is *standard* if (F is monotone and) for all $a \subseteq b$, $Fi_{a,b} = i_{Fa,Fb}$.

We only use the notation $i_{a,b}$ when $a \subseteq b$. So the condition that $Fi_{a,b} = i_{Fa,Fb}$ implies monotonicity; this is why our definition of standardness is phrased as it is.

Observe that the monotone functors are closed under composition, and so are the standard functors.

Proposition 2.6 *Let F be standard.*

1. *Let $f : a \to b$, let $a_0 \subseteq a$, and suppose that $b_0 \subseteq b$ is such that $f[a_0] \subseteq b_0$. Let $g : a_0 \to b_0$ be obtained by restricting f. Then for all $x \in Fa_0$, $(Ff)x = (Fg)x$.*

2. *Suppose that $f : a \to b$, and let $b_0 = f[a]$. Then $(Ff)[Fa] \subseteq Fb_0$.*

Proof For (1), let $i : a_0 \hookrightarrow a$ and $j : b_0 \hookrightarrow b$ so that $j \circ g = f \circ i$. By functoriality and standardness, $Fj \circ Fg = Ff \circ Fi$. But Fj and Fi are inclusions. This implies our result.

For (2), let g be as in (1), with $a = a_0$. By the result of part (a), $(Ff)[Fa] = (Fg)[Fa]$. But as $Fg : Fa \to Fb_0$, so $(Fg)[Fa] \subseteq Fb_0$. ⊣

Theorem 2.7 (V. Trnková 1969) *Every standard functor F on* Set *preserves intersections of non-empty sets. If $A \cap B \neq \emptyset$, then $F(A \cap B) = FA \cap FB$.*

Proof By monotonicity, $F(A \cap B) \subseteq F(A) \cap F(B)$.

Consider the diagram below:

$$\begin{array}{ccc} A \cap B & \xrightarrow{\ i\ } & A \\ {\scriptstyle j}\downarrow & & \downarrow{\scriptstyle k} \\ B & \xrightarrow[\ l\ ]{} & A \cup B \end{array}$$

All of the morphisms are inclusions. By the assumption that $A \cap B \neq \emptyset$, let $x \in A \cap B$. Now define $r : A \to A \cap B$ and $s : A \cup B \to B$ by

$$r(a) = \begin{cases} a & \text{if } a \in A \cap B \\ x & \text{if } a \notin A \cap B \end{cases} \qquad s(c) = \begin{cases} c & \text{if } c \in B \\ x & \text{if } a \in A \setminus B \end{cases}$$

Note that $j \circ r = s \circ k$, and also that $s \circ l = id_B$. Let $a : FA \cap FB \to FA$ and $b : FB \to FA \cap FB$ be the inclusions. Then $Fk \circ a = Fl \circ b$, and

$$\begin{array}{rcl} Fl \circ Fj \circ Fr \circ a & = & Fl \circ Fs \circ Fk \circ a \\ & = & Fl \circ Fs \circ Fl \circ b \\ & = & Fl \circ b \end{array}$$

Thus $Fr \circ a$ is an inclusion, since when we follow it by the inclusion of $F(A \cap B)$ in $F(A \cup B)$, we get the inclusion of $F(A) \cap F(a)$ in $F(A \cap B)$. Because of this, we see that $F(A) \cap F(B) \subseteq F(A \cap B)$. ⊣

Set The objects are the sets, and the morphisms are triples $\langle x, y, f \rangle$ where $f : x \to y$. That is, each triple $\langle x, y, f \rangle$ is a morphism from x to y. The identity morphism id_a for a set a is $\langle a, a, f \rangle$, where f is the identity function on a and the composition operation of morphisms is given by:

$$\langle y, z, g \rangle \circ \langle x, y, f \rangle = \langle x, z, g \circ f \rangle$$

Functors on Set The polynomial operators on sets extend to endofunctors on Set. The way that these operations are defined on morphisms is straightforward and may be found in any book on category theory. Here is a brief summary: For any set s, the constant functor with value s is a functor on Set. It takes every function to id_s. For any two functors F and G, we have a functor $F \times G$ defined by $(F \times G)(a) = Fa \times Ga$; here we use the cartesian product on sets. If $f : a \to b$, then

$$(F \times G)f(x, y) \quad = \quad (Ff(x), Gf(y)).$$

We also have a functor $F+G$ defined by $(F+G)(a) = Fa+Ga$ using the coproduct on sets, that is, the disjoint union. Here the action on morphisms is by cases $(F+G)f(\mathsf{inl}x) = Ff(x)$, and $(F+G)f(\mathsf{inr}x) = Gf(x)$. A special case is $Fx = x+1$. That is, Fx is the disjoint union of x with a singleton. And if $f : x \to y$, then $Ff : Fx \to Fy$ works in much the same way, taking the new point in x to the new point in y, and otherwise behaving like f.

The power polynomial operators also extend to endofunctors on Set: on morphisms $f : x \to y$, the function $Pf : Px \to Py$ takes each subset $a \subseteq x$ to its direct image $f[a] = \{f(z) : z \in a\}$.

Class Here the objects are formulas in the language of set theory $\varphi(x, y_1, \ldots, y_n)$ together with n sets $a_1, \ldots, a_n$. (We think of this as $\{b : \varphi[b, a_1, \ldots, a_n]\}$.) The morphisms are then triples consisting of two formulas with parameters defining the domain and codomain, and a third one with two free parameters defining the action of the morphism.

Functors on Class The functors of interest are again the power polynomials. They are defined on Class similarly to the way they are defined on Set. The main difference between Set and Class for our purposes is that in Set we cannot solve $\mathcal{P}(x) = x$, while we can do so in Class.

2.7 Algebras for a functor

Let F be an endofunctor on a category C. An *algebra for* F is a pair (c, f), where c is an object of C, and $f : Fc \to c$.

Here is a basic example that illustrates why these are called *algebras*. Let's take the category Set of sets, and the functor

$$Ha = (a \times a) + (a \times a)$$

For the object N of natural numbers, HN is thus two copies of $N \times N$.

One example of an algebra for this functor is (N, α), where $\alpha(a, b) = a + b$ for $\langle a, b\rangle$ red, and $\alpha(a, b) = a \times b$ for $\langle a, b\rangle$ blue.

Getting back to the terminology of "algebra", the point is that the function α does the work of the two tables. The function "is" the tables.

Here is another example of an algebra. This time we are concerned on Set with $Fx = x+1$, as defined above. The algebra we have in mind is (N, s). Here $s : N + 1 \to N$ takes the natural number n to its successor $n + 1$, and the new point in $N + 1$ to the number 0.

The advantage of the categorical formulation is that the usual notions of a *morphism of algebras* turn out to be special case of a more general definition.

Let (c, f) and (d, g) be algebras for the same functor. A *morphism of algebras from* (c, f) *to* (d, g) is a morphism $\alpha : c \to d$ in the category C so that the diagram below commutes:

$$\begin{array}{ccc} Fc & \xrightarrow{f} & c \\ {\scriptstyle F\alpha}\downarrow & & \downarrow{\scriptstyle \alpha} \\ Fd & \xrightarrow[g]{} & d \end{array}$$

It now is clear that we have a category of algebras for a given functor. And so we immediately have the concept of *initial* and *final* algebras. There is no guarantee that these exist, but in many interesting cases they do. The reason we are interested in initial algebras is their connection to *recursion*.

To see this in detail, we return to the functor $Fx = x + 1$ on Set. We saw the algebra (N, s) above. We claim that this is an initial algebra. What this means is that for any algebra (A, a), there is a unique algebra morphism from $h : (N, s) \to (A, a)$. That is, the diagram below commutes:

$$\begin{array}{ccc} N+1 & \xrightarrow{s} & N \\ {\scriptstyle h+1}\downarrow & & \downarrow{\scriptstyle h} \\ A+1 & \xrightarrow[a]{} & A \end{array}$$

Now the map a is the same as a map $i : A \to A$ together with a choice of some element $b \in a$. And to say that the diagram above commutes is the same thing as saying that $h(0) = b$, and for all $n \in N$, $h(s(n)) = a(h(n))$.

Stepping back, the purported initiality of (N, s) is the same as the following assertion:

> For every set A, every $b \in A$, and every $a : A \to A$, there is a unique function $h : N \to A$ such that $h(0) = b$, and for all $n \in N$, $h(s(n)) = a(h(n))$.

This is the standard form of the Principle of Recursion on N. The upshot is that this principle is equivalent to the assertion that (N, s) is an initial algebra of the functor $Fx = x + 1$.

One way to interpret this equivalence is that we can take the existence of an initial algebra for $Fx = x+1$ as an *axiom of set theory*, in place of the usual Axiom of Infinity. That axiom says that there is an algebra for the singleton functor $Sx = \{x\}$ on sets which contains $\emptyset$ as an element and whose structure is the inclusion. This principle is easier to state than the algebraic reformulation. It takes a bit of work to use the simpler standard formulation to derive the Recursion Principle, and this is one of the basic topics in any course on axiomatic set theory.

Two general facts: First, the structure map of an initial algebra on Set is always a bijection. By a very general result in category theory due to Lambek, structure maps of initial algebras are always categorical isomorphisms. And categorical isomorphisms on Set always are such. And from this we see that $\mathcal{P}$ has no initial algebras on Set, by Cantor's Theorem.

Initial algebras for polynomial functors on Set Let $F : \mathsf{Set} \to \mathsf{Set}$ be a power polynomial functor. We know that F is monotone (it preserves the subset relation on sets), and it is not hard to check a slightly stronger property: F preserves inclusion maps between classes: An *inclusion* is a map $i_{a,b} : a \to b$ on classes which "doesn't do anything": $a \subseteq b$, and $i(x) = x$ for all $x \in a$. We say that F is *standard* if it preserves inclusions in the sense that $Fi_{a,b} = i_{Fa,Fb}$. Once again, every power polynomial endofunctor on Set is standard.

The polynomial operations on sets (without power) are also *continuous*: they preserve countable unions of sets.

Let $F : \mathsf{Set} \to \mathsf{Set}$ be a polynomial endofunctor. We sketch the proof that the least fixed point F_* carries the structure of an initial algebra, together with the identity on it.

One forms the increasing sequence

$$0 \xrightarrow{!} F0 \xrightarrow{F!} F^2 0 \xrightarrow{F^2!} \cdots$$

We write 0 for $\emptyset$. Each of the maps shown is an inclusion, by standardness. Let F_* be the union of the increasing sequence $F^n 0$ of sets. Then $F(F_*) = F_*$ by continuity. So (F_*, id) is an algebra for F. To check initiality, let (A, a) be an algebra for F, so $a : Fa \to a$. Define maps $g_n : F^n 0 \to A$ by recursion, with $g_0 : 0 \to A$ the empty function (this is what initiality of $\emptyset$ amounts to), and $g_{n+1} = a \circ Fg_n$. Check that we have an increasing sequence of functions

$$g_0 \subseteq g_1 \subseteq g_2 \subseteq \cdots$$

then take the union to get $\varphi : F_* \to A$. It turns out that this φ is a morphism of F-algebras, and indeed is the only such.

2.8 Coalgebras for a functor

We now turn to coalgebras. Again, let F be an endofunctor on a category C. A *coalgebra for* F is a pair (c, f), where c is an object of C, and $f : c \to Fc$. Comparing this to the definition of an algebra, we can see that a coalgebra is the same kind of structure, except that the direction of the arrow is reversed.

For example, every graph is a coalgebra of $\mathcal{P}$ on Set. That is, every graph $(G, \to)$ may be re-packaged as (G, e), with $e : G \to \mathcal{P}G$ given by $e(x) = \{y \in G : x \to y\}$. In words, we trade in the edge relation of a graph with the function that assigns to each point its set of children. This re-packaging has an inverse, and so the notions of "graph as set with relation" and "graph as coalgebra of $\mathcal{P}$" are in this sense notational variants.[8]

Let (c, f) and (d, g) be coalgebras for the same functor. A *morphism of coalgebras from* (c, f) *to* (d, g) is a morphism $\alpha : c \to d$ in the category C so that the diagram below commutes:

$$\begin{array}{ccc} c & \xrightarrow{f} & Fc \\ {\scriptstyle\alpha}\downarrow & & \downarrow{\scriptstyle F\alpha} \\ d & \xrightarrow[g]{} & Fd \end{array}$$

A coalgebra (c, f) is a *final* (or *terminal*) coalgebra if for every coalgebra (d, g), there is a *unique* morphism of coalgebras $\alpha : (d, g) \to (c, f)$.

Here is another example as we wind our way back to set theory. These are based on discussions at the beginning of these notes, concerning streams of numbers (Section 1.1). We are dealing with the functor $Fa = N \times a$. Then a system of stream equations is a coalgebra for F. To see how this works in a concrete case, we return to (2), reiterated below:

$$\begin{array}{rcl} x & \approx & \langle 0, y\rangle \\ y & \approx & \langle 1, z\rangle \\ z & \approx & \langle 2, x\rangle \end{array}$$

We re-package this as (X, e), where $X = \{x, y, z\}$, $e(x) = \langle 0, y\rangle$, $e(y) = \langle 1, z\rangle$, and $e(z) = \langle 2, x\rangle$. So now we know examples of coalgebras for this F. Another coalgebra for F uses the set N^∞ of streams as its carrier set. The coalgebra itself is $(N^\infty, \langle \mathsf{head}, \mathsf{tail}\rangle)$. We claim that this coalgebra is final. This means that for every stream system (X, e), there is a unique $e^\dagger : X \to N^\infty$ such that the diagram below commutes:

$$\begin{array}{ccc} X & \xrightarrow{e} & N \times X \\ {\scriptstyle e^\dagger}\downarrow & & \downarrow{\scriptstyle Fe^\dagger} \\ N^\infty & \xrightarrow[\langle \mathsf{head}, \mathsf{tail}\rangle]{} & N \times N^\infty \end{array}$$

To see what this is saying, the key is to understand the map $Fe^\dagger$. (This is exactly the place where category theory comes in, since we apply a functor to a morphism.) It takes a pair $\langle i, s\rangle$ to $\langle i, e^\dagger(s)\rangle$. Now the commutativity of the diagram above means that for every $x \in X$,

$$\langle \mathsf{head}, \mathsf{tail}\rangle(e^\dagger(x)) \quad = \quad \langle \mathsf{first}(e(x)), e^\dagger(\mathsf{second}(e(x)))\rangle$$

(We are using first and second for the projection maps from $N \times X$ to N and X, respectively.) So if $e(x) = \langle 13, z\rangle$, then we would have $e^\dagger(x) = \langle 13, e^\dagger(z)\rangle$. The upshot is that a solution to a stream system is a morphism into $(N^\infty, \langle \mathsf{head}, \mathsf{tail}\rangle)$. Thus the assertion that the latter coalgebra is final is the same as the assertion that every stream system has a unique solution.

Much the same applies to the tree example from Section 1.2.

[8] This point is weakened a bit when one considers the morphisms. For graphs, there are several natural notions of morphism; one is a map between the nodes that preserves edges (in one direction). The notion of a morphism of $\mathcal{P}$-coalgebras gives a different concept.

2.9 The axioms again

At this point we rephrase *FA* and *AFA* to make a comparison. Recall that V is the class of all sets, and that $V = \mathcal{P}V$. This means that (trivially) the identity on the universe maps V onto $\mathcal{P}V$, and vice-versa. Despite this, we want to introduce notation for these two maps that makes them different. We shall write

$$\begin{array}{l} i : \mathcal{P}V \to V \\ j : V \to \mathcal{P}V \end{array}$$

Thus i takes a multiplicity (a set of sets) and regards it as a unity (a set); j takes a set and regards it as a set of sets.

The Foundation Axiom in Algebraic Form Except for not being a set, (V, i) is an initial algebra for $\mathcal{P}$: for all sets a and all $e : \mathcal{P}(a) \to a$, there is a unique $s : V \to a$ such that

$$\begin{array}{ccc} \mathcal{P}V & \xrightarrow{i} & V \\ {\scriptstyle \mathcal{P}s}\downarrow & & \downarrow{\scriptstyle s} \\ \mathcal{P}a & \xrightarrow[e]{} & a \end{array} \tag{18}$$

That is, for all sets x,

$$s(x) \quad = \quad e(\{s(x') : x' \in x\}) \tag{19}$$

The Anti-Foundation Axiom in Coalgebraic Form Except for not being a set, (V, j) is a final coalgebra for $\mathcal{P}$: for every set b and every $e : b \to \mathcal{P}b$, there exists a unique $s : b \to V$ such that $s = \mathcal{P}s \circ e$:

$$\begin{array}{ccc} b & \xrightarrow{e} & \mathcal{P}b \\ {\scriptstyle s}\downarrow & & \downarrow{\scriptstyle \mathcal{P}s} \\ V & \xrightarrow[j]{} & \mathcal{P}V \end{array} \tag{20}$$

The map s is called the *solution* to the *system* e.

Class forms We only mentioned forms of the axioms pertaining to sets. They are a little nicer when stated as axioms on Class:
FA is equivalent to the assertion that (V, i) is an initial algebra for $\mathcal{P}$ on Class.
AFA is equivalent to the assertion that (V, j) is a final coalgebra for $\mathcal{P}$ on Class.

2.10 Conceptual comparison

We indicate in Figure 2 a kind of conceptual comparison of iterative and coiterative ideas. The entries towards the top are *dualities* in the categorical sense. Moving downwards, the rows in the chart are more like research directions than actual results. So spelling out the details in the chart

algebra for a functor	coalgebra for a functor
initial algebra	final coalgebra
least fixed point	greatest fixed point
congruence relation	bisimulation equivalence relation
equational logic	modal logic
recursion: map out of an initial algebra	corecursion: map into a final coalgebra
Foundation Axiom	Anti-Foundation Axiom
iterative conception	coiterative conception
set with operations	set with transitions and observations
useful in syntax	useful in semantics
bottom-up	top-down

Figure 2: The conceptual comparison

For many functors on Set, especially polynomial functors and the finite power set functor, the initial algebra is the least fixed point together with the identity. For the polynomial functors, this least fixed point is itself an algebra of terms.

The connection between greatest fixed points and final coalgebras is the content of the following result.

Theorem 2.8 (Aczel) *For every power polynomial F on* Class*, the greatest fixed point together with the identity on it, (F^*, id), is a final coalgebra of F on* Class*. Moreover, if F is a polynomial functor, then F^* is a set and (F^*, id), is a final coalgebra of F on* Set.

The original result used much weaker hypotheses on F, using notions which we did not define, so our statement is rather weaker than in Aczel's book. Several papers have gone on to weaken strengthen this Final Coalgebra Theorem.

Bisimulation We have given the definition of bisimulation earlier. We discussed it in connection with graphs, but the reader may also know of a notion with the same name coming from modal logic. Actually, the theory of coalgebra studies a more general notion, that of bisimulation on a coalgebra for a given functor, defined first in Aczel and Mendler [4]. This more general notion specializes to several concepts which had been proposed in their own fields. In addition, it is (nearly) the dual concept of a congruence on an algebra; this explains our line in the conceptual comparison chart.

Equational logic and modal logic A great deal of work has shown ways in which equational logic and modal logic are "dual", but to spell this out in detail would require quite a bit more category theory than we need in the rest of these notes.

There is a growing field of coalgebraic generalizations of modal logic. For a survey of this area, see Kurz [16].

The final coalgebra of a functor may be regarded as a space of *complete observations*. (As with all our points in this section, this statement is mainly for functors on Set, and the notion of "complete observation" is, of course, merely suggestive.) For example, let At be a set whose elements are called *atomic propositions*, and consider the functor $F(a) = \mathcal{P}_{fin}(a) \times \mathcal{P}(\mathsf{At})$. A coalgebra for this is a set a together with one map of a into its finite subsets, and another into the collection of sets of atomic propositions. Putting the two maps together gives a *finitely-branching* Kripke model: each point has finitely many children and some set of atomic propositions. Now modal logic gives us a way of "observing" properties of points in coalgebras (Kripke models). And the record of everything that one could observe from a point is the modal theory of that point. Further, one may take the collection of all theories of all points in all finitely-branching Kripke models and make this collection (it is a set) into the carrier of a final coalgebra for the functor. Indeed, this would be one way to construct a final coalgebra.

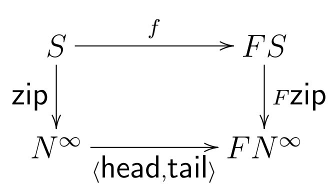
Corecursion Returning now to the chart, we present an example of a corecursive definition. We mentioned in (3) how the zip function on streams is to work. It should satisfy

$$\mathsf{zip}(s,t) \quad = \quad \langle \mathsf{head}(s), \mathsf{zip}(t, \mathsf{tail}(s)) \rangle \tag{21}$$

Here is how zip is uniquely defined via a corecursive definition. Write $N^\infty \times N^\infty$ as S in this discussion. We want a map from S to N^∞. We are dealing with S as the final coalgebra of the functor $Fa = N \times a$, and we'll write the structure on the final coalgebra as $\langle \mathsf{head}, \mathsf{tail} \rangle$, just as we did it in Section 1.1. The idea is to turn S into the carrier set of a coalgebra for, say (S, f). Then zip will be the unique coalgebra morphism from (S, f) to $(S, \langle \mathsf{head}, \mathsf{tail} \rangle)$. It remains to define f. Let

$$f(s,t) \quad = \quad \langle \mathsf{head}(s), \langle t, \mathsf{tail}(s) \rangle \rangle.$$

As mentioned, by finality there is a unique $\mathsf{zip} : S \to N^\infty$ so that the diagram below commutes:



To make sure that this works, we follow an arbitrary pair of streams, say $\langle s, t \rangle$ around the square, starting in the upper-left. Going down, we have $\mathsf{zip}(s,t) \in N^\infty$. From this, the structure takes this to $\langle \mathsf{head}(\mathsf{zip}(s,t)), \mathsf{tail}(\mathsf{zip}(s,t)) \rangle \in FN^\infty$. But we could also take our $\langle s, t \rangle$ across the top via f to get $\langle \mathsf{head}(s), \langle t, \mathsf{tail}(s) \rangle \rangle$. Now $F\mathsf{zip}$ applies to this pair, and this is where the action of F as a functor enters. We get $\langle \mathsf{head}(s), \mathsf{zip}(\langle t, \mathsf{tail}(s)\rangle)\rangle$. So overall, we have

$$\begin{array}{rcl} \mathsf{head}(\mathsf{zip}(s,t)) & = & \mathsf{head}(s) \\ \mathsf{tail}(\mathsf{zip}(s,t)) & = & \mathsf{zip}(t, \mathsf{tail}(s)) \end{array}$$

just as desired. It says: to zip two streams, start with the head of the first, and then repeat this very process on the second followed by the tail of the first.

Sets, again We have already discussed at length the lines in the table concerning the Foundation and Anti-Foundation Axioms, and their attendant conceptual backgrounds. The point of this section is to situation that entire discussion inside of a larger one.

Examples of final coalgebras and corecursive definitions Our conceptual comparison makes the point that algebras embody sets with operations. This point is almost too easy: the reason behind the terminology of "algebras" in category theory is that sets with operations may be modeled as algebras in the categorical sense. For coalgebras, it is harder to make the case that they directly correspond to sets with either "transitions" or "observations". However, we present a few examples that motivate this point.

We collect in Figure 3 a list of functors on Set or Class along with final coalgebras or other data from Figure 2. We assume *AFA* in this discussion.

First, for any set S, the functor $Fa = S \times a$. A coalgebra for this F is a *stream system of equations* as we saw it in Section 1.1, except that there we made things concrete and took S to be the set of natural numbers. The final coalgebra is the set $S^\infty = S \times S^\infty$ of *streams over* S. The logical language for this functor would be a sentential (propositional) language whose sentences are either true or of the form $s : \varphi$, where $s \in S$. The semantics would be the obvious one; for example

$$(0, 1, 2, 3, \ldots) \models 0 : 1 : 2 : \mathsf{true}.$$

One should note that carrier of the final coalgebra may be taken to be certain theories in this language. These may be described extrinsically as the theories of all points in all coalgebra. It is more informative, however, to set down a logical system and then consider the maximal consistent sets in the system. With the right definition, the maximal consistent sets do turn out to be the carrier of a final coalgebra for the functor.

Second, we consider $Fa = (S \times a) + 1$. Here $1 = \{0\}$ and $+$ is the disjoint union. However, it is more common for people to represent the one and only element of 1 using a symbol like $*$. The coalgebras are like stream systems of equations, except now an equation might ask for a stream to "stop" by having $*$ on the right-hand side. So an example of a coalgebra would be $x \approx \langle s, y\rangle$, $y \approx *$. Then the solution would take $x^\dagger$ to be the one-term sequence s. The logic for this functor would be the same logic (HML) as before, except that now we add an atomic sentence to detect the ends of finite sequences.

Turning to the last two lines, we already know that *AFA* is equivalent to the assertion that (V, id) is a final coalgbra of $\mathcal{P}$; also, even without *AFA*, we have a final coalgebra whose

Functor	coalgebra	final coalgebra	logic
$S \times a$	stream system	streams over S	Hennessy-Milner logic (HML)
$(S \times a) + 1$	stream system, allowing ends	finite sequences over S and streams over S	add “end” to HML
$\mathcal{P}a$	set equations; graph	hyperset; pointed graph modulo bisimulation	infinitary modal, no atoms
$\mathcal{P}_{fin}a \times \mathcal{P}(\mathsf{At})$	finitely-branching Kripke model	a certain subset of the canonical model of K	modal logic over At

Figure 3: Examples of functors and related notions from coalgebra

carrier set is the pointed graphs modulo bisimulation. The logic in this case is infinitary modal logic. It turns out that two points in a given coalgebra have the same infinitary modal theory iff they are bisimilar.

The line concerning $\mathcal{P}_{fin}a \times \mathcal{P}\mathsf{At}$ is the closest to the Kripke semantics of modal logic. One might hope that the final coalgebra would turn out to be the canonical model of the modal logic K, but this is not quite right. One needs to cut down to those maximal consistent sets which are realized by some point in some *finitely branching* model.

The lines in at the bottom of the conceptual comparison chart are the most programmatic of all.

Doing without ***AFA*****: final coalgebras in** ***ZF*** We mentioned in footnote 2 that it is possible to alter the pairing operation in such a way that one may prove many of the results that our treatment obtains only by using *ZFA*. This points is mentioned in Forster [13] and developed in detail in Paulson [23] (and in other papers by Paulson). One replaces the Kuratowski $\langle x, y\rangle$ with a variant, $(\{0\} \times a) \cup (\{1\} \times b)$. (This is the usual disjoint union operation, also called the *coproduct* on sets.) Then one defines variant oother things: the cartesian product, functions, etc. And in terms of these one can indeed study streams and infinite trees, and many other sets of interest. Even more, one can prove the *final coalgebra theorems* which we shall state later.

One might think that this move undermines much of the interest in *AFA*. For Paulson, the reduction is important since he wants to use an automatic theorem prover to work with assertions in set theory. It makes sense to work out detailed reductions so as to avoid changing the set theory.

I don't think that others will find this conclusive, for two reasons. First, the method doesn't apply to equations like $x = \{x\}$, or to collections like $x = \mathcal{P}_{fin}(x)$. The latter kind of equation is especially useful in applications. But even more, what will be of interest will be the whole assembly of what we might call *coalgebraic concepts*: coinduction, corecursion, and top-down treatments of various phenomena. Someone who is using these concepts and is also worried about modeling in set theory would probably find it convenient to work with *AFA*, even if many of the end applications could be done in standard set theory.

3 Coalgebraic Bisimulation

In this preliminary section, we discuss the main idea behind the construction so that you can refer back to it as you read the details. Although we are most centrally interested in $\mathcal{P}_{fin}$, much of what we do will hold more generally. In this section, F is any functor on sets, but you should read things with special attention to the case when $F = \mathcal{P}_{fin}$.

The most basic overall idea is to take the *disjoint union* $\mathcal{U}$ of all the F-coalgebras in the entire world. Now this big collection has something going for it: every coalgebra certainly maps into it. But it has two problems:

1. It is too big to be a set, hence it won't really be a coalgebra in the first place.

2. A given coalgebra has many morphisms into it, and so it won't be final.

All of the rest of the work here is about solving these problems. Let's think about the second one first. Our plan will be to find a certain equivalence relation $\equiv$ on the huge disjoint union and then to take the *quotient* by it. Clearly we want to identify two points, say x and y, if there is a coalgebra, say (A, a) and morphisms, say $\varphi : A \to \mathcal{U}$ and $\psi : A \to \mathcal{U}$ such that $\varphi(a) = x$ and $\psi(a) = y$. In the case of $F = \mathcal{P}_{fin}$ and $F = \mathcal{P}$, we'll have a very nice description of the equivalence relation $\equiv$ that we are after. And then it will turn out that $\mathcal{U}/\equiv$ has the property that every coalgebra has exactly one morphism into it. However, the quotient might still not be a set, and so there is a little more work to be done to get around this.

The main idea behind the equivalence relation that we'll study comes from thinking about *kernel relations of coalgebra morphisms.* In general, if $f : A \to B$ is any function, then the *kernel of* f is

$$K_f \quad = \quad \{a_1, a_2) \in A \times A : f(a_1) = f(a_2)\}.$$

Suppose now that A is the carrier of a coalgebra (A, α) for some functor F. Let (C, γ) be a final coalgebra for the same functor F. If (a_1, a_2) belongs to K_f for *any* coalgebra morphism $\varphi : (A, \alpha) \to (B, \beta)$, then γa_1 and γa_2 must be the same, since the final coalgebra morphism for α must factor through the final coalgebra morphism for β.

Hope 1 For any coalgebra (A, α) for F, the relation

$$K_A = \{a_1, a_2) \in A \times A : \text{for some coalgebra morphism } \varphi,\ \varphi(a_1) = \varphi(a_2)\}$$

is an *equivalence relation on* A.

As we shall see, this holds for many functors F of interest to us.

Hope 2 We can describe this relation K_A in concrete terms. That is, we can tell whether $(a_1, a_2) \in K_A$ just by looking at the coalgebra (A, α), without having to look at all of the coalgebras and morphisms in the entire world.

Again, this turns out to hold. It also will be of permanent interest for our study of coalgebra.

Hope 3 Although the collection of F-coalgebras is too large to be a set, there should be only a set $\mathcal{X}$ of *one-point-generated* coalgebras. This is a special kind of

We let $\mathcal{U}$ be the coproduct (=disjoint unioin) of all coalgebras in $\mathcal{X}$, and let $\mathcal{C}$ the quotient of $\mathcal{U}$ by the relation $\equiv$ in Hope 1. Then this quotient $\mathcal{U}$ itself has a coalgebra structure. Indeed, $(\mathcal{U}, \rightarrow)$ is a final coalgebra:

Hope 4 For any F-coalgebra $A = (A, \alpha)$ there unique coalgebra morphism from A to $\mathcal{C}$. It is given by taking, for each $a \in A$, some $(B, \beta) \in \mathcal{X}$ isomorphic to the sub-coalgebra A_a of A generated by a, and then taking the image of a in the composite

$$A_a \longrightarrow B \longrightarrow \mathcal{U} \longrightarrow \mathcal{C} \,.$$

This is our plan.

We are going to see in Section 3.1 how all of this works for $F = \mathcal{P}_{fin}$. We then propose a generalized theory in Section 3.5. Finally, we see how the work of Section 3.1 generalizes.

3.1 Bisimulation relations on graphs

At this point, we turn to the study of a certain relation on graphs which is intimately related to morphisms of coalgebras. It's of independent interest; that is, there are reasons to study it beyond the matter of constructing final coalgebras. So we take up bisimulation here, and return in the next section to the application of it to the bigger project in this chapter.

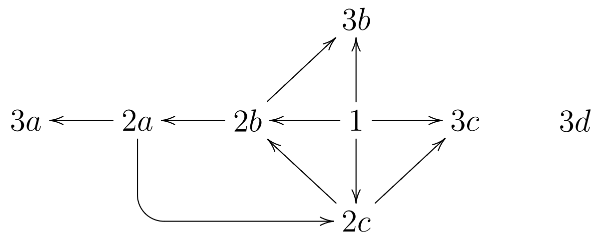
Let $(G, \rightarrow)$ be a graph. A relation R on G is a *bisimulation* iff the following holds: whenever xRy,

(zig) If $x \rightarrow x'$, then there is some $y \rightarrow y'$ such that $x'Ry'$.

(zag) If $y \rightarrow y'$, then there is some $x \rightarrow x'$ such that $x'Ry'$.

Bisimulation between graphs Before giving examples, we should clarify some usage. At a few points, we'll speak of bisimulation *between* two graphs G and H, rather than *on* a single graph. This can be defined in the same general way. Note also that one can take the *disjoint union* $G + H$ of the graphs G and H, and then a bisimulation between G and H would be a bisimulation on $G + H$.

Returning to bisimulation on a graph For an example, let's look at the following graph G:



All of the 3-points have no children. (Point $3d$ is not reached from any other point, but the arrows *into* a node are of no interest.) So every relation which only relates 3-points is a bisimulation on G. Concretely,

$$\{(3a, 3b), (3c, 3a), (3d, 3d)\}$$

is easily seen to be a bisimulation.

For that matter, the empty relation is also a bisimulation on G.

Another bisimuation is

$$\{(2a, 2b), (2b, 2c), (2c, 2a)\} \cup \{(3a, 3b), (3b, 3c), (3c, 3a)\}.$$

Let's call this relation R. It would take a lot of checking to actually verify that R is a bisimulation. Here is just two items of it: We see that $2b \; R \; 2c$. Now $2c \to 2b$. Thus we need some node x so that $x \; R \; 2b$ and $2b \to x$. For this, we take $2a$. For our second point of verification, again note that $2b \; R \; 2c$. Since $2b \to 3b$, we need some node x so that $2c \to x$ and $3b \; R \; x$. We take $x = 3c$ for this.

The largest bisimulation on our graph G is the relation that relates 1 to itself, all 2-points to all 2-points, and all 3-points to all 3-points. Note that this is an equivalence relation: reflexive, symmetric, and transitive. This is not an accident, in view of the next two results.

Proposition 3.1 *Let $(G, \to)$ be a graph.*

1. *The diagonal relation $\Delta = \{(g, g) : g \in G\}$ is a bisimulation on G.*
2. *If R is a bisimulation on G, so is R^{-1}.*
3. *If R and S are bisimulations on G, so is $R \circ S$.*
4. *If $\mathcal{S}$ is any set of bisimulations on G, then $\bigcup_{R \in \mathcal{S}} R$ is also a bisimulation on G.*

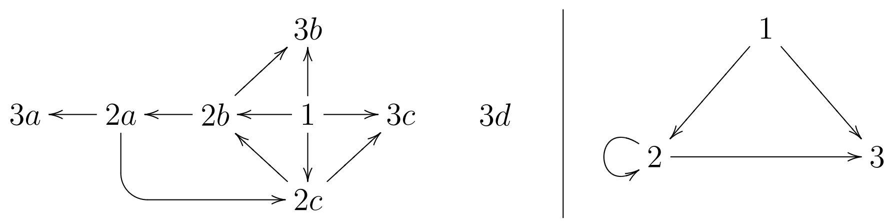
Proof These are easy verifications. ⊣

Lemma 3.2 *For any graph* $(G, \to)$*, there is a largest bisimulation on* G*. This relation is an equivalence relation called* bisimilarity *and denoted* $\equiv$*. It is also characterized by*

$$x \equiv y \quad \textit{iff} \quad \textit{there is a bisimulation on } G \textit{ relating } x \textit{ to } y.$$

Proof Let $\equiv$ be the union of all bisimulation relations on G. By Proposition 3.1, $\equiv$ is an equivalence relation. ⊣

We can always form the *quotient graph* using the largest bisimulation. Here is how this works, using G from above as an example. In $G/{\equiv}$, we would have three nodes, corresponding to the three equivalence classes under the largest bisimulation; let's call these 1, 2 and 3. We put an arrow between two equivalence classes if some (every) element of the first has an arrow to some element of the second. In this way, we construct the quotient. Here is a picture of G again, along with its quotient $G/{\equiv}$ under the largest bisimulation:



The map from G to $G/{\equiv}$ takes the 2-points to 2 and the 3-points to 3.

Lemma 3.3 *Let* $(G, \to)$ *be a graph, let* $\equiv$ *be the largest bisimulation on it, and consider the quotient* $G/\equiv$*. Every bisimulation relation* R *on* $G/\equiv$ *is a subrelation of the diagonal* Δ*.*

Proof Let S be the relation on G defined by $g\ S\ h$ iff $[g]\ R\ [h]$. We check that S is a bisimulation on G. From this it follows that if $[g]\ R\ [h]$, then actually $[g] = [h]$.

For the bisimulation verification, we only consider the (zig) part. Suppose that $g\ S\ h$, and also $g \to g'$. We know that $[g]\ R\ [h]$. In $G/\equiv$, $[g] \to [g']$. Since R is a bisimulation, there is some $[x]$ such that $[h] \to [x]$ in $G/\equiv$, and also $[g']\ R\ [x]$. Because $[h] \to [x]$, we have some $h' \equiv h$ and some $x' \equiv x$ such that $h' \to x$. Since $\equiv$ is a bisimulation, and since $h' \equiv h \to x$, there is some h'' such that $h \to h''$ and $h'' \equiv x' \equiv x$. Now h'' is the point we want: $h \to h''$, and $g\ S\ h''$; the last point holds because $[g']\ R\ [h''] = [x]$. ⊣

This result leads to a definition.

Definition A graph G is *simple* if every bisimulation relation R on G is a subrelation of the diagonal Δ.

Using Lemma 3.3, every quotient of a graph by its bisimilarity relation is simple, and every simple graph arises in this way.

3.2 Generated subgraphs

Proposition 3.4 *Let $(G, \to)$ be a graph, let $g \in G$, and let*

$$G_g \quad = \quad \{h \in G : \text{for some } n \geq 0, \text{ there is a path of length } n \text{ in } G \text{ from } g \text{ to } h\}.$$

Regard G_g as an induced subgraph of G. Then the inclusion $i : G_g \to G$ is a $\mathcal{P}$-coalgebra morphism. Moreover, if G is finitely branching, then G_g is countable and also finitely branching, and the inclusion is a morphism of $\mathcal{P}_{fin}$-coalgebras.

Proof It is easy to check that i is a $\mathcal{P}$-coalgebra morphism. The only thing that requires some checking is the assertion that G_g is countable. But for each n, the set of nodes of G reachable from g by a path of length n is finite. So G_g is a countable union of finite sets, and thus is countable. ⊣

G_g has various names: as a set, it would be the *set of descendants of g in G*. As a graph it could be called the *the subgraph generated by g*, the *part of G accessible from g*, or the *subgraph of G induced by the descendants of g*.

Let G be a graph, and consider the disjoint union $\coprod_{g \in G} G_g$ of all generated subgraphs of G. Technically, this set is

$$\{(G_g, h) : h \in G_g \text{ and } g \in G\}.$$

It carries a coalgebra structure given by

$$(G_g, h) \to (G_k, l) \quad \text{iff} \quad g = k \text{ and } h \to l \text{ in } G_g.$$

Lemma 3.5 *Concerning coproducts of generated subgraphs,*

1. *For each $h \in G$, $i_h : G_h \to \coprod_{g \in G} G_g$ is an injective coalgebra morphism, where*

$$i_h(k) \quad = \quad (G_h, k).$$

2. *The map*

$$\epsilon : \coprod_{g \in G} G_g \to G$$

given by $\epsilon(G_g, h) = h$ is a surjective coalgebra morphism. For each $g \in G$, the preimages of g are exactly the pairs (G_h, g) where h is an ancestor of g in G.

3.3 Bisimulations and kernels

Up until now, we have said what bisimulation is, but we did not describe its relation to kernels. To rectify matters, here is the main result.

Theorem 3.6 *Let G be a graph, considered as a coalgebra for $\mathcal{P}$. Let $R \subseteq G \times G$. Then the following are equivalent:*

1. R is the kernel of some $\mathcal{P}$-coalgebra morphism φ.

2. R is a bisimulation relation and an equivalence relation.

Moreover, if G is finitely branching, then the same equivalences hold, where we require in (1) that R be the kernel of some $\mathcal{P}_{fin}$-coalgebra morphism.

Proof Suppose first that R is the kernel of $\varphi : (G, e) \to (H, f)$. Clearly R is an equivalence relation. To check that it is a bisimulation on G, suppose that $g\ R\ h$, and $g \to g'$; thus $g' \in e(g)$. We know that $\varphi(g) = \varphi(h)$, so also $f\varphi(g) = f\varphi(h)$. And as φ is a coalgebra morphism, $\mathcal{P}\varphi \circ e = f \circ \varphi$. This means that $\mathcal{P}\varphi(e(g)) = \mathcal{P}\varphi(e(h))$. Note that $\varphi(g') \in \mathcal{P}\varphi(e(g)) = \mathcal{P}\varphi(e(h))$. This means that there is some h' such that $h \to h'$ and $\varphi(g') = \varphi(h')$. (Please be sure that you understand this last point. It is the key to the argument.) And so $g'\ R\ h'$, since R is the kernel of φ.

The other part of the bisimulation verification is similar, mutatis mutandis.

In the other direction, let R be a bisimulation which is also an equivalence relation. Consider the quotient G/R. It carries a coalgebra structure by

$$[g] \to [h] \quad \text{iff} \quad (\exists g'\ R\ g)(\exists h'\ R\ h)(g' \to h').$$

Since R is a bisimulation and an equivalence relation, this relation on G/R is well-defined, and it gives a coalgebra structure on G/R. We check that the natural map $\nu : G \to G/R$ is a coalgebra morphism, where $\nu(g) = [g]$. That is, we check that for all $g \in G$,

$$\{[h] : g \to h \text{ in } G\} \quad = \quad \{[k] : [g] \to [k] \text{ in } G/R\}.$$

In one direction, take an element of the set on the left, say $[h]$, where $g \to h$. Then clearly $[g] \to [h]$, so $[h]$ belongs to the set on the right. In the other direction, suppose that $[g] \to [k]$. Let $g'\ R\ g$ and $k'\ R\ k$ be such that $g' \to k'$. Since R is a bisimulation, we in fact have some g'' such that $g \to g''$ and $g''\ R\ k'$. And then $[g''] = [k'] = [k]$. So $[k] = [g'']$ belongs to $\{[h] : g \to h\}$, as desired.

Finally, the kernel of ν is exactly R. ⊣

We also need the following result. It has nothing to do with kernels, but rather with images of *parallel pairs* of morphisms.

Lemma 3.7 *Let $(G, \to)$ and $(H, \rightsquigarrow)$ be $\mathcal{P}_{fin}$-coalgebras, and let $\varphi, \psi : G \to H$ be two coalgebra morphisms. Then*

$$R \quad = \quad \{(\varphi g, \psi g) : g \in G\}$$

is a bisimulation on H.

Proof Suppose that $\varphi g\ R\ \psi g$. We check the (zig) condition. Suppose that $\varphi g \rightsquigarrow h$. Since φ is a coalgebra morphism, there is some $g' \in G$ such that $g \to g'$ and $\varphi g' = h$. But then by the definition of R, $\varphi g'\ R\ \psi g'$. And since $g \to g'$, $\psi g \rightsquigarrow \psi g'$. To summarize: given a child h of φg in G, we found a child $\psi g'$ of ψg in H such that h and $\psi g'$ are related by R. ⊣

3.4 The final coalgebra of $\mathcal{P}_{fin}$

Definition Consider the collection $\mathcal{X}$ of finitely branching graphs whose node set is a set of natural numbers. The main thing to note is that $\mathcal{X}$ is a set; indeed

$$\mathcal{X} \subseteq \mathcal{P}(\omega) \times \mathcal{P}(\mathcal{P}(\omega) \times \mathcal{P}(\omega)).$$

Let $(\mathcal{U}, \rightarrow) = \coprod_{G \in \mathcal{X}} G$ be the coproduct (=disjoint union) of all graphs in $\mathcal{X}$. Thus the nodes in the graph $\mathcal{U}$ are the pairs (G, g), $G = (G, \rightarrow)$ where $G \in \mathcal{X}$ and $g \in G$. (We might write these nodes as triples $(G, \rightarrow, g)$, but we prefer to keep the notation simple by taking the first G to be the full graph structure.) The structure on $\mathcal{U}$ is

$$(G, g) \rightarrow (H, h) \quad \text{iff} \quad G = H \text{ and } g \rightarrow h \text{ in } G.$$

As always, we have a bisimilarity relation $\equiv$ on $\mathcal{U}$. $\mathcal{C}$ is the quotient graph $\mathcal{U}/\equiv$.

Theorem 3.8 *$(\mathcal{C}, \rightarrow)$ is a final coalgebra for $\mathcal{P}_{fin}$.*

Proof Let $(G, \rightsquigarrow)$ be a graph. We first check that there is at least one coalgebra morphism $\varphi : G \rightarrow \mathcal{C}$. For each $g \in G$, the generated subgraph G_g is a countable graph. Choose a graph H_g whose node set is included in N, fix an isomorphism $j_g : G_g \rightarrow H_g$, and also let φ_g be the composite

$$G_g \xrightarrow{j_g} H_g \xrightarrow{i_g} \mathcal{U} \xrightarrow{\nu} \mathcal{C}$$

The map $i_{H_g} : H_g \rightarrow \mathcal{U}$ is given by $i_g(x) = (H_g, x)$. As earlier, the natural map ν takes an element of $\mathcal{U}$ to its equivalence class under bisimilarity. Overall, for $k \in G_g$, $\varphi_k(g) = [(H_g, j_g(k)]$.

We then define $\varphi : G \rightarrow \mathcal{C}$ by

$$\varphi(g) \quad = \quad \varphi_g(g) \quad (= [(H_g, j_g(g)]).$$

We need to check that φ is a coalgebra morphism.

First, let $g \rightsquigarrow h$. Then G_h is a subgraph of G_g, and indeed $(G_g)_h = G_h$. The idea at this point is that H_h is thus isomorphic to $(H_g)_h$. Isomorphisms are bisimulations; rather, the graph of an isomorphism is a bisimulation relation, and so the image of g will have as a child the image of h.

In more detail, we consider the relation R on $\mathcal{U}$ given by

$$R \quad = \quad \{((H_g, j_g(k)), (H_h, j_h(k))) : k \in G_h\}.$$

It is easy to check that R is a bisimulation, and it is immediate that $(H_g, j_g(h)) \; R \; (H_h, j_h(h))$. Using this fact, we see that in $\mathcal{C}$,

$$[(H_g, j_g(g))] \quad \rightarrow \quad [(H_g, j_g(h))] \quad = \quad [(H_h, j_h(h))].$$

Therefore

$$\varphi(g) = \varphi_g(g) = [(H_g, j_g(g))] \to [(H_h, j_h(h))] = \varphi_h(h) = \varphi(h).$$

This is the (zig) part of bisimulation.

For the (zag) part, we need to check that every child of $\varphi(g)$ is of the form $\varphi(g)$ for some child h of h in G. Take such a child w of $[(H_g, j_g(g))]$. By the definition of the quotient, $w = [(H_g, j_g(x))]$ for some $x \in H_g$ such that $j_g(g) \to j_g(x)$. Since H_g is isomorphic to G_g by j_g, there is some h such that $j_g(h) = x$. Tracing through the definitions, we see that $\varphi(h) = w$.

We conclude with the uniqueness of φ. Suppose $\psi : G \to \mathcal{C}$ were a coalgebra morphism. By Lemma 3.7,

$$\{(\varphi(g), \psi(g)) : g \in G\}$$

is a bisimulation on $\mathcal{C}$. Since $\mathcal{C} = \mathcal{U}/\equiv$, we see from Lemma 3.3 that our set above is a subset of the diagonal on $\mathcal{C}$. This means that for all $g \in G$, $\varphi(g) = \psi(g)$. Hence $\varphi = \psi$. $\dashv$

3.5 Aczel-Mendler bisimulations

There are several aspects to our work on bisimulations:

1. A general definition of bisimulation between coalgebras for a given functor.
2. Examples of what bisimulations look like for different functors.
3. Some properties of this general definiton, leading to a construction of a final coalgebra.
4. A set of other results which depend on a hypothesis on the functor, that it *preserve weak pullbacks*.

Recall that a relation R on a set c is a subset of $c \times c$. For example, we always have the *diagonal relation* Δ_c on c, where $\Delta_c(x, x)$ for all $x \in c$. As such, we have projection maps $\pi_1, \pi_2 : R \to c$.

Let $F : \mathsf{Set} \to \mathsf{Set}$ be a functor, and let $f : c \to Fc$ be a coalgebra. A *bisimulation on c* is a relation $R \subseteq c \times c$ such that there exists $r : R \to FR$ such that both projections π_1 and π_2 give coalgebra morphisms:

$$\begin{array}{ccc} R & \xrightarrow{r} & FR \\ {\scriptstyle \pi_1}\downarrow & & \downarrow{\scriptstyle F\pi_1} \\ c & \xrightarrow[f]{} & Fc \end{array} \qquad \begin{array}{ccc} R & \xrightarrow{r} & FR \\ {\scriptstyle \pi_2}\downarrow & & \downarrow{\scriptstyle F\pi_2} \\ c & \xrightarrow[f]{} & Fc \end{array} \tag{22}$$

More generally, if $f : c \to Fc$ and $g : d \to Fd$ are coalgebras, then a *bisimulation between c and d* is again $R \subseteq c \times d$ for which there is some $r : R \to FR$ such that both projections π_1 and π_2 give coalgebra morphisms.

These definitions are due to Aczel and Mendler.

3.6 The power set and finite power set

Recall that a $\mathcal{P}$-coalgebra $\langle G, f\rangle$ is essentially a graph, where we write $x \to y$ iff $y \in f(x)$.

Proposition 3.9 *Let $\langle G, f\rangle$ be a $\mathcal{P}$-coalgebra. Then a relation R on G is a $\mathcal{P}$-bisimulation according to the Aczel-Mendler definition iff it is a bisimulation in our earlier sense. That is, iff the following holds: whenever xRy,*

a. If $x \to x'$, then there is some $y \to y'$ such that $x'Ry'$.

b. If $y \to y'$, then there is some $x \to x'$ such that $x'Ry'$.

Proof Suppose first that R is a $\mathcal{P}$-bisimulation, say via r. Let xRy and $x \to x'$. We follow the pair (x, y) around both squares in (22), suitably specialized to our setting. We get

$$\begin{array}{lcl} f(x) & = & \{\pi_1(a,b) : (a,b) \in r(x,y)\} \\ f(y) & = & \{\pi_2(a,b) : (a,b) \in r(x,y)\} \end{array}$$

Since x' belongs to $f(x)$, the first equation implies that there is some y' such that $(x', y') \in r(x, y)$. Moreover, the second equation tells us that each such y' must belong to $f(y)$. At this point, we have verified condition (a) from the assumption that R is a $\mathcal{P}$-bisimulation. The verification of (b) is similar.

Now assume (a) and (b). Define $r : R \to \mathcal{P}R$ by

$$r(x,y) \quad = \quad \{(x',y') \in R : x \to x', y \to y'\}.$$

We check that the left square of (22) commutes; the right is similar. Fix (x, y) so that $R(x, y)$. Then

$$(f \circ \pi_1)(x,y) \quad = \quad f(x) \quad = \quad \{x' : x \to x'\}.$$

By (a), for every x' such that $x \to x'$, there is some y' such that $y \to y'$ and $x'Ry'$. That is, $(x', y') \in r(x, y)$. Thus $x' \in (F\pi_1 \circ r)(x, y)$. This for all $x' \to x$ shows that $f(x) \subseteq (F\pi_1 \circ r)(x, y)$. For the reverse inclusion, suppose that $x' \in (F\pi_1 \circ r)(x, y)$. Then there is some y' such that $r(x, y)$ contains (x', y'). By definition of r, we see that $x \to x'$; that is, $x' \in f(x)$. This shows that $(F\pi_1 \circ r)(x, y) \subseteq f(x)$. We conclude that $f \circ \pi_1 = F\pi_1 \circ r$. ⊣

3.7 Product with a fixed set

Fix a set C in this section, and consider the functor $F(a) = C \times a$.

Proposition 3.10 *Let $\langle a, f\rangle$ be an F-coalgebra. A relation $R \subseteq a \times a$ is a bisimulation iff the following holds: whenever xRy,*

a. $(\pi_1 \circ f)x = (\pi_1 \circ f)y$.

b. $(\pi_2 \circ f)x \, R \, (\pi_2 \circ f)y$.

Here π_1 *and* π_2 *are the projections from* $C \times a$.

Proof Suppose first that R is an F-bisimulation, say via r. Let xRy. Then

$$\begin{array}{rcl} (\pi_1 \circ f)x & = & (\pi_1 \circ f \circ \pi_1^R)(x,y) \\ & = & (\pi_1 \circ F\pi_1^R \circ r)(x,y) \\ & = & (\pi_1^{A\times R} \circ r)(x,y) \end{array}$$

By the same reasoning $(\pi_1 \circ f)y = (\pi_1^{A\times R} \circ r)(x,y)$. This verifies (a). As for (b), we similarly calculate $(\pi_2 \circ f)x =$

In the other direction, assume (a) and (b). Define $r : R \to C \times R$ by

$$\begin{array}{rcl} r(x,y) & = & ((\pi_1 \circ f)x, ((\pi_2 \circ f)x, (\pi_2 \circ f)y)) \\ & = & ((\pi_2 \circ f)x, ((\pi_2 \circ f)x, (\pi_2 \circ f)y)) \end{array}$$

We check that $F\pi_1 \circ r = f \circ \pi_1$; the work for π_2 is similar. For $(x,y) \in R$,

$$\begin{array}{rcl} (F\pi_1 \circ r)(x,y) & = & F\pi_1((\pi_1 \circ f)x, ((\pi_2 \circ f)x, (\pi_2 \circ f)y)) \\ & = & ((\pi_1 \circ f)x, (\pi_2 \circ f)x)) \\ & = & f(x) \\ & = & (f \circ \pi_1)(x,y) \end{array}$$

⊣

3.8 Some results on bisimulation

Lemma 3.11 *For any coalgebra* $\alpha : A \to FA$, Δ_A *is a bisimulation on* A.

Proof We'll write Δ for Δ_A here. The projections π_1 and π_2 are identical, and we just write π for them. Let $d : A \to \Delta$ be given by $d(x) = (x,x)$. It is easy to check that $\pi = d^{-1}$. Let $\delta : \Delta \to F\Delta$ be $Fd \circ \alpha \circ \pi$. Then $\pi : (\Delta, \delta) \to (A, \alpha)$ is a coalgebra morphism, since

$$F\pi \circ \delta \quad = \quad F\pi \circ (Fd \circ \alpha \circ \pi) \quad = \quad (F\pi \circ Fd) \circ \alpha \circ \pi \quad = \quad \alpha \circ \pi.$$

⊣

Lemma 3.12 *Let* $f : c \to Fc$ *and* $g : d \to Fd$ *be coalgebras, and let* $R \subseteq c \times d$ *and* $r : R \to FR$ *witness that* R *is a bisimulation. Let* $i : R \to R^{-1}$ *be* $i(x,y) = (y,x)$. *Let* $s : R^{-1} \to FR^{-1}$ *be*

$$s \quad = \quad Fi \circ r \circ i^{-1}.$$

Then S *is a bisimulation between* d *and* c.

Proof We show $g \circ \pi_1 = F\pi_1 \circ s$, and $f \circ \pi_2 = F\pi_2 \circ s$. Note that $\pi_1 \circ i = \pi_2$, and $\pi_2 \circ i = \pi_1$. And

$$\begin{array}{rcl} F\pi_1 \circ s & = & F\pi_1 \circ Fi \circ r \circ i^{-1} \\ & = & F\pi_2 \circ r \circ i^{-1} \\ & = & f \circ \pi_1 \circ i^{-1} \\ & = & f \circ \pi_2 \end{array}$$

The other equation is similar. ⊣

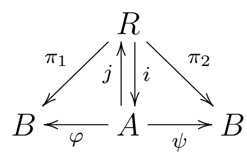
The lemma below is essentially Lemma 5.3 in Rutten's paper.

Lemma 3.13 (Generalizing Lemma 3.7) *Let (A, α) and (B, β) be F-coalgebras, and let $\varphi, \psi : A \to B$ be two coalgebra morphisms. Then the* joint image relation

$$R \quad = \quad \{(\varphi a, \psi a) : a \in A\}$$

is a bisimulation on B.

Proof Consider the diagram below:



Here $j(a) = (\varphi(a), \psi(a))$. And i is any inverse for it; i exists because j is surjective. The two triangles above commute. Let

$$\gamma \quad = \quad Fj \circ \alpha \circ i.$$

We have a coalgebra (R, γ), and we check that both projections are coalgebra morphisms.

$$\begin{array}{rcl} F\pi_1 \circ \gamma & = & F\pi_1 \circ Fj \circ \alpha \circ i \\ & = & F\varphi \circ \alpha \circ i \\ & = & \beta \circ \varphi \circ i \\ & = & \beta \circ \pi_1 \end{array}$$

Changing π_1 to π_2 shows that $F\pi_1 \circ \gamma = \beta \circ \pi_1$. ⊣

Lemma 3.14 *The union of any family $\mathfrak{F}$ of bisimulations on a coalgebra (A, α) is again a bisimulation on (A, α).*

Proof Here is a proof which uses Lemma 3.13. As we start, please remember that relations are sets of pairs. Let

$$W \quad = \quad \{(R, x, y) : R \in \mathfrak{F} \text{ and } (x, y) \in R\}.$$

So W is the disjoint union of the relations in $\mathcal{F}$. More to the point, for each $R \in \mathcal{F}$ we have $\mathsf{in}_R : R \to W$ given by $\mathsf{in}_R(x,y) = (R,x,y)$. then W is the coproduct $\coprod_{R\in\mathcal{F}} R$, and it enjoys the following universal property: if $k_R : R \to X$ is a family of maps indexed by elements of $\mathcal{F}$ and with the same codomain, then there is a unique $k : W \to X$ such that for all $R \in \mathcal{F}$, $k_R = k \circ \mathsf{in}_R$.

For each $R \in \mathcal{F}$, let $r_R : R \to FR$ witness the bisimulation condition. So $F\mathsf{in}_R \circ r_R : R \to FW$. By our universal property, let $w : W \to FW$ be such that for all $R \in \mathcal{F}$, $F\mathsf{in}_R \circ r_R = w \circ \mathsf{in}_R$.

Let $\varphi : W \to A$ be $\varphi(R,x,y) = x$. Thus $\varphi \circ \mathsf{in}_R = \pi_1$. (We should use a notation like π_1^R here, since different R will involve different relations. But we omit this for readability.) We claim that $F\varphi \circ w = \alpha \circ \varphi$ so that φ is a coalgebra morphism. For this, again we use the universal property. For each $R \in \mathcal{F}$, $F\pi_1 \circ r_R = \alpha \circ \pi_1$; this is a map from R to FA. Thus there is a unique $z : W \to FA$ such that for all $R \in \mathcal{F}$, $F\pi_1 \circ r_R = z \circ \mathsf{in}_R$. We prove that $F\varphi \circ w = \alpha \circ \varphi$ by showing that both satisfy the condition defining z. First,

$$\begin{array}{rcl} F\varphi \circ w \circ \mathsf{in}_R & = & F\varphi \circ F\mathsf{in}_R \circ r_R \\ & = & F\pi_1 \circ r_R \end{array}$$

Second,

$$\begin{array}{rcl} \alpha \circ \varphi \circ \mathsf{in}_R & = & \alpha \circ \pi_1 \\ & = & F\pi_1 \circ r_R \end{array}$$

So at this point we see that φ is a coalgebra morphism. We get a second coalgebra morphism ψ by $\varphi(R,x,y) = x$. Then using Lemma 3.13, the joint image relation of φ and ψ is a bisimulation on (A,α). This joint image relation is just $\bigcup\mathcal{F}$. In more detail, if $(x,y) \in \bigcup\mathcal{F}R$, then for some $R \in \mathcal{F}$, $(x,y) \in R$. In this case, $(R,x,y) \in W$. $\varphi(R,x,y) = x$, $\psi(R,x,y) \in R$, and so (x,y) is in the joint image. Conversely, if (x,y) is in the joint image, let (S,u,v) be such that $\varphi(S,u,v) = x$ and $\psi(S,u,v) = y$. Then $u = x$ and $v = y$. So for some $S \in \mathcal{F}$, $(x,y) \in S$. This means that $(x,y) \in \bigcup\mathcal{F}$. ⊣

Theorem 3.15 (from Rutten's paper) *Let $f : c \to Fc$ and $g : d \to Fc$ be coalgebras, and let $h : c \to d$ be any function. Then h is a morphism of coalgebras iff its graph G_h is a bisimulation between c and d, where*

$$G_h \quad = \quad \{(x,y) : h(x) = y\}.$$

Proof Let $\pi_1 : G_h \to c$ and $\pi_2 : G_h \to d$ be the projections. The key points are that π_1 is bijective (since h is a function), and $h = \pi_2 \circ \pi_1^{-1}$.

Assuming that G_h is a bisimulation, let $k : G_h \to FG_h$ be a coalgebra such that π_1 and π_2 are morphisms. To see that h is a coalgebra morphism, we need only check that π_1^{-1} is one, since then $h = \pi_2 \circ \pi_1^{-1}$ will again be one. As for π_1^{-1}, note that $F\pi_1^{-1} = (F\pi_1)^{-1}$. Then since $f \circ \pi_1 = F\pi_1 \circ k$, we have $k \circ \pi_1^{-1} = F\pi_1^{-1} \circ k$.

And assuming that h is a morphism of coalgebras, we need some k as above. We take $k = F\pi_1^{-1} \circ f \circ \pi_1$. We need to check that $f \circ \pi_1 = F\pi_1 \circ k$ and $f \circ \pi_2 = F\pi_2 \circ k$. The first of these is clear. And for the second,

$$\begin{array}{lcl} F\pi_2 \circ k & = & F\pi_2 \circ F\pi_1^{-1} \circ f \circ \pi_1 \\ & = & F(\pi_1 \circ \pi_2^{-1}) \circ f \circ \pi_1 \\ & = & Fh \circ f \circ \pi_1 \\ & = & g \circ h \circ \pi_1 \\ & = & g \circ \pi_2 \end{array}$$

This completes the proof. ⊣

This also gives another proof that the diagonal Δ_A is a bisimulation on A: it is the graph of the identity.

3.9 Functors preserving (weak) pullbacks

Definition A functor $F : C \to D$ *preserves pullbacks* if the image of every pullback square is a pullback square.

Lemma 3.16 *Concerning preservation of pullbacks:*

1. *Constant functors preserve pullbacks.*
2. *If F and G preserve pullbacks, so do $F + G$, $F \times G$, and $F \circ G$.*
3. *$\mathcal{P}$, $\mathcal{P}_{fin}$, and $\mathcal{D}$ do not preserve pullbacks.*

Proof Here is a counterexample to $\mathcal{P}_{fin}$. We use the notation from (12). Let $B = \{a, b\}$, let $C = \{1, 2\}$, and let $D = \{*\}$. Then $h : B \to D$ and $j : C \to D$ are the unique maps. The pullback A is $B \times C$ with the projections. We claim that the image of (12) under $\mathcal{P}$ is not a pullback. To see this, note that

$$\begin{array}{lcccl} \mathcal{P}\pi_1(\{(a,1),(b,2)\}) & = & B & = & \mathcal{P}\pi_1(B \times C) \\ \mathcal{P}\pi_2(\{(a,1),(b,2)\}) & = & C & = & \mathcal{P}\pi_1(B \times C) \end{array}$$

We turn this into a formal counterexample as follows. Let X be a one-point set $\{x\}$, and let $f'(x) = B$ and $g'(x) = C$. Then $\mathcal{P}h \circ f' = \mathcal{P}j \circ g'$. But there are at least two morphisms $k : X \to A$ satisfying $\mathcal{P}\pi_1 \circ k = f'$ and $\mathcal{P}\pi_2 o k = g'$, namely the morphisms determined by $\{(a,1),(b,2)\}$ and by $B \times C$. ⊣

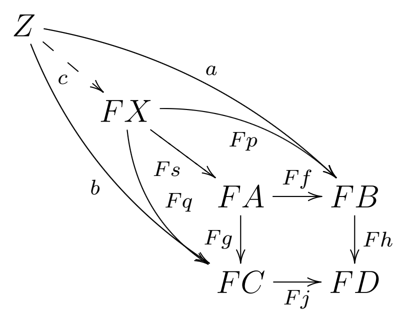
Because of the negative examples in Lemma 3.16, part 3, we need a weaker concept.

Definition A functor F *weakly preserves pullbacks* if the image of every pullback square is a weak pullback square. F *preserves weak pullbacks* if the image of every weak pullback square is a weak pullback square.

Lemma 3.17 *If F preserves pullbacks, then it weakly preserves pullbacks. F weakly preserves pullbacks iff it preserves weak pullbacks.*

Proof The first assertion is obvious. Half of the second is also easy: if F preserves weak pullbacks, then it weakly preserves pullbacks. The only thing to check is that if F weakly preserves pullbacks, then it preserves weak pullbacks. For this, consider a weak pullback as in (12). There is a pullback of h and j, say $(X, p : Xt \circ B, q : X \to C)$. The universal property of the weak pullback give us $s : X \to A$ such that $f \circ s = p$, and $g \circ s = q$.

We check that the image of (12) under F is a weak pullback. Suppose Z with a and b make the outside of the diagram below commute:



Since F weakly preserves pullbacks, the square involving FX is a weak pullback, and so there is a map c such that $Fp \circ c = a$ and $Fq \circ c = b$. The map we want is $Fs \circ c$; the verification that it does what we want is an easy diagram chase. $\dashv$

Lemma 3.18 *Concerning preservation of weak pullbacks:*

1. *Constant functors preserve weak pullbacks.*

2. *$\mathcal{P}$, $\mathcal{P}_{fin}$, and $\mathcal{D}$ also preserve weak pullbacks.*

3. *If F and G preserve weak pullbacks, so do $F + G$, $F \times G$, and $F \circ G$.*

Proof Here we only check that $\mathcal{P}$ and $\mathcal{P}_{fin}$ weakly preserve pullbacks. Consider a pullback square as in (12). We may assume that $A \subseteq B \times C$, $f = \pi_1$, and $g = \pi_2$. Suppose that we have

$$(X, f' : X \to \mathcal{P}B, g' : X \to \mathcal{P}C)$$

such that $\mathcal{P}h \circ f' = \mathcal{P}j \circ g'$. We want to define some $k : X \to \mathcal{P}A$ so that $f' = \mathcal{P}f \circ k$ and $g' = \mathcal{P}g \circ k$. We can do it by

$$k(x) \quad = \quad f'(x) \times g'(x).$$

For each $x \in X$, $\mathcal{P}\pi_1(k(x)) = f'(x)$ and $\mathcal{P}\pi_2(k(x)) = g'(x)$. $\dashv$

We shall be interested in knowing whether the image of a pullback under a certain functor F is again a pullback. That is, we would like to *assume* this about F and then develop some of the theory which we saw previously. However, this assumption that pullbacks are preserved is too strong: it fails for $\mathcal{P}_{fin}$, for example. So we shall study the weaker condition that F *preserve weak pullbacks*: the image under F of every weak pullback is again a weak pullback.

Digression: a functor which does not preserve weak pullbacks, even weak kernels Consider the *Aczel-Mendler* functor

$$F(A) \quad = \quad \{(a, b, c) \in A \times A \times A : (a = b) \text{ or } (a = c) \text{ or } (b = c)\}.$$

Note that if $(a, b, c) \in F(A)$ and $f : A \to B$, then $(fa, fb, fc) \in F(B)$. We therefore take the action of F on morphisms is by pointwise application, as expected.

This functor F does not preserve weak pullbacks. To see this, let $A = \{a, b\}$, and let $f : A \to 1$ be the unique map. The pullback of f with itself is the product $A \times A$ with its projections. We claim that the image of this pullback under F is not a weak pullback. To see this, note that

$$Ff(a, b, b) \quad = \quad (*, *, *) \quad = \quad Ff(a, a, b).$$

But there is no $x \in F(A \times A)$ such that

$$F\pi_1(x) = (a, b, b) \text{ and } F\pi_2(x) = (a, a, b).$$

The only possibility for x is $((a, a), (b, a), (b, b))$, but this triple does not belong to $F(A \times A)$.

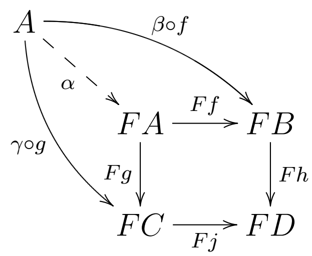
3.10 Properties of bisimulation for functors preserving weak pullbacks

Lemma 3.19 *Assume that F preserves weak pullbacks. Let $h : (B, \beta) \to (D, \delta)$ and $j : (C, \gamma) \to (D, \delta)$ be coalgebra morphisms. Suppose also that we have a pullback square in* Set *making use of h and j:*

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ {\scriptstyle g}\downarrow & & \downarrow{\scriptstyle h} \\ C & \xrightarrow[j]{} & D \end{array}$$

Then there is a coalgebra structure $\alpha : A \to FA$ such that f and g are coalgebra morphisms.

Proof Consider the diagram below:



The inside square is a weak pullback. The outside square starting at A commutes, since

$$Fh \circ \beta \circ f \quad = \quad \delta \circ h \circ f \quad = \quad \delta \circ j \circ g \quad = \quad Fj \circ \gamma \circ g$$

By the weak pullback condition, there is some $\alpha : A \to FA$ so that the top two triangles commute. This just means that α is a coalgebra morphism. $\dashv$

Lemma 3.20 *Assume that F preserves weak pullbacks. If R and S are bisimulations on a coalgebra (A, α), then $R \circ S$ is again a bisimulation on (A, α).*

Proof We have a pullback square as in (14). (In our application of the lemma, $B = C = A$.) By Lemma 3.19, there is some $\chi : X \to FX$ such that $\pi_1^X : X \to R$ and $\pi_2^X : X \to S$ are coalgebra morphisms. Further, π_1^R and π_2^S are also coalgebra morphisms. And so we have coalgebra morphisms

$$\pi_1^R \circ \pi_1^X : X \to A \quad \text{and} \quad \pi_2^S \circ \pi_2^X : X \to A.$$

The joint image is thus a bisimulation on A, by Lemma 3.13. As we know, this joint image is the relational composition $R \circ S$. $\dashv$

Lemma 3.21 *Assume that F preserves weak pullbacks. Let (A, α) be a coalgebra for F. Then the union of all bisimulations on (A, α) is an equivalence relation and a bisimulation on (A, α).*

Proposition 3.22 *Let F be the Aczel-Mendler functor from Section 3.9. There is a coalgebra (A, α) for which the largest bisimulation on F is not transitive.*

Proof Let $A = \{a, b, c\}$, and let α be given by

$$\begin{array}{rcl} \alpha(a) & = & (a, a, b) \\ \alpha(b) & = & (a, a, a) \\ \alpha(c) & = & (a, b, b) \end{array}$$

We claim that the following relation R is the largest bisimulation. We take

$$R \quad = \quad (A \times A) \setminus \{(a, c), (c, a)\},$$

so that R relates seven of the nine possible pairs. And we take $\rho : R \to FR$ to be given in the obvious way. Here are some examples:

$$\begin{array}{lcl} \rho(a,a) & = & ((a,a),(a,a),(b,b)) \\ \rho(a,b) & = & ((a,a),(a,a),(b,a)) \\ \rho(b,c) & = & ((a,a),(a,b),(a,b)) \end{array}$$

It is routine to check that R is a bisimulation. To see that it is the largest, we must check that (a,c) is not in any bisimulation. The reason is that $((a,a),(a,b),(b,b)) \notin FA$. We conclude that R really is the largest bisimulation. It is not transitive, since $a\ R\ b\ R\ c$, but $\neg(a\ R\ c)$. $\dashv$

Theorem 3.23 *Assume that F preserves weak pullbacks. Let (A,α) be a coalgebra for F, and let $R \subseteq A \times A$. Then the following are equivalent:*

1. *R is the kernel of some F-coalgebra morphism φ.*
2. *R is a bisimulation relation and an equivalence relation.*

Moreover, if (2) holds, then the natural map from A to A/R is a coalgebra morphism.

Proof (1) $\Longrightarrow$ (2): Let (B,β) be a coalgebra and $\varphi : A \to B$ be a morphism. The kernel R gives a (weak) pullback square

$$\begin{array}{ccc} R & \xrightarrow{\pi_1} & A \\ {\scriptstyle \pi_2}\downarrow & & \downarrow{\scriptstyle \varphi} \\ A & \xrightarrow[\varphi]{} & B \end{array}$$

By Lemma 3.19, there is some $r : R \to FR$ such that the projections π_1 and π_2 are coalgebra morphisms from (A,α) to (R,r). This checks that R is a coalgebraic bisimulation. Obviously it is an equivalence relation.

(2) $\Longrightarrow$ (1): Let R be a bisimulation and an equivalence relation. Consider the diagram below:

$$R \underset{\pi_2}{\overset{\pi_1}{\rightrightarrows}} A \xrightarrow{\nu} A/R$$

A/R is the quotient, and ν is the natural map into it. This ν is the *coequalizer* of π_1 and π_2. This means that $\nu \circ \pi_1 = \nu \circ \pi_2$, and that if $\mu : A \to B$ is any morphism such that $\mu \circ \pi_1 = \mu \circ \pi_2$, there is a unique $j : A/R \to B$ such that $\mu = j \circ \nu$. Now consider the following diagram:

$$\begin{array}{ccccc} R & \underset{\pi_2}{\overset{\pi_1}{\rightrightarrows}} & A & \xrightarrow{\nu} & A/R \\ & & {\scriptstyle \alpha}\downarrow & & \downarrow{\scriptstyle r} \\ & & FA & \xrightarrow[F\nu]{} & F(A/R) \end{array}$$

We get r from the coequalizer property mentioned above. It is a coalgebra morphism, and its kernel is R. $\dashv$

Lemma 3.24 (from Rutten's paper) *Assume that F preserves weak pullbacks. Let φ : $A \to B$ be a morphism of coalgebras. If R is a bisimulation on R, then $f^{-1}(B)$ is a bisimulation on A, where*

$$f^{-1}(R) \quad = \quad \{(a_1, a_2) \in A \times A : (\varphi(a_1), \varphi(a_2)) \in R\}.$$

Proof We express $f^{-1}(R)$ as a relational composition:

$$f^{-1}(R) \quad = \quad Gr(\varphi) \circ R \circ (Gr(\varphi))^{-1}.$$

We have here the graph relation of φ, R, and finally the converse of the graph relation. We use Lemmas 3.20 and 3.12, and Theorem 3.15. ⊣

Lemma 3.25 *Assume that F preserves weak pullbacks. Let (A, α) be a coalgebra for F, and let $\equiv$ be bisimilarity on A. The quotient coalgebra $A/\equiv$ is* simple*: every bisimulation on it is a subrelation of the diagonal.*

Proof Let R be a bisimulation on $A/\equiv$. By Lemma 3.24, $\nu^{-1}(R)$ is a bisimulation on A, where $\nu : A \to A/\equiv$ is the natural map. So $\nu^{-1}(R)$ is included in $\equiv$, the kernel of ν. Hence $R \subseteq \Delta_{A/\equiv}$. ⊣

3.11 Small functors

Definition A functor F on sets is *small* if there is a set M such that for all $X \neq \emptyset$,

$$FX \quad = \quad \bigcup_{f:M \to X} \text{image}(Ff).$$

Lemma 3.26 *Concerning small functors:*

1. *Constant functors are small.*
2. *$\mathcal{P}_{fin}$ and $\mathcal{D}$ are also small.*
3. *For all C and M, the functor $G_{C,M}(X) = C \times X^M$ is small.*
4. *If F and G are small, so are $F + G$ and $F \times G$.*

Proof For the functors in part 2, we use $M = N$, the set of natural numbers. Here are the details for $\mathcal{P}_{fin}$. Let $a \in \mathcal{P}_{fin}(X)$. Let $f : N \to a$ be such that its image is a, and let $s \subseteq X$ be a finite set such that $\mathcal{P}_{fin} f(s) = a$. Then $a \in \text{image}(\mathcal{P}_{fin} f)$.

Here is the verification of part 3. Let $(c, \mu) \in C \times X^M$, so that $\mu : M \to X$. $G\mu(c, id_M) = (c, \mu)$. Hence $(c, \mu) \in \text{image}(G\mu)$ for some $\mu : M \to X$.

It is easy to see that if F and G are small, so is $F + G$. The parallel verification for $F \times G$ is a bit harder, and we leave it as an exercise. ⊣

Note that in Lemma 3.26 we did not obtain the result that small functors are closed under composition. The easiest way to get this is via a separate characterization of small functors which turns out to be of independent interest.

Before we do this, we mention that the functors $G_{C,M}$ from part 3 of the lemma are about to play an important role in the rest of this section. Coalgebras for $G_{C,M}$ are deterministic automata on the alphabet A with ouputs in the set C.

Lemma 3.27 *For all C and M, there is a final coalgebra of $G_{C,M}$.*

Proof Each $G_{C,M}$ is a generalized polynomial functor, corresponding to a signature with C function symbols, each of arity M. So the carrier of the final coalgebra is the set of infinite trees whose nodes are labeled with the elements of C and whose edges are labeled with the elements of M. $\dashv$

Theorem 3.28 (V. Trnková 1969) *The following are equivalent:*

1. *F is small.*
2. *For some C and M, there is a natural transformation $\eta : G_{C,M} \to F$ such that for all non-empty X, $\eta_X : C \times X^M \to FX$ is surjective.*

Proof For (1)$\Longrightarrow$(2), let $C = FM$ and use the same M. Let $\eta : G \to F$ be $\eta(c, \mu) = F\mu(c)$. Here is why η is natural. Let $f : X \to Y$. Let $c \in FM$ and $\mu \in X^M$. Then

$$\begin{array}{rcl}(\eta_Y \circ Gf)(c, \mu) & = & \eta_Y(c, f \circ \mu) \\ & = & F(f \circ \mu)(c) \\ & = & Ff((F\mu)(c)) \\ & = & (Ff \circ \eta_X)(c, \mu)\end{array}$$

The assumption that F is small translates to the surjectivity of η_X for non-empty X.

For (2)$\Longrightarrow$(1), we know by Lemma 3.26 that $G_{C,M}$ is small. We check that this and the property of η implies that F is also small. Let $X \neq \emptyset$, and let $y \in FX$. Since η_X is surjective, let $z \in GX$ be such that $\eta_X(z) = y$. There is some $f : M \to X$ and some $a \in GM$ so that $Gf(a) = z$. By naturality, $Ff \circ \eta_M = \eta_X \circ Gf$. So

$$Ff(\eta_M(a)) \quad = \quad \eta_X \circ Gf(a) \quad = \quad \eta_X(z) \quad = \quad y.$$

Thus $y \in \text{image}(Ff)$. $\dashv$

In view of Theorem 3.28, we say that a natural transformation η is *surjective* if for all $X \neq \emptyset$, η_X is a surjective function.

Corollary 3.29 *If F and K are small, so is $F \circ K$.*

Proof Let C, M, D, and N be such that there are natural transformations $\eta : G_{C,M} \to F$ and $\rho : G_{D,N} \to K$. Now

$$(G_{C,M} \circ G_{D,N})(X) \quad = \quad C \times (D \times X^N)^M \quad \approx \quad (C \times D^N) \times X^{NM}.$$

We shall show that there is a surjective natural transformation κ from $G_{C\times D^N,NM}$ onto $F \circ K$. Let

$$\sigma \; : G_{C\times D^N,NM} \to G_{C,M} \circ G_{D,N}$$

be the evident natural transformation whose components are all bijections. Let

$$\begin{array}{rcl} \kappa & = & (\eta * \rho) \circ \sigma \\ & = & (\eta G_{C,M} \circ F\rho) \circ \sigma \end{array}$$

So κ is a natural transformation of the right type. We only need to check that for $X \neq \emptyset$, the component κ_X is surjective. First, ρ_X is surjective. Next, *every* functor F on Set preserves all surjective maps (by Lemma 2.3), so $F\rho_X$ is surjective. Also, each map $\eta_{G(X)}$ is surjectve. So overall, κ_X is a composition of three surjective maps and is thus itself surjective. ⊣

3.12 Final coalgebras for small functors preserving weak pullbacks

Theorem 3.30 *Let F be small and preserve weak pullbacks. Then F has a final coalgebra.*

Proof Fix F, and let C and M be as in Theorem 3.28. We'll write G for $G_{C,M}$. So we have $\eta : G \to F$ such that η_X is surjective for non-empty X.

Let $(A, \alpha : A \to G_{C,M}A)$ be a final coalgebra for G. We claim that every coalgebra for F has a morphism into $(A, \eta_A \circ \alpha)$. To see this, let (B, β) be a coalgebra for F. If $B = \emptyset$, then our result is trivial. Otherwise, $\eta_B : GB \to FB$ is surjective. Let $\rho : FB \to GB$ be a one-sided inverse: $\eta_B \circ \rho = id_{FB}$. Consider the following diagram:

$$\begin{array}{ccccc} B & \xrightarrow{\rho\circ\beta} & GB & \xrightarrow{\eta_B} & FB \\ \downarrow{\scriptstyle \varphi} & & \downarrow{\scriptstyle G\varphi} & & \downarrow{\scriptstyle F\varphi} \\ A & \xrightarrow[\alpha]{} & GA & \xrightarrow[\eta_A]{} & FA \end{array}$$

We get φ by finality, so the left square commutes. The right square commutes by naturality of η. Finally, the map across the top is $\eta_B \circ \rho \circ \beta = \beta$. So indeed φ is a morphism of F-coalgebras.

At this point, we know that the F-coalgebra $A = (A, \eta_A \circ \alpha)$ is *weakly final*. By Lemma 3.21, the bisimilarity relation $\equiv$ on A is an equivalence relation, and we have a quotient coalgebra $A/equiv$. The natural map $\nu : A \to A/\equiv$ is a coalgebra morphism by Theorem 3.23. Let (B, β) be an F-coalgebra. By what we showed above, there is a coalgebra morphism $\varphi : B \to A$. Thus $\nu \circ \varphi : B \to A/\equiv$ is also a coalgebra morphism. We complete the proof by checking the uniqueness of $\nu \circ \varphi$. Let $\psi : B \to A/\equiv$ be any coalgebra morphism. By Lemma 3.13, the joint image of $\nu \circ \varphi$ and ψ is a bisimulation on $A \equiv$. But $A/\equiv$ is simple by Lemma 3.25. And so the joint image is a subrelation of the diagonal relation. This means that $\nu \circ \varphi = \psi$. ⊣

Corollary 3.31 *Every functor built from the constants, $\mathcal{P}_{fin}$, and $\mathcal{D}$ using $+$, $\times$, and composition has a final coalgebra.*

Proof By Lemma 3.18, every such functor preserves weak pullbacks. By Lemma 3.26 and Corollary 3.29, every such functor is small. Now our result follows by Theorem 3.30. ⊣

The result of this section can actually be improved a fair amount. One doesn't really need the hypothesis that F preserve weak pullbacks. This would appear to be surprising given our development, but it means that the more general proof goes via different reasoning. The reasons why we did not present the better result are that (1) the work for it is somewhat specialized, and we would not need the many details of it; (2) on the other hand, some of the ideas which we saw concerning functors preserving weak pullbacks really will be used again; and (3) there are many results concerning coalgebras which do depend on some property of the functor involved, including applications of the results which we've seen. But to emphasize our point, here is a statement of the theorem:

Theorem 3.32 *Let F be small. Then F has a final coalgebra.*

4 The Foundation and Anti-Foundation Axioms

The set theoretic side of our story is connected to two axioms, the *Foundation Axiom* and the *Anti-Foundation Axiom*. We present them here, and discuss some related conceptions of set.

4.1 Background from set theory

We start with a reminder of a few basic facts of set theory. One can find more in any textbook on the subject.

Power sets For any set s, the power set of s is the set of subsets of s. We write this set as $\mathcal{P}s$.

Pairing The Kuratowski ordered pair $\langle a, b\rangle$ of two sets a and b is $\{\{a\}, \{a, b\}\}$.[9] The standard presentation of set theory defines and studies relations, functions, and the like in terms of this pairing operation. All mathematical facts about these notions can then be proved in set theory.

Natural numbers One also defines versions of the natural numbers by: $0 = \emptyset$, $1 = \{\emptyset\}$, etc. Again, all facts about numbers and functions on them can be proved in set theory. In fact, essentially all mathematical facts whatsoever can be stated formally and proved in set theory.

[9] In these notes, we sometimes write $\langle a, b\rangle$ as (a, b), especially in connection with structures like graphs $(G, \to)$.

Union and transitive closure For any set a, $\bigcup a$ is the set of elements of elements of a. A set is *transitive* if every element of it is also a subset of it. The *transitive closure* of a is $a \cup \bigcup a \cup \bigcup \bigcup a \cup \cdots$. This set is denoted $tc(a)$. It is the smallest transitive set which includes a as a subset.

Theorem 4.1 (Cantor) *For all sets s, and all functions $f : s \to \mathcal{P}s$, f is not surjective. In fact, $\{x \in s : x \notin f(x)\}$ is not in the image set $f[s]$.*

Proof Let $c = \{x \in s : x \notin f(x)\}$. Suppose towards a contradiction that $c \in f[s]$. Fix $a \in s$ such that $c = f(a)$. Then $a \in c$ iff $a \notin f(a)$ iff $a \notin c$. $\dashv$

Corollary 4.2 *For all sets s, $\mathcal{P}s$ is not a subset of s.*

Proof If $\mathcal{P}s \subseteq s$, we construct a function f from s onto $\mathcal{P}s$: let $f(a) = a$ if $a \in s$, and otherwise let $f(a) = \emptyset$. So we cannot have $\mathcal{P}s \subseteq s$, lest we contradict Cantor's Theorem. $\dashv$

Corollary 4.3 (Russell's Paradox) *There is no set R such that every set belongs to R.*

Proof Such a set would have $\mathcal{P}a \subseteq R$ for all sets a. In particular $\mathcal{P}R \subseteq R$, contradicting our last result. $\dashv$

We call the last result *Russell's Paradox* in view of its content. Neither our statement nor our proof are the most standard ones.

Well-ordered sets and ordinal numbers We need the concept of ordinal numbers at a few places.

A *well-ordered set* is a pair $W = (W, <)$ where $<$ is a relation on W which is a strict linear order and with the property that every non-empty subset of W has a $<$-least element. For example, $(N, \prec)$ is a well-order, where $\prec$ is given by

$$0 \prec 2 \prec 4 \prec \cdots 1 \prec 3 \prec 5 \prec \cdots \tag{23}$$

One can show using the Replacement Axiom that every well-ordered set W has a unique decoration d. An *ordinal number* (or *ordinal*) is a set of the form $d(w)$, for some well-ordered set $(W, <)$ and some $w \in W$.

One usually uses Greek letters such as α and β for ordinal numbers, and one also writes $\alpha < \beta$ if $\alpha \in \beta$. There are a number of standard facts about ordinal numbers, including the following:

1. If α is an ordinal, so is $\alpha \cup \{\alpha\}$.

2. The standard modeling of natural numbers renders them as ordinals, and the set ω of natural numbers is also an ordinal.

3. The collection of ordinal numbers is not a set.

4. Except for not being a set, the ordinal numbers with $<$ has all the properties of a well-ordered set.

An ordinal α is a *successor ordinal* if $\alpha = \beta \cup \{\beta\}$ for some (other) ordinal β. Ordinals which are neither 0 nor successor ordinals are called *limit ordinals*. The smallest limit ordinal is ω; it is $d(1)$ for the well-order indicated in equation (23).

The cumulative hierarchy There is a unique operation $\alpha \mapsto V_\alpha$ taking ordinals to sets such that

$$\begin{array}{lcl} V_0 & = & \emptyset \\ V_{\alpha+1} & = & \mathcal{P}(V_\alpha) \\ V_\lambda & = & \bigcup\{V_\beta : \beta < \lambda\} \quad \text{for } \lambda \text{ a limit ordinal} \end{array}$$

The *ZFC* axioms We are not going to state them here, but see any book on set theory. The Z is for Ernst Zermelo, the first to list these axioms. The F is for Abraham Fraenkel, the one who formulated the Replacement Axioms and the Foundation Axiom.

Classes The axioms of set theory are not about sets as much as they are about the *universe of sets*. One of the intuitive principles of the theory is that arbitrary collections of mathematical objects "should be" sets. Due to paradoxes, this intuitive principle is not directly formalized in standard set theories. In a sense, the axioms one does have are intended to give enough sets to constitute a mathematical universe while not having so many as to risk inconsistency. But it is natural in this connection to consider some collections of objects which are demonstrably not sets. These are called *proper classes*. The term *class* informally refers to a collection of mathematical objects. Classes are usually not first-class objects in set theory. (Certainly they are not in the most standard set theory, *ZFC*. One source for alternative axiomatic theories of sets is the SEP entry on this. It mentions quite a few theories which treat classes as first-class objects.) Instead, a statement about classes is regarded as a paraphrase for some other (more complicated and usually less intuitive) statement about sets. This is probably not a good place to discuss the details of the formalization; one useful source is Chapter 1 of Azriel Levy's book [17] on set theory.

For our purposes, classes may be taken as *definable subcollections of the universe of sets*. For example, if a is any set, then the class of all sets which do not contain a as an element is $\{x : a \notin x\}$. In specifying a class, one may use the first-order language with the membership symbol and the rest of the syntax from logic, and one may also use particular sets as parameters, as we have just done.

The class V of all sets is $\{x : x = x\}$. The definability here is in the first-order logic with just a symbol $\in$ for membership, and the quantifiers range over sets (not classes). Another class of interest is WF, the class of all well-founded sets. This is the same as $\bigcup_\alpha V_\alpha$, the sets that belong to V_α for some α.

If C is a class, we define the *power class of* C, $\mathcal{P}C$ by

$$\mathcal{P}C \quad = \quad \{x : (\forall y)(y \in x \to \varphi_C(y)\},$$

where φ_C is the formula that defines the class C. It is important that in this definition x ranges over *sets* and not classes; the formal language used does not directly talk about classes in the first place. For example, $\mathcal{P}V = V$ and $\mathcal{P}(WF) = WF$. We also define the action of other operations on classes in the same general way. For example, the finite power set $\mathcal{P}_{fin}$ takes a class C to the class of finite subsets of C.

4.2 The Foundation Axiom

The Foundation Axiom (*FA*) may be stated in different ways. Here are some formulations; their equivalence in the presence of the other axioms is a standard result of elementary set theory.

1. There are no infinite sequences of sets

$$x_0 \ni \; x_1 \ni x_2 \ni \cdots \ni x_n \ni x_{n+1} \ni \cdots$$

each of whose terms is an element of the previous term.

2. For every non-empty set x, there is some $y \in x$ such that $y \cap x = \emptyset$.

3. Let a be any set, and let $f \; : \mathcal{P}(a) \to a$. Let b be any transitive set. Then there is a unique function $g \; : b \to a$ so that for all $x \in b$,

$$g(x) \quad = \quad f(g[x]).$$

4. For every set x, there is an ordinal number α such that $x \in V_\alpha$.

5. $V = WF$.

The first of these is probably the easiest to remember and think about. The second is important because it is the one most easily expressed in first-order logic. The third is a recursion principle; we shall consider a closely related principle in Section 2.9.

The iterative conception of set As we have seen, one formulation of *FA* says that every set belongs to some V_α. This is a mathematical formulation of the *iterative concept of set*: sets are just what one gets by iterating the power set operation on the well-ordered class of ordinal numbers. We start with nothing, the empty set.[10] This is V_0. Then we form $V_1 = \mathcal{P}V_0$. Then $V_2 = \mathcal{P}V_1$. Going on, when we come to ω, we take the union of all the

[10]Perhaps the similarity of "starting from nothing", *creatio ex nihilo*, and the Big Bang are part of the appeal of the iterative conception.

sets V_n from earlier. This is V_ω. Then we proceed to $V_{\omega+1} = \mathcal{P}V_\omega$. We continue like this absolutely forever, going through "all the ordinal numbers". The collection so described is the universe V of sets.

This way of describing the iterative picture suggests that the ordinal numbers were somehow present "before" all the iteration takes place, or at least that they have a life apart from the rest of the sets. There is a different way of understanding the iterative conception, one that emphasizes the harmony between the iteration of the power set operation and the Replacement Axiom: as one iterates the power set axiom, more and more well-ordered sets appear. Replacement allows us to decorate these well-ordered sets, creating new ordinals in the process. Thus the whole picture is one of balance. Indeed, this point about balance can be phrased without reference to any "iteration" at all: there is an equilibrium in the set theoretic universe between the "sideways" push of the Power Set Axiom and the "upward" push of the Replacement Axiom.[11]

Using the Foundation Axiom *FA* plays no role in the formalization of mathematics or in the study of infinity. It is an "optional extra" for mathematics. For conceptual purposes, *FA* is used to clarify our picture of sets, just as we have described. This often comes with an implicit argument of roughly the following shape:

An argument One is tempted to justify *FA* along the following lines:

1. Russell's Paradox shows that there is no set of all sets: the class V cannot be a member of itself.

2. The iterative picture shows why *no* set can be a member of itself.

3. The iterative picture also suggests *FA* in the form that every set belongs to some V_α.

4. Thus *FA* reflects a picture which avoids Russell's Paradox, hence it is sensible to accept it.

The rejoinder here is that there might be other intuitive pictures or conceptions of sets that also explain, or draw lessons from, the paradoxes. So they would be as sensible as *FA* in this regard.

Since *FA* plays a conceptual role but no mathematical role, it is not surprising that there are widely different views on whether it is an important part of standard set theory *ZFC* or not. For a collection of quotes on the role of *FA*, see [7].

[11] I realize that my presentation of the iterative conception does mix temporal and spatial metaphors. In this connection, I mention from John Burgess' review [10] of papers (including Boolos [9]: "The informal description of this [iterative or cumulative hierarchy] almost demands resort to a metaphor, spatial imagery being preferable to temporal as being less likely to be (mis)taken literally." A criticim of the iterative conception and specifically of the metaphors used may be found in Parsons [22].

The Foundation Axiom and object circularity We mentioned in connection with streams that according to standard set theory, streams of numbers do not exist. Here is the reasoning. Recall that we defined a stream to be a pair of a number and another stream. Suppose that a stream s exists, so that the set N^∞ of streams is non-empty. Recall that we have a function $f_s : N \to N^\infty$ by recursion

$$\begin{array}{lcl} f_s(0) & = & s \\ f_s(n+1) & = & \mathsf{tail}(f_s(n)) \end{array}$$

To save on some notation at this point, let's write h_n for $\mathsf{head}(f_s(n))$ and t_n for $\mathsf{tail}(f_s(n))$. For all n,

$$f_s(n) = \langle h_n, t_n \rangle = \{\{h_n\}, \{h_n, t_n\}\};$$

this is true of any pair whatsoever. Notice that

$$f_s(n+1) = t_n \in \{h_n, t_n\} \in f_s(n).$$

So we have

$$f_s(0) \ni \{h_0, t_0\} \ni f_s(1) \ni \{h_1, t_1\} \ni f_s(2) \ni \cdots$$

This is a descending sequence in the membership relation, something forbidden by *FA*.

The same kind of remark applies to infinite trees as we discussed them, and certainly to hypersets. The conclusion is that if one wants to work with such objects in a set theory with *FA*, then one must do so indirectly.

4.3 The Anti-Foundation Axiom

The Anti-Foundation Axiom *AFA* is stated as follows:

$$\text{Every graph has a unique decoration} \tag{24}$$

The theory *ZFA* is *ZFC* with *FA* replaced by *AFA*. It includes the Axiom of Choice, even though there is no "*C*" in the acronym.

The coiterative conception of set *AFA* gives rise to, or reflects, a conception of set that is at odds with the iterative conception. For lack of a better name, we call it the *coiterative conception*. According to this, a set is an abstract structure obtained by taking a graph G (a set with a relation on it), and then associating to each node x in the graph a set in such a way that the set associated to x is the set of sets associated to the children of x in G. This association is what we called *decoration* earlier. This association might be thought of procedurally, but it need not be so construed. One can instead posit a harmony between decoration and power sets.[12]

[12] As an interesting aside, note that the ordinals can be obtained via decoration of well-ordered sets. In *ZF*, one needs the Replacement Axioms for this.

What changes with *AFA*, and what does not change? *AFA* gives us unique solutions to systems of set systems; this is almost immediate from the axiom and the close relation of set systems and graphs. But it also gives us unique solutions for stream systems and tree systems. The details of this are suggested by our work on the decoration of the graph in (9).

All of the results in set theory which do not use *FA* go through when one replaces it by *AFA*. In particular, the following topics are unchanged: Russell's Paradox and the Separation (Subset) Axioms; the modeling of ordered pairs, relations and functions; the natural numbers, real numbers, etc.; well-orderings and the ordinal numbers; transfinite recursion on well-orders and well-founded relations; the Axiom of Choice; problems and results concerning the sizes of infinite sets. The only difference would be in modeling questions for circularly defined objects of various sorts, as we have been discussing them.

In terms of modeling circularity, *AFA* gives several new concepts and techniques. These are described in our next section.

5 Using *AFA*

This section offers a quick introduction to the central parts of the theory non-wellfounded sets: what one would need to know to use the theory and to read papers on it.

5.1 Bisimulation

The topic of *bisimulation* is one of the earliest goals in a treatment of non-welfounded sets.

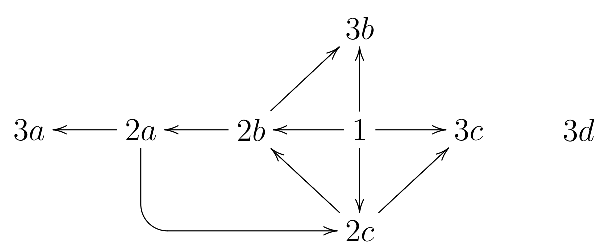
Let $(G, \to)$ be a graph. A relation R on G is a *bisimulation* iff the following holds: whenever xRy,

a. If $x \to x'$, then there is some $y \to y'$ such that $x'Ry'$.

b. If $y \to y'$, then there is some $x \to x'$ such that $x'Ry'$.

Bisimulation between graphs Before giving examples, we should clarify some usage. At a few points, we'll speak of bisimulation *between* two graphs G and H, rather than *on* a single graph. This can be defined in the same general way. Note also that one can take the *disjoint union* $G + H$ of the graphs G and H, and then a bisimulation between G and H would be a bisimulation on $G + H$.

Returning to bisimulation on a graph For an example, let's look at the following graph G:



All of the 3-points have no children. (Point $3d$ is not reached from any other point, but the arrows *into* a node are of no interest.) So every relation which only relates 3-points is a bisimulation on G. Concretely,

$$\{(3a, 3b), (3c, 3a), (3d, 3d)\}$$

is easily seen to be a bisimulation.

For that matter, the empty relation is also a bisimulation on G.

Another bisumuation is

$$\{(2a, 2b), (2b, 2c), (2c, 2a)\} \cup \{(3a, 3b), (3b, 3c), (3c, 3a)\}.$$

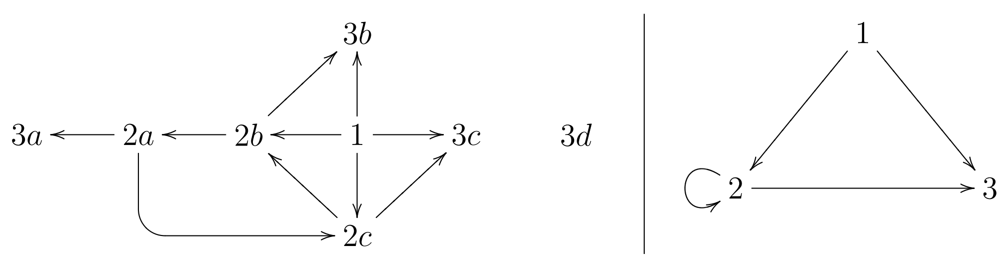
Let's call this relation R. It would take a lot of checking to actually verify that R is a bisimulation. Here is just two items of it: We see that $2b\ R\ 2c$. Now $2c \to 2b$. Thus we need some node x so that $x\ R\ 2b$ and $2b \to x$. For this, we take $2a$. For our second point of verification, again note that $2b\ R\ 2c$. Since $2b \to 3b$, we need some node x so that $2c \to x$ and $3b\ R\ x$. We take $x = 3c$ for this.

The largest bisimulation on our graph G is the relation that relates 1 to itself, all 2-points to all 2-points, and all 3-points to all 3-points. Note that this is an equivalence relation: reflexive, symmetric, and transitive. This is not an accident.

Proposition 5.1 *For any graph H, there is a largest bisimulation on H. This relation is an equivalence relation called bisimilarity and denoted $\equiv_b$. It is also characterized by*

$$x \equiv_b y \quad \textit{iff} \quad \textit{there is a bisimulation on } H \textit{ relating } x \textit{ to } y.$$

We can always form the *quotient graph* using the largest bisimulation. Here is how this works, using G from above as an example. In $G/\equiv_b$, we would have three nodes, corresponding to the three equivalence classes under the largest bisimulation; let's call these 1, 2 and 3. We put an arrow between two equivalence classes if some (every) element of the first has an arrow to some element of the second. In this way, we contruct the quotient. Here is a picture of G again, along with its quotient $G/\equiv_b$ under the largest bisimulation:



The map from G to $G/\equiv_b$ takes the 2-points to 2 and the 3-points to 3.

Up until now, we have said what bisimulation is, but we did not describe its relation to anything else. To rectify matters, here is the main result.

Theorem 5.2 *Assume AFA. Let G be a graph, let x and y belong to G, and let d be the decoration of G. Then the following are equivalent.*

1. $d(x) = d(y)$.
2. *There is a bisimulation relating x and y.*

We are not going to prove this theorem in full here, but instead here are two hints. To prove that (1) implies (2), check that the *kernel relation of* d,

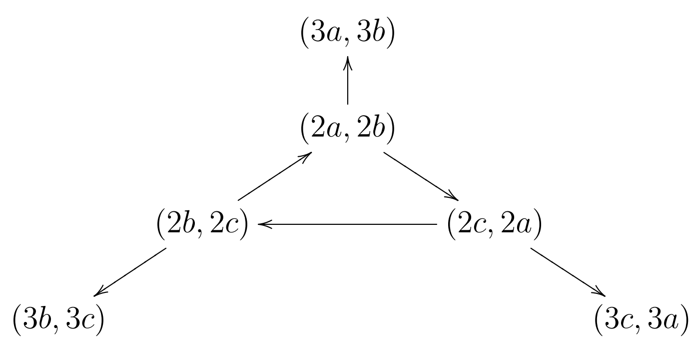
$$\{(u, v) : d(u) = d(v)\}$$

is a bisimulation on G.

In the other direction, the idea is to turn a bisimulation into a graph itself, and then extract two decorations of it; by the uniqueness part of *AFA* these must coincide. Here is how this is done in a concrete example. We saw above that

$$R \quad = \quad \{(2a, 2b), (2b, 2c), (2c, 2a)\} \cup \{(3a, 3b), (3b, 3c), (3c, 3a)\}.$$

is a bisimulation. We make it into a graph by taking the product relation. This gives the following graph which we call H:



Let d be a decoration (no, *the* decoration) of G. We get two decorations of H: $k(u, v) = d(u)$, and $l_2(u, v) = d(v)$. (It is good to check that these really are decorations of H.) But H can have only one decoration. So $k = l$. And then, corresponding to the fact that $2a \; R \; 2b$, for example, we have

$$d(2a) \quad = \quad k(2a, 2b) \quad = \quad l(2a, 2b) \quad = \quad d(2b).$$

5.2 Doing without *AFA*

Our work on bisimulation above can be used to effect a reduction of the of non-wellfounded sets to that of ordinary sets, much in the spirit of what we saw for streams and functions in Section 1.1. There are several ways to describe such a reduction.

A *pointed graph* is a triple $(G, \to, g)$ such that $\to$ is a relation on G and $g \in G$. A *bisimulation* between pointed graphs $(G, \to, g)$ and $(H, \leadsto, h)$ is a bisimulation between R between $(G, \to)$ and $(H, \leadsto)$ such that $g \; R \; h$.

In the remainder of this discussion, we let p, q, ... denote pointed graphs. We write $p \equiv q$ if there is a bisimulation between p and q. We also write $p \, \epsilon \, q$ if there is a pointed graph $(G, \to, g)$ and some $g \to h$ in G so that

$$p \equiv (G, \to, g) \qquad \text{and} \qquad q \equiv (G, \to, h).$$

Sentences in the language of set theory talk about sets, and we translate them to sentences about pointed graphs by restricting all quantifiers to the class of pointed graphs, and then translating $\in$ to ϵ, and $\equiv$ for $=$. For example, the Axiom of Extensionality

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

would translate to

$$(\forall \text{ pointed graphs } p, q)(p \equiv q \leftrightarrow (\forall \text{ pointed graphs } r)(r \; \epsilon \; p \leftrightarrow r \; \epsilon \; q)).$$

This last sentence is then provable. (Hint: the union of any set of bisimulations on a graph is again a bisimulation on it.)

Indeed, all of the axioms of *ZFA* are provable, including *AFA*. This is a fairly long and tedious, but not so tricky. A version of it (for set theory with *urelements*, objects which are not sets) is the topic of a chapter in Barwise and Moss [8].

One can also go further: instead of translating the identity relation = into something more complex, we may keep the language simple and complicate the interpretation. We would like to replace "pointed graph" by "$\equiv$-class of pointed graph". Since these are not sets, we instead employ *Scott's trick* and instead use "set of well-founded pointed graphs whose node sets are $\equiv$-equivalent and with the property that no pointed graph of smaller rank is also $\equiv$-equivalent to them."

Doing all of this leads to relative consistency result:

Theorem 5.3 *If ZFC is consistent, then so is ZFA, and vice-versa.*

5.3 Extended graphs

The way we have presented graphs, decorations, and *AFA* is a very "minimalist" presentation. If one would like some node of a graph G to be decorated by some set a, the most obvious way would be to add all the elements of $tc(\{x\})$ as fresh nodes in G, with $y \to z$ iff $z \in a$. This means that one must take new copies if some of the sets in $tc(\{x\})$ already happen to be nodes in G. This is often cumbersome: when working with graphs and decorations, one might well want to *pre-specify* as much as possible the value of the decoration on a node. There are several ways to do this with *AFA*, and we'll indicate one here.

Two sets are *disjoint* if their intersection is empty. When one takes the union of two sets, say a and b, it is sometimes a good idea to make sure that no element occurs in both sets. The way to do this is to replace one or both of a and b by copies.

The *disjoint* union of sets a and b is $a + b$, defined by

$$a + b \quad = \quad (\{0\} \times a) \cup (\{1\} \times b).$$

It is easy to see that the two sets in the union are disjoint: the elements of $a + b$ "wear on their sleeve" a mark of which set they come from.

The disjoint union comes with two natural functions:

$$\mathsf{inl} : a \to a + b \qquad \mathsf{inr} : b \to a + b$$

defined by $\mathsf{inl}(x) = \langle x, 0\rangle$ and $\mathsf{inr}(x) = \langle x, 1\rangle$.[13]

A *graph extended with set parameters* (or *extended graph* for short) is a set G together with a function $e : G \to \mathcal{P}G + V$. If $e(g)$ is of the form $\langle 0, s\rangle$ for some $s \subseteq G$, then we think of it as a node in the same way we did so earlier. In particular, we want to decorate it with the set of decorations of its children. If $e(g) = \langle 1, x\rangle$, then we want a decoration to be forced to have the value x on g.

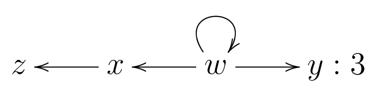
Formally, a decoration of an extended graph is a function d defined on G so that

$$d(g) \quad = \quad \begin{cases} \{d(h) : h \in s\} & \text{if } e(g) = \langle 0, s\rangle \\ x & \text{if } e(g) = \langle 1, x\rangle \end{cases}$$

Here is an example: Let G be the extended graph with node set $\{w, x, y, z\}$ and with e given by

$$e(w) = \langle 0, \{w, x, y\}\rangle, \quad e(x) = \langle 0, \{z\}\rangle, \quad e(z) = \langle 0, \emptyset\rangle, \quad e(y) = \langle 1, 2\rangle$$

It is natural to draw this as



Then a decoration d of this extended graph would satisfy the following conditions:

$$\begin{array}{rclcrcl} d(z) & = & 0 & \qquad & d(y) & = & 3 \\ d(x) & = & 1 & & d(w) & = & \{d(w), 1, 3\} \end{array}$$

Theorem 5.4 *Assume AFA. Then every extended graph has a unique decoration.*

The point we are trying to make is that there is quite a bit of theory around to facilitate working in *ZFA* in order to do modeling of various forms of circular phenomena.

[13] The "in" stands for *injection*, and the "l" and "r" are from *left* and *right*.

5.4 Collection circularity in *ZFA*

The fact that *AFA* allows us to solve various kinds of systems of set equations is only the beginning. When we discussed infinite trees in Section 1.2, we noted that the collection Tr of infinite trees *should* satisfy

$$Tr \quad = \quad \{x, y\} \cup (\{\bullet\} \times Tr) \cup (\{*\} \times Tr \times Tr). \tag{25}$$

A similar equation should hold of streams from Section 1.1:

$$N^{\infty} \quad = \quad N \times N^{\infty}. \tag{26}$$

For that matter, the universe V of sets should satisfy

$$V \quad = \quad \mathcal{P}V \tag{27}$$

Assuming the power set axiom and the formulation of $\mathcal{P}$ as an operator on classes, the universe V does satisfy (27).

We are free to step back and think of these as equations which we hope to solve. For example, we could take the set N as known, regard N^{∞} as a variable, and then consider an equation like $x = N \times x$. However, the none of these is the kind of equation that we could hope to solve in a perfectly general way using *AFA*: the right-hand sides are not given in terms of sets of objects on the left. Solving more complicated systems takes special additional work. Here is what is known on this.

First, under *FA*, (25) and (26) have a unique solution: the empty set. Under *AFA*, they have many solutions. For example, for the stream equation, the set of streams corresponding to functions which are eventually 0 is a solution. However, the *largest* solutions are of special interest. For these, one can prove that the largest solutions are in one-to-one correspondence with what we have called the unraveled forms. And for other reasons which we shall see, there is a good reason to accept the claim that the largest solutions are good mathematical models of the intuitive concepts.

Under *AFA*, things are different. Here is the general picture: An operator on sets F is *monotone* if whenever $a \subseteq b$, then also $Fa \subseteq Fb$. This is a very common feature for operators on sets. The *polynomial operators on sets* are the smallest collection containing the constant operators, and closed under cartesian product, disjoint union, and functions from a fixed set. For example, if A and B are fixed sets, then $Fs = (A \times s) + {}^{B}(s + A)$ is a polynomial operator on sets. If one also allows the power set operator to occur, then we get the *power polynomial operators*. Every power polynomial operator is monotone. And now, we have the following results of Aczel:

Proposition 5.5 *Then monotone operator F on sets has a least fixed point F_* and a greatest fixed point F^*. In particular, every polynomial operator on classes has least and greatest fixed points. On classes, the same is true for the larger collection of power polynomial operators.*

Assuming FA, the fixed points are unique; frequently they are the empty set. With AFA, the greatest fixed points usually have non-wellfounded members. We shall study this in more detail when we turn to coalgebra. For now, we return to the last of the example equations at the top of this section, (27). This equation has no solutions in sets due to Cantor's Theorem 4.1. However, in terms of classes, this equation does have solutions, as we know. The universal class V is a solution, as we have seen. And the class WF of well-founded sets is a solution. This is the smallest solution $\mathcal{P}_*$, and V is the largest. Under FA, $\mathcal{P}^* = V = \mathcal{P}_*$. Under AFA, $\mathcal{P}_*$ and $\mathcal{P}^*$ are different: $\mathcal{P}_* = WF$, and $\mathcal{P}^* = V$ and thus contains sets such as $\Omega = \{\Omega\}$.

6 Additional related modeling of circularity

6.1 Universal Harsanyi type spaces and coalgebraic modal logic

We mentioned in Section 1.4 the problem of finding *universal type spaces*. We point out here that this amounts to finding a final coalgebra for a certain endofunctor on a certain category. Then we discuss what is known about this matter. Finally, we have some conceptual points about this matter.

The first relevant category is Meas, the category of measurable spaces and measurable functions between them. We have seen the definitions in Section 1.4. We also saw the definition of $\Delta(M)$, the set of probability measures on the space M. What we did not see is the way Δ acts on morphisms to give a functor. If $f : M \to N$ is measurable, then for $\mu \in \Delta(M)$ and $A \in \Sigma'$, $(\Delta f)(\mu)(A) = \mu(f^{-1}(A))$. That is, $(\Delta f)(\mu) = \mu \circ f^{-1}$.

The category Meas has products and coproducts. The product space is the categorical product, and the coproduct is given by disjoint unions. However, Meas does not preserve weak pullbacks, and (worse) it does not preserve ω^{op} limits [27].

The class of *measure polynomial functors* is the smallest class of functors on Meas containing the identity, the constant spaces M and closed under products, coproducts, and Δ.

Theorem 6.1 *Every measure polynomial functor has a final coalgebra.*

This result is not quite what we want in the formalization of universal type spaces. For one thing, there are no "players" around. We'll discuss matters for the case of two players. We repeat our earlier definition of a two-player type space: it is a tuple (M, σ, N, τ), where M and N are measurable spaces, and

$$\begin{aligned} \sigma &: M \to \Delta(S \times N) \\ \tau &: N \to \Delta(S \times M) \end{aligned}$$

Now we first of all want to reformulate this in terms of coalgebras. We fix S and take as our category C the category of pairs (M, N) of measurable spaces, with a morphism from (M_1, N_1) to (M_2, N_2) just being a pair of morphsims (f, g), where $f : M_1 \to M_2$ and $g : N_1 \to N_2$. We have an endofunctor $\Delta : C \to C$ given by

$$\begin{array}{lcl} \Delta(M, N) & = & (\Delta(S \times N), \Delta(S \times M)) \\ \Delta(f, g) & = & (\Delta(1_S \times g), \Delta(1_S \times f)) \end{array}$$

In these terms, we can say that a type space is a coalgebra of Δ on C. Further, a morphism

There are some clear conceptual clues that coalgebra could be connected to type spaces. The first is that the notion of "belief" in the game theory literature is typically a probabilistic one. If we replace "belief" with "knowledge" above, then we have a very-well-studied notion, the formalization of knowledge by *possible worlds semantics*. The mathematical structures for possible worlds semantics are sets W of *worlds* with two functions, one giving for each world $w \in W$ some set of "atomic propositions" true at w, and the other giving for each w some set of worlds which are said to be "possible from w". These structures are essentially the coalgebras in the category Set of sets of the functor $F(W) = A \times \mathcal{P}(W)$, where A is the power set of the set of atomic propositions and $\mathcal{P}$ is the power set functor on sets. Perhaps the primary contribution of coalgebra to this area to date is to show that *modal logic*, the natural logical language for the structures, generalizes to *coalgebraic versions of modal logic*. We'll return to this point shortly.

The second clue has to do with the role of the universal type space in this field. Types are taken to be elements of the universal type space. In the universal type space all possible types are uniquely represented, and the idea is that two types with exactly the same beliefs about the underlying world of "nature" plus the types of the other players are taken to be undistinguishable. This is the same ideology as we find concerning *final coalgebras*, in which (as is well known) points with the same behavior are identified.

6.2 Fractal sets and completely iterative algebras

Completely iterative algebras Let F be an endofunctor on a category C with coproducts. A *completely iterative algebra* (*cia*) for F is an algebra (A, a) for F such that for all $e : X \to FX + A$, there is a unique $e^\dagger : X \to A$ such that

$$\begin{array}{ccc} X & \xrightarrow{e} & FX + A \\ {\scriptstyle e^\dagger}\downarrow & & \downarrow{\scriptstyle Fe^\dagger + id_A} \\ A & \xleftarrow[{[a, id_A]}]{} & FA + A \end{array}$$

There are two main kinds of examples for us. The first is from set theory. We take C to be Class, F to be $\mathcal{P}$, and (A, a) to be (V, i). (Here, as earlier, $i : \mathcal{P}V \to V$ is the identity, regarding a set of sets as a set.) Assuming AFA, we have a cia. This is the content of Theorem 5.4: what we called an extended graph corresponds to a function of the form $e : G \to \mathcal{P}G + V$, and a decoration corresponds to $e^\dagger$.

A *metric space* is a pair (X, d) where X is a set, and $d : X \times X \to R$ has the following distance-like properties:

1. $d(x, y) = 0$ iff $x = y$.
2. $d(x, y) = d(y, x)$.
3. $d(x, z) \leq d(x, y) + d(x, z)$.

A sequence $x_0, x_1, x_2, \ldots$ *converges to* y if for every $\epsilon > 0$ there is some N such that for all $n > N$, $d(x_n, y) < \epsilon$. A *Cauchy sequence* in X is a sequence of points with the property that for every $\epsilon > 0$ there is some N such that for $m, n > N$, $d(x_m, x_n) < \epsilon$. A metric space is *complete* if every Cauchy sequence in it has a limit; this limit is then unique.
Let (X, d) be a complete metric space. Consider the set $C(X)$ of all non-empty compact subspaces of X together with the *Hausdorff metric* h; for two compact subsets A and B of X

$$h(A, B) = \max\{\, d(A \to B), d(B \to A)\,\},$$

where $d(A \to B) = \max_{a \in A} \min_{b \in B} d(a, b)$. $(C(X), h)$ is then a complete metric space.

Figure 4: Supplement on definitions concerning metric spaces

The fact that the universe (V, i) is a cia extends in two ways. First, every final coalgebra of any functor on any category is a cia, provided the category has a coproduct operation $+$. And returning to Class, the universe (V, i) is a cia of many other functors, including all of the power polynomials.

It is the second kind of cia examples which connect our topic to fractal sets. For the definitions, you may wish to consult the supplement on definitions concerning metric spaces.

CMS is the category of complete metric spaces with distances measured in the interval $[0, 1]$ together with maps $f : X \to Y$ such that $d_Y(fx, fy) \leq d_X(x, y)$ for all $x, y \in X$. These maps are called *non-expanding*. A stronger condition is that f be ϵ-*contracting*: for some $\epsilon < 1$ we have that

$$d_Y(fx, fy) \leq \epsilon \cdot d_X(x, y)$$

for all $x, y \in X$. If $f, g : X \to Y$, we measure the distance between f and g by taking the supremum of $d_Y(f(x), g(x))$ as x ranges over X. We write this as $d(f, g)$. A functor F on CMS is called ϵ-*contracting* if there exists a constant $\epsilon < 1$ such that for all non-expanding maps $f, g : X \to Y$, $d(Ff, Fg) \leq \epsilon \cdot d(f, g)$.

Theorem 6.2 (Adámek and Reitermann [6]) *Let F be a contracting endofunctor on the category* CMS *of complete metric spaces. Then any non-empty F-algebra (A, a) is completely iterative.*

An example Here is an example which is mathematically suggestive. We use the cia property to show that every system of equations like

$$\begin{array}{lcl} x_1 & \approx & .5x_2 \\ x_2 & \approx & .5x_3 + .5 \\ x_3 & \approx & .5x_4 + .5 \\ x_4 & \approx & .5x_5 \\ & \cdots & \end{array} \tag{28}$$

has a *unique* solution in the unit interval of real numbers $I = [0, 1]$. On the right, one either has .5 times some variable, or else .5 times some variable with .5 added on. The pattern may be arbitrary. The set of variables may be infinite, indeed of arbitrary size.

Let F be the functor on CMS taking a space (X, d) to $(X+X, d^*)$, where d^* measures distances in the same copy of X by taking half the corresponding distance in (X, d); otherwise, $d^*(x, y) = 1$.

We use the following example (I, a) of an algebra for F: $I = [0, 1]$, and $a : FI \to I$ is given by $a(\mathsf{inl}(x)) = .5x$, and $a(\mathsf{inr}(x)) = .5x + .5$.

So (I, a) is a cia for F. We use this, taking X to be our set of variables, made into a metric space by setting $d(x_i, x_j) = 1$ for $i \neq j$. We define $e : X \to FX + I$ as suggested from our example above: $e(x_1) = \mathsf{inl}(\mathsf{inl}(x_2))$, $e(x_2) = \mathsf{inl}(\mathsf{inr}(x_3))$, etc. This particular system does not use the I summand, and the inr and inl just code whether the "+.5" appears on the right of the equation.

Then the cia property gives a unique solution map $e^\dagger : X \to I$. Tracing through the definitions of the functions involved shows that this is the desired solution to (28).

The Cantor set, again Let I be the unit interval $[0, 1]$, an object in CMS. The Cantor "middle third" set c may be obtained via the cia structure on $C(I)$, the space of compact subsets of I with a metric as defined in our supplement on metric spaces.

We take for $F : \mathsf{CMS} \to \mathsf{CMS}$ the functor which takes a metric space (X, d) to $(X, \frac{1}{3}d)$. This is $\frac{1}{3}$-contracting, and we use the non-empty F-algebra $(C(I), \alpha)$, where

$$\alpha(S) \quad = \quad \frac{1}{3}S \cup (\frac{1}{3}S + \frac{2}{3}).$$

To use the cia structure, let X be a one-point space, and then let $e : X \to X + C(I)$ be inl. Then the solution is a map $e : X \to C(I)$. Its value on the point of X is then some $c \in C(I)$ such that $c = \frac{1}{3}c \cup (\frac{1}{3}c + \frac{2}{3})$, just as desired. The uniqueness of c then also follows from the uniqueness of solutions in cias.

The explanation It might be useful to see why an equation like (11) is related to the procedure of removing the middle third from each element in a collection of intervals.

Let X be the set of non-empty compact subsets of $[0, 1]$. We define a function $f : X \to X$ by

$$f(a) \quad = \quad \frac{1}{3}a \ \cup \ \left(\frac{2}{3} + \frac{1}{3}a\right).$$

We want to solve $x \approx f(x)$. We start with an arbitrary $a \in X$. Then we form a sequence

$$\begin{array}{lcl} c_0(a) & = & a \\ c_{n+1}(a) & = & f(c_n(a)) \end{array}$$

The space X has enough limits (this is where compactness enters), and the sequence $c_n(a)$ is a Cauchy sequence in it. And so we have around to guarantee that it will have a limit. Call this limit a^*. Since f is continuous, we have

$$f(a^*) = f(\lim_n c_n(a)) = \lim_n f(c_n(a)) = \lim_n c_{n+1}(a) = a^*.$$

Thus a^* is a solution to our equation, as desired. But the function f is actually, contracting in certain precise sense, and so by a standard metric argument, a^* will not depend on a at all: starting the iteration at any non-empty compact set would give the same result.

This follows iterative definition of the Cantor set. Let's take a to be $[0, 1]$ itself. Note that removing the middle-third of $[0, 1]$ is gives $f(a)$. That is, rather than think of removal, think of shrinking the interval by one-third, putting one piece on $[0, \frac{1}{3}]$ and the other on $[\frac{2}{3}, 1]$. It just so happens that these operations produce the same thing. And this coincidence persists at the second step: $f(f(a))$ consists of the four pieces $[0, \frac{1}{9}]$, $[\frac{2}{9}, \frac{1}{3}]$, $[\frac{2}{3}, \frac{7}{9}]$, and $[\frac{8}{9}, 1]$. So the iterative method is giving us the terms of the iteration sequence beginning with $[0, 1]$. Finally, the iteration sequence *beginning with* $[0, 1]$ is a shrinking sequence of sets, and then by the way limits are calculated in X, the limit is exactly the intersection of the shrinking sequence.

7 Conclusion

References

[1] Peter Aczel, *Non-Well-Founded Sets*. CSLI Lecture Notes Number 14, CSLI Publications, Stanford, 1988.

[2] Peter Aczel, Jiří Adámek, Stefan Milius, Jiri Velebil, "Infinite trees and completely iterative Theories: a coalgebraic view." *Theoretical Computer Science*, 300 (2003), 1-45.

[3] Peter Aczel, Jiří Adámek, and Jiří Velebil, "A coalgebraic view of infinite trees and iteration," Electronic Notes in Theoretical Computer Science 44.1 (2001).

[4] Peter Aczel and Nax Mendler, "A final coalgebra theorem", in D. H. Pitt et al (eds.) *Category Theory and Computer Science*, Springer-Verlag, Heidelberg, 1989, 357–365.

[5] Jiří Adámek, Stefan Milius, and Jiří Velebil, "On coalgebra based on classes," *Theoretical Computer Science* 316 (2004), no. 1-3, 3–23.

[6] J. Adámek and J. Reitermann, Banach's Fixed-Point Theorem as a Base for Data-Type Equations, *Appl. Categ. Structures* 2 (1994), 77–90.

[7] Jon Barwise and Larry Moss, Hypersets, *The Mathematical Intelligencer*, 13, No. 4 (1991), 31–41.

[8] Jon Barwise and Lawrence Moss, *Vicious Circles*. CSLI Lecture Notes Number 60, CSLI Publications, Stanford, 1996.

[9] George Boolos, "The iterative conception of set." *The Journal of Philosophy*, vol. 68 (1971), 215–231.

[10] John Burgess, reviews in the *Journal of Symbolic Logic*, vol. 50, No. 2 (1985), 544–547.

[11] Daniela Cancila, Furio Honsell, and Marina Lenisa, "Properties of set functors," in F. Honsell et al. eds., *Proceedings of COMETA'03*, ENTCS, 104 , 2004, pp. 61-80.

[12] Keith Devlin, *The Joy of Sets*, second edition. Springer-Verlag, 1993.

[13] Thomas Forster, Why Set theory without the axiom of foundation? *Journal of Logic and Computation*, 4 (1994), no. 4, 333–335.

[14] Marco Forti and Furio Honsell, "Set theory with free construction principles", Annali Scuola Normale Superiore di Pisa, Classe de Sciencze 10:493–522.

[15] Peter Freyd, "Real coalgebra", post on categories mailing list, 22 December 1999, available via `www.mta.ca/∼cat-dist`.

[16] Alexander Kurz, "Coalgebras and Their Logics", *ACM SIGACT News*, Volume 37 , Issue 2 (June 2006), 57 – 77

[17] Azriel Levy, *Basic Set Theory*, Springer-Verlag 1979.

[18] Stefan Milius, "Completely iterative algebras and completely iterative monads," *Inform. and Comput.* 196 (2005), 1–41.

[19] Stefan Milius and Lawrence S. Moss, "The category theoretic solution of recursive program schemes," in J. L. Fiadero et al (eds.) the *Proceedings of CALCO 2005*, Springer LNCS 3629, 2005, 293–312.

[20] Lawrence S. Moss, "Coalgebraic logic," *Annals of Pure and Applied Logic* 96 (1999), no. 1-3, 277–317.

[21] Lawrence S. Moss, "Parametric corecursion," *Theoretical Computer Science* 260 (1–2), 2001, 139–163.

[22] Charles Parsons, “What is the iterative conception of set?”, in Hintikka and Butts (eds.) **Logic, foundations of mathematics and computability theory**. (Proc. Fifth Internat. Congr. Logic, Methodology and Philos. of Sci., Univ. Western Ontario, London, Ont., 1975), Part I, pp. 335–367. Univ. Western Ontario Ser. Philos. Sci., Vol. 9, Reidel, Dordrecht, 1977.

[23] Lawrence C. Paulson, Final coalgebras as greatest fixed points in ZF set theory. Mathematical Structures in Computer Science 9 (1999), 545567.

[24] J.J.M.M. Rutten, “Universal coalgebra: a theory of systems,” *Theoretical Computer Science* 249(1), 2000, pp. 3-80.

[25] Daniele Turi, *Functorial Operational Semantics and its Denotational Dual*, (Ph.D. thesis, CWI, Amsterdam, 1996).

[26] Daniele Turi and J.J.M.M. Rutten, “On the foundations of final semantics: non-standard sets, metric spaces, partial orders,” *Mathematical Structures in Computer Science* 8 (1998), no. 5, 481–540.

[27] I. D. Viglizzo, Coalgebras on measurable spaces, Ph.D. thesis, Indiana University (July 2005).