

# Categories and Constructs

A THESIS  
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL  
OF THE UNIVERSITY OF MINNESOTA  
BY

Shay Allen Logan

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF  
Doctor of Philosophy

Geoffrey Hellman  
Roy T. Cook

August, 2015

# Acknowledgements

I would like to acknowledge support from the National Science Foundation (Grant No. 00006595), the Graduate School (for a first-year fellowship), the Minnesota Center for Philosophy of Science, and my family – especially my wife, Kim.

## **Abstract**

The essays in this collection concern two subjects, each of which falls within the purview of the philosophy of mathematics. The first three essays concern the philosophical status of category theory. The last essay concerns the possibility of a social constructivism regarding mathematicalialia.

# Contents

<b>Acknowledgements</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>List of Tables</b>	<b>vi</b>
<b>List of Figures</b>	<b>vii</b>
<b>1 Category Theory is a Contentful Theory</b>	<b>9</b>
1.1 Categories as Autonomous . . . . .	9
1.2 Objection 1 . . . . .	12
1.3 Objection 2 . . . . .	14
1.4 Objection 3 . . . . .	15
<b>2 Categories for the Neologicist</b>	<b>18</b>
2.1 Abstractionism . . . . .	18
2.2 Definitions and Important Examples . . . . .	22
2.2.1 Poset Categories . . . . .	22
2.2.2 When are Categories the Same? . . . . .	23
2.3 BLC . . . . .	24
2.3.1 Categories Without Objects . . . . .	25
2.3.2 The Abstraction Principle . . . . .	26
2.3.3 Building Ax . . . . .	27
2.3.4 Building Iso . . . . .	31
2.3.5 Defining Cat . . . . .	32

2.4	“New” C . . . . .	33
2.4.1	Small=Finite . . . . .	34
2.4.2	Small=Exponentially Small . . . . .	35
2.5	What Remains to Be Done . . . . .	35
<b>3</b>	<b>Abstractionist Categories of Categories</b>	<b>38</b>
3.1	Introduction . . . . .	39
3.1.1	Abstractionism . . . . .	39
3.2	Relative Abstraction Theory . . . . .	43
3.3	Generators and Functorial Abstraction . . . . .	45
3.4	Language and Definitions . . . . .	47
3.5	Consistency and Size Issues . . . . .	52
3.6	Features of the Category of $U$ -small Categories . . . . .	53
3.6.1	Preliminaries . . . . .	54
3.6.2	Theorems . . . . .	57
3.7	Philosophical Conclusion . . . . .	59
<b>4</b>	<b>The Semantics of Social Constructivism</b>	<b>64</b>
4.1	The Cole-Dieterle Debate . . . . .	64
4.2	Regimenting Cole’s Discussion . . . . .	67
4.3	The Analysis of Temporal Sentences . . . . .	69
4.3.1	Syntax . . . . .	69
4.3.2	Semantics . . . . .	70
4.3.3	An Example . . . . .	71
4.3.4	$\mathcal{L}^P$ as a Solution . . . . .	72
4.3.5	“Since...” and “Until...” . . . . .	72
4.3.6	$\mathcal{L}_{SU}^P$ as a Solution . . . . .	74
4.4	Two-Dimensional Solutions . . . . .	76
4.4.1	Semantics . . . . .	77
4.4.2	An Example . . . . .	79
4.4.3	The $\mathcal{L}^{2P}$ -Solution . . . . .	79
4.5	The Real World as an $\mathcal{L}^{2P}$ -model . . . . .	82
4.5.1	A Complicated World . . . . .	82

4.5.2	A Simple World . . . . .	84
4.5.3	. . . . .	85
4.6	Fictionalism . . . . .	87
4.6.1	Syntax . . . . .	90
4.6.2	Semantics . . . . .	90
4.6.3	$\mathcal{L}^{\infty P}$ as a Solution . . . . .	91
	<b>References</b>	<b>94</b>

# List of Tables

3.1	Multiplication Table for $\mathbf{1}$ . . . . .	56
-----	---	----

# List of Figures

2.1	The poset category associated to the ordinal 4 . . . . .	23
2.2	Some Useful Translations . . . . .	27
2.3	Proof that Left Identities are Loops . . . . .	31
3.1	Multiplication Table for and Picture of <b>2</b> . . . . .	51
4.1	$\mathcal{L}^P$ Semantics . . . . .	70
4.2	$\mathcal{L}_{SU}^P$ Semantics . . . . .	73
4.3	$\mathcal{L}^{2P}$ -Semantics . . . . .	79
4.4	Guide to Important Temporal Quantifiers . . . . .	80
4.5	Key to Translation . . . . .	86
4.6	$\mathcal{L}^{\infty P}$ Semantics . . . . .	90



# Introduction

There are four essays in this collection. All four fall under the broad umbrella of philosophy of mathematics. Three of the essays concern philosophical aspects of category theory. The fourth demonstrates the *unacceptability* of social constructivism as an account of the ontology of mathematics. In this introduction I will summarize each of the papers and draw what connections there are to be drawn among them.

## Chapter 1

The first paper in the collection, *Category Theory is a Contentful Theory*, examines complaints that were made in [1] about the *contentfulness* of category theory. The claim made in that paper was that, unlike set theory, category theory suffered from a lack of *content*. By this the authors meant roughly that one cannot, from the axioms of a theory like that presented in [2] alone, determine what type of thing a category might be.

In this chapter, I refute the claim that category theory suffers from a lack of content by demonstrating that categories can be perfectly well understood as “ways of combining two things to make a third”. This is then extended to include the second component of CCAF, namely, functors. Functors can be understood simply as “ways of associating one thing to another”. CCAF, then, is to be understood as having two sorts of content: some of what it describes is to be understood as ways of combining two things to make a third, some of it as ways of associating one thing to another.

## Chapter 2

The second paper in the collection, *Categories for the Neologist*, produces a neologist theory of categories. Neologism, which is a contemporary re-imagining of the project begun by Frege in his *Grundgesetze* (see [3]), has seen a recent spike in interest. A major reason for this is the *philosophical* insights provided by the apparently purely *mathematical* results it encourages one to explore. It is worth saying a few words about this.

The general “game” of neologism involves finding sentences in second- and higher-order logical languages that have as their consequences theories that are bi-interpretable (in a strict sense) with theories that describe various families of important mathematical objects. The paradigm example of this is Hume’s Principle, **HP**: the second order sentence that (intuitively) says two concepts have the same number just if there is a bijection from the objects falling under one to the objects falling under the other.

If we let  $T_{\mathbf{HP}}$  be the second order theory generated by **HP** and  $\mathbf{PA}^2$  be the second-order theory generated by the second-order Peano axioms for arithmetic, then there are translations between  $T_{\mathbf{HP}}$  and  $\mathbf{PA}^2$  that preserve theoremhood and commute in such a way that one can see  $T_{\mathbf{HP}}$  proves exactly what  $\mathbf{PA}^2$  “says it should” and vice versa (these technical details are explained more carefully below). The upshot is this: in terms of *just HP*, one can prove (an analogue of) every theorem of arithmetic one might want, and can prove nothing more than this. So the single sentence **HP** somehow “contains” the entirety of arithmetic. One has reason to expect, then, that **HP** in some sense gives the “essential features” of the objects of arithmetic – that is, of numbers. The fact that **HP** can intuitively be read as saying that numbers correspond to bijection-classes of concepts is an added bonus, since one intuitively expects that number *ought to* have this feature.

Given the success of **HP** in providing philosophical insight into numbers, then, one avenue of contemporary work in neologism involves proving similar results for other mathematical theories. In my second chapter, I do this for categories – I produce an abstraction principle that gives us objects corresponding in an intuitive way to categories. The fact that the natural way to do this involves the arrows-only definition of a category (the definition of category underwriting the presentation of categories as

“ways of combining two things to make a third”) suggests that the arrows-only definition of categories – which might initially strike one as *ad hoc* or *stilted* in some way – is something like “the right” way to think of categories.

## Chapter 3

The third paper in the collection is by far the most technical. It can be seen, to some extent, as the completion of the project begun in Chapter 2. Rather than producing an abstraction principle aimed at providing a theory of categories alone, Chapter 3 aims to produce an abstraction principle that provides us with a theory of categories of categories. The philosophical insights in this case are far more rewarding.

The inspiration for this project (to which the project in Chapter 2 was a warm-up) came from an examination of Lawvere’s dissertation, reprinted in [4]. In the course of laying out the axioms for his category of categories, Lawvere very early on points out the importance of the fact that  $\mathbf{2}$  is a generator for the category of categories.

In categorial terms, this means that functors can be distinguished by examining the collection of functors that result from precomposing them with a functor from the category that looks like this:

$$\bullet \longrightarrow \bullet$$

If functors  $A$  and  $B$  are *distinct* but have the same domain, there will be a functor  $F$  from a category with this shape to the common domain of  $A$  and  $B$  so that when  $F$  is composed with  $A$  the result is different from when  $F$  is composed with  $B$ .

In his dissertation, Lawvere provides a formal language in which this result corresponds to the following sentence:

$$\forall \mathcal{A} \forall \mathcal{B} \forall f \forall g [\mathcal{A} \xrightarrow{f} \mathcal{B} \wedge \mathcal{A} \xrightarrow{g} \mathcal{B} \wedge \forall u (\mathbf{2} \xrightarrow{u} \mathcal{A} \Rightarrow fu = gu) \Rightarrow f = g]$$

The syntactic form of this sentence was sufficient similar to that of an abstraction principle that I immediately suspected one could build an abstractionist theory of categories of categories directly from an abstraction principle of essentially the same form. Chapter three is the (extended, technical) demonstration that this basic insight is correct. In terms of the results provided in Chapters 1 and 2, it lends support to a very simple

description of categories of categories: a category of categories is a way of combining two things to make a third in which one of the objects is “**2**-ish” and acts as a generator.

It is, to me, entirely surprising that “**2** is a generator” is to categories of categories as “numbers correspond to bijection-classes” is to arithmetic.

## Chapter 4

In Chapter 4 I turn my attention to matters ontological. In [5], [6], and [7], there is a conversation about the possibility of seeing mathematical objects as social constructs. Very roughly, the dialogue is as follows:

- In [5], Julian Cole proposes a carefully crafted theory of mathematical objects as social constructs.
- In [6], Jill Dieterle points out and carefully argues the relevance of a *temporal* objection to social constructivism about mathematical objects. Specifically, Dieterle points out that mathematical objects are paradigm examples of atemporal objects, which seems to conflict with their being social constructs.
- In [7], Julian Cole responds by taking his social constructivism further: not only are mathematical objects socially constructed, so are their “temporal profiles”. That is, not only do we construct the objects, we also declare at what times the objects are to have existed.

The problem with Cole’s response is that it would seem that it is susceptible to a “higher-order” version of Dieterle’s complaint. That is, if temporal profiles are also social constructs, then they too must be objects that were constructed at some particular time. This, in turn, forces us to confront temporal oddities of the following form: for some objects  $x$  and times  $t_1$ ,  $t_2$  and  $t_3$ , it will be the case that, at  $t_1$  the sentence “ $x$  existed at  $t_3$ ,” will be true but at  $t_2$  the sentence “ $x$  existed at  $t_3$ ” will be false. In particular, since Cole produced his account of socially constructed temporal profiles in order to allow for *atemporal* mathematical objects, we must say that there are times  $t_1$  and  $t_2$  such that at  $t_1$  the number seven existed atemporally, but at  $t_2$ , the number seven did not exist atemporally.

In this chapter I provide a formal language and formal semantics with which we can make sense of this type of multi-temporal claim. Unfortunately for Cole, the results are not promising: neither the account he provides, nor any natural extension (and even some unnatural extensions – I show in an appendix that the result extends even to transfinite multi-temporalities) of it will allow him to claim mathematical objects as social constructs and avoid what amount to roughly the same problems as those Dieterle pointed out.

## Remarks on Some Common Themes

The papers found in this dissertation were not meant to form a coherent whole, so I won't pretend they do. Nonetheless, they were all written by me and written in a roughly fifteen-month period, so it would be somewhat surprising if there were *no* common threads whatsoever. Here I'll comment on a few of these threads. I group my comments into two families: remarks on categories and remarks on mathematical objects.

### Remarks on Categories

There was an impressively long-lasting philosophical debate that took place around the status of category theoretic foundations of mathematics. It began with [8], was most recently summarized in [9], and in between was (directly) touched upon by [10], [11], [12], [13], and [14] among others.

The conversation has made at least one thing clear: the appropriateness of category theory as a setting for doing foundational work is, at best, suspect. What one must *not* conclude from this is that categories – and categories of categories – are in themselves somehow “defective”. The first of my papers examines one instance where such a conclusion seems to have been drawn, and corrects the error. The second and third papers demonstrate that both category theory and category-of-category theory are perfectly acceptable parts of mathematics even from the exacting standards given by the neologicists.<sup>1</sup>

---

<sup>1</sup> Admittedly, one could adopt more exacting standards than the neologicists have, though they're far from being the *least* exacting.

All of that, however, is more of a metaphilosophical than a properly philosophical point. The properly philosophical conclusions to draw from my examination of category theory are (a) that categories without objects are, in some sense, the appropriate objects of category theory, and that (b) the central status in category theory of the sentence “**2** is a generator” deserves more philosophical attention.

## Remarks on Mathematical Objects

Regardless of the status of categorial foundations, the constructions Lawvere gave in his thesis shed a different light on many mathematical objects. A group, if we are speaking set-theoretically, is a set together with a certain type of binary operation. A group, if we are speaking categorially, is a one-object category all of whose arrows are invertible. These are remarkably different ways of looking at “the same” mathematical object.

Now, let’s suppose for a moment that we’ve settled on ZFC as the “one true foundation” for mathematics. Let Definition 1 be “a group is a one-object category all of whose arrows are invertible” and let Definition 2 be “a group is a set with an associative binary operation that admits an identity and inverses”. Even having settled on foundational matters, if we take the constructions implied by the Definitions 1 and 2 seriously – that is, if we build groups in the first case by building categories, or in the second place by equipping a set with a binary operation – the resulting objects will be, in general, different. Thus there is a strong sense in which what given a mathematical object *is* seems to vary with one’s *perspective*.

One could be tempted to explain this phenomenon by saying that mathematical objects are social constructs – after all, it seems that, given a singular term  $t$  occurring in my group theoretic practice, the referent of  $t$  might be one thing if I’m working category-theoretically and something entirely different if I’m working set-theoretically. It *appears* to be a small step from here to social constructivism – if the referent of the singular term  $t$  is determined not by  $t$ , nor by the language  $t$  figures in, but instead by the theory we choose to work in, then it seems what  $t$  *is* varies with a characteristically social choice.

It’s important to get a handle on where this argument goes wrong. The position I advocate is similar to the position Eli Hirsch has argued we should adopt regarding *quantifier variance* in a series of essays collected in [15]. It can be put quite simply: the

sort of “semantic pluralism” just pointed out does seem to have as a consequence that what a singular term refers to can vary, and that this variance is a function of social actions. This does not, however, entitle us to conclude that the things a singular term manages to refer to are themselves social constructs.

The point is subtle and requires being made delicately. Let  $t$  be a singular term occurring somewhere in theory  $T$ . Suppose we have on hand a family of ways of *doing*  $T$ -theory, and let the referent of the term  $t$  when doing  $T$ -theory in way  $i$  be  $s_i$ . If it is a social action that determines whether one does  $T$ -theory in way  $i$  or in way  $j$ , then it is a social action that determines whether  $t$  means  $s_i$  or  $s_j$ . This is, I admit a limited form of social construction – what particular object  $t$  picks out varies with social action. But this is not a social constructivism worth writing home about – nobody ever doubted that social actions determined the meanings of words, and that had we used words differently, or should we decide to do so tomorrow, words would mean different things. This much is obvious.

A social constructivism that *would* be worth writing home about would be a demonstration that (at least some of) the  $s_i$  are themselves socially constructed. Nothing in the realm of semantic pluralism warrants our making this further conclusion – it is one thing for us to take a vote to decide whether, starting tomorrow, the word “glurb” ought to refer to the number seven or the Supreme Court. It is another thing for us to take a vote to decide what the number seven or the Supreme Court is. In the former case we are constructing a new fragment of our language – we are deciding what, from among a range of options, we are going to take the word “glurb” to mean. In the latter case we are deciding what various objects *actually are*.

Returning to the case at hand: category theory and set theory provide different settings in which one can do large amounts of mathematics. It is plausible that the referents of many of our mathematical terms will vary with our choice of setting for doing mathematics. We thus can – to an admittedly limited degree – *choose* what the referents of our mathematical terms are. But this should not lead us to conclude the referents of our mathematical terms are themselves the kind of thing that depend on social action. Social constructivism about mathematical objects is a position that cannot be sensibly maintained – this is the conclusion of the final essay in this collection. The

fact that we can *choose* to use “group” in either a “category-theoretic way” or a “set-theoretic way” ought not lead us to conclude that groups are social constructs, but only that the meaning of the term “group” is to some extent determined by social action.



# Chapter 1

# Category Theory is a Contentful Theory

## Introduction

In [1], some objections to category theory as an autonomous foundation are presented. The authors of that paper do a commendable job making clear several distinct senses of “autonomous” as it occurs in the phrase “autonomous foundation.” Unfortunately, the paper seems to treat the “categorist” perspective rather unfairly. Several infelicities of this sort were addressed by [16]. I wish in this note to address yet another apparent infelicity.

## 1.1 Categories as Autonomous

The subject of this chapter is the comments in [1] concerning the *contentfulness* of William Lawvere’s axiomatic system CCAF.<sup>1</sup> For details of this system itself the reader is encouraged to consult [2] or [4]. No technical details from these expositions will be needed, however.

The authors of [1] are willing to admit “CCAF asserts the existence of certain

---

<sup>1</sup> A side note: CCAF is often assumed to abbreviate either “category of categories and functors” or “category of categories as foundation.” In Lawvere’s thesis, however, the system normally called CCAF appears under the heading “category of categories and adjoint functors”.

categories and describes some of the functors between them.” However, as they correctly point out, this by itself is insufficient for CCAF to serve as a foundation of mathematics.

As the authors point out, a necessary condition for CCAF to serve as a foundation of mathematics is that be contentful, and, “if it is to make a contentful assertion, we need to be able to identify its subject matter – namely, categories – independently of the theory.” The authors provide us an illustrative example of a non-contentful theory to help make this objection more clear.

“[A lack of content] would be the problem, for instance, with the theory that consists of the following sentence: ‘the mome raths outgrabe’. The reason that this theory lacks content is that there is no way of identifying, independently of the theory, what mome raths are, or what it is to outgrabe. Without identifying the subject matter of the theory in such a way, the statements of the theory do not make contentful assertions.” [1, p. 231]

Presumably (and the authors are certainly correct here) a theory that runs afoul this objection (which I will call the Contentful Theory Objection [CTO]) cannot serve as an autonomous foundation for mathematics.

Of course, a proponent of CCAF can respond to CTO by simply providing a CCAF-free identification of the subject matter of category theory. The problem, the authors hold, is that this cannot be accomplished without running afoul what they call the Logical Dependence Objection [LDO]: if any theory  $T$  is to provide an alternative foundation for mathematics to the foundation provided by set theory, it must not be the case that  $T$  “depend[s] logically on a prior theory of classes and functions in order to ground [its] existential assertions.” That is, if one cannot identify what the content of  $T$  is without appealing to classes and functions, then  $T$  does not qualify as an *autonomous* foundation for mathematics. In the case of CCAF, Linnebo and Pettigrew seem to suggest that any attempt to avoid CTO by providing a CCAF-free identification of the subject matter of category theory will, in the process of describing this subject matter, make appeal to a prior theory of classes and functions. That is, they claim any attempt to avoid CTO is bound to run afoul LDO. In the remainder of this note I show this is simply false.

First let’s get clear on the content of LDO. As it applies to CCAF it’s fairly clear: if we claim the objects studied by CCAF are, e.g. “things like the category with groups

for objects and group homomorphisms as arrows” then the identification we’ve made is logically dependent on a prior set theory since groups and group homomorphisms are defined set-theoretically.

It’s less clear what LDO has to say about set theory itself. Presumably set theory is contentful, so there is some way of identifying its subject matter independent of set theory itself. Also, since if anything is an independent foundation for mathematics, set theory is, one would hope this identification would not also run afoul LDO.

So what, then, *is* the content of set theory? Geoffrey Hellman has argued for a particularly compelling and parsimonious answer in his [17]:

Set theory is about the operation of collecting. (1)

Thus if set theory manages to avoid CTO without running afoul LDO, it must be the case that I have a set-theory-free way  $W$  of understanding what “collecting” is, and  $W$  itself must not depend on prior understanding of some mathematical theory. Let’s examine how the set theory supporter might go about supplying such a  $W$ .

To begin, she could claim “collecting” simply *is* sufficiently definite to serve as the content of set theory. This amounts to claiming that

- (a) The notion of “collecting” needs no further explanation to be understood,<sup>2</sup> and
- (b) Using this notion we can identify, independently of set theory, the content of set theory.

If we allow set theory this recourse, surely CCAF can help itself to a similar one; the CCAF-theorist can then easily avoid CTO without running afoul LDO by claiming

Category theory is about the operation of combining two things to make a third. (2)

Actually, the phrase “ways of combining two things to make a third” only suffices as a *partial* specification of the subject matter of CCAF. More specifically, CCAF is a first

---

<sup>2</sup> Note this claim seems extremely dubious in light of the antinomies of naïve set theory – we will not address this point in this paper.

order theory about those ways of combining two things to make a third that satisfy the following two conditions:

- (a) Whenever the combination of  $a$  with  $b$  (which we will write  $a \circ b$ ) and the combination of  $b$  with  $c$  (which we will write  $b \circ c$ ) are defined, then both  $a \circ (b \circ c)$  (that is, the combination of  $a$  with  $b \circ c$ ) and  $(a \circ b) \circ c$  (that is, the combination of  $a \circ b$  with  $c$ ) are defined and further these two combinations of things are actually the same thing; and
- (b) It must admit, for each “combinee,” both a left combination-identity and a right combination-identity.

Of course CCAF is supposed to be a first order theory of a category of categories, so to those unfamiliar with category theory it may not be obvious at first that “ways of combining two things to make a third” could be its subject matter. After all, “ways of combining two things to make a third” makes mention of neither objects nor arrows. One might thus wonder where its categorial content is supposed to come from.

These worries are easily laid to rest – it has been long known (actually since the beginning of category theory in [18]) that objects are superfluous to the definition of a category; their role can be played by arrows. When characterizing categories in an object-free way, the only axioms that matter are those specifying that the composition of arrows be associative and that it admit, for each arrow, a left and a right identity element. Categories viewed from this perspective are nothing more than a type of algebra of arrows – that is, a way of combining two things (arrows) to make a third (their composite). Since the object-free characterization of categories is in fact precisely the characterization that underlies CCAF, it is especially appropriate to use here. I turn now to examining a few objections to the idea that (2) is sufficiently definite to serve as the subject matter of category theory.

## 1.2 Objection 1

One might see the word “things” in (2) and object that “ways of combining two things to make a third” is not sufficiently definite to characterize anything until we’ve been

told what the membership-structure of the “thing-space” is. This seems to be what Linnebo and Pettigrew address in one of their footnotes:

“the proponent of CCAF as a foundation may complain that a category need not involve a *set* of objects and a *set* of arrows but rather a *collection* or *aggregate* of objects and a *collection* or *aggregate* of arrows. But this will not buy him much time, since our best theory of collections or aggregates or pluralities of any sort is set theory.”[1, p. 232, footnote 5]

But why must the CCAF-supporter specify at all what structure is formed by the relation “ $x$  is an arrow of category  $y$ ” just because the Set Theory-supporter is in the habit of doing so? It seems perfectly coherent to study “ways of combining things” without needing information about the structure formed by the membership pattern that holds between the things being combined and their totality.

For example: one can easily imagine building physical instantiations of associative ways of combining two things to make a third that admit left and right identities. These would be machines that take in two things and output a third, and which behave as they ought to in order to model this description. A trained mechanic could probably build many such machines, repair such machines when they broke, identify when two such machines were essentially the same, etc. All of this could be done without a moment’s thought being given to what structure was being instantiated by relation “ $x$  is a possible input to machine  $y$ ,” or whether there is such a structure at all – the mechanic could work perfectly well with the relation “ $x$  can combine with something to produce an output” instead, for example.

Perhaps an alternative argument will make this clearer: if we take as primitive the membership relation, then it can appear the CCAF-supporter has failed to specify the membership-structure of her space of arrows.<sup>3</sup> But by the same token, if we take as primitive “ways of combining two things to make a third,” then it can appear as if the set-theorist has similarly failed to specify the combination-structure of her spaces

---

<sup>3</sup> This may *appear* to be the case, but it is at least debatable whether it *is* the case. One can make perfectly coherent the notion of a discrete category within CCAF; and can use this notion to define the words “set,” “membership,” and the like (again, the interested reader is urged to consult Lawvere’s work for the details). Nonetheless, arguing for the claim that CCAF admits a membership relation that is as robust as the membership relation present in, say, ZFC is not relevant to the details of this particular paper.

of elements. So when one begins either from the *assumption* that “ $x$  is a member of  $y$ ” is primitive or that “combining  $x$  with  $y$  gives  $z$ ” is primitive, those structures specified only in terms of the other relation seem insufficiently specified.

### 1.3 Objection 2

One could perhaps object that, even if “ways of combining two things to make a third” is sufficiently definite to define the content of *some* theory, nonetheless in order to single out those ways that are associative and admit left and right identities (which we must do to specify the content of CCAF in particular) one must have a prior understanding of sets. Thus, even if we can understand “ways of combining two things to make a third” without relying on a prior theory of collection and membership, we nonetheless need such a theory to be able to state which particular ways we are interested in as CCAF-theorists.

But we can state versions of the usual category-axioms specifying the associativity of composition and the existence of identity morphisms without quantifying over collections at all as follows:

Let  $W$  be a way of combining two things to make a third. Let  $a \circ_W b$  stand for the result of combining  $a$  and  $b$  (in this order) in way  $W$ . Then we say  $W$  is an associative way of combining two things to make a third that admits left and right identities if

- Whenever  $a \circ_W b$  and  $b \circ_W c$  are defined, so are  $a \circ_W (b \circ_W c)$  and  $(a \circ_W b) \circ_W c$  and these two are equal.
- For every  $a$  there are things  $1_{sa}$  and  $1_{ta}$  so that

$a \circ_W 1_{sa} = a$  and for any  $b$  if either of  $b \circ_W 1_{sa}$  or  $1_{sa} \circ_W b$  is defined then it equals  $b$ ; and

$1_{ta} \circ_W a = a$  and for any  $b$  if either of  $1_{ta} \circ_W b$  or  $b \circ_W 1_{ta}$  is defined then it equals  $b$ .<sup>4</sup>

---

<sup>4</sup> It may appear that these two by themselves leave open the possibility that the identities are not unique. However, if  $1_{sa}$  and  $1'_{sa}$  were two right identities for  $a$ , then we would have that  $1_{sa} = 1_{sa} \circ_W 1_{sa'} = 1'_{sa}$ . Thus, right identities are unique. A similar argument shows left identities to also be unique.

Since we can thus explain which ways of combining two things to make a third are the associative ones that admit both a left and a right identity for each element without even *quantifying over* collections, surely we can understand this without *relying on a theory of* collections.

Altogether, then, it seems if we justify set theory’s status as an autonomous foundation by claiming (1) is sufficiently definite to serve as the subject matter of set theory, then we can also justify CCAF as an autonomous foundation by claiming (2) is sufficiently definite to serve as the subject matter of that theory.

The proponent of set theory may at this point wish to go down another road altogether and say the reason (1) is an acceptable identification of the content of set theory is because it relies only on purely logical notions. That is, “a collection of things” – which, it would seem, is the result when one does some collecting – can be understood by simply grasping what it means for an object to hold a particular property – the collecting the blahs is just gathering up all those  $x$ ’s for which “ $x$  is a blah” is a truth.

If this is the road the set theorist goes down, it is extremely difficult to see how she will block the CCAF theorist from taking the same path. A proponent of CCAF can do this most easily by pointing out that “ways of combining two things to make a third” can be perfectly well explained in terms of ternary relations. Thus, if the set theorist is allowed to claim set theory as an autonomous foundation because the content of set theory can be grasped using pure logic, then it seems the CCAF theorist should be allowed to make the same claim – unless the set theorist has some reason to claim that properties are logical while ternary relations are not.

### 1.4 Objection 3

Finally, the set theorist can object to (2) as a characterization of the subject matter of CCAF on the grounds that, in addition to categories, CCAF contains further elements known as *functors*, and unless one can describe what functors are, one has only succeeded in giving a description of part of the content of the CCAF axioms. This, however, is easily accommodated: a functor is just a way of associating one thing to another. Actually, just as with “way of associating two things to make a third”, we need a few extra details here. In words, we have the following:

Given two ways of combining two things to make a third  $\circ_1$  and  $\circ_2$ , a functor  $f$  from the first way to the second is a way of associating combinees from the first way to combinees from the second way such that  $f(a \circ_1 b) = f(a) \circ_2 f(b)$ . That is, such that one associates to the result of combining (in the first sense) two things the same thing that results from combining (in the second sense) what one has associated the two things to.

As with “ways of combining two things to make a third”, “ways of associating one thing to another” avoids both Objection 1 and Objection 2. CCAF is then a combined first-order theory describing certain ways of combining two things to make a third and ways of associating one thing to another.

## Conclusion

Of course, Linnebo and Pettigrew’s arguments are closely related to arguments in [8] that have been revisited repeatedly. A summary of “the ways in which Feferman’s (1977) arguments have been used (and misused) in the philosophical literature” can be found in [9]. While the argument above may appear to merely contribute to this collection of uses (and hopefully not to the misuses), it should be pointed out that one can read in the argument I have offered an *agreement* with Feferman’s basic point that “the general concepts of operation and collection have logical priority with respect to structural notions.” Admittedly, the agreement is rather tenuous, as the theory proposed above seems to do Feferman one better by relying only on a previous understanding of the general concept of an operation. Nonetheless, I can agree to “the general concepts of operation and collection” having logical priority while still maintaining that this leaves set theory and category theory on equal footing *as regards their contentfulness*; the difference between the two cases amounts only to the following: where the operation assumed in set-theoretic foundations is a binary membership relation, the operation assumed in category-theoretic foundations is a ternary composition relation. Thus, if set theory is taken to avoid LDO and if, to the extent that it does so, set theoretic foundations are to be labeled as autonomous (in the sense of Linnebo and Pettigrew),



category theoretic foundations should be given the same honors.<sup>5</sup>

The lesson to be learned from all this is that remaining silent sometimes is a perfectly fine answer. The mere fact that the things the set theorist likes to talk about are things the category theorist does not find interesting does not make the category theorist's contributions dependent on the set theorist's – not everyone has to have something to say about sets.

---

<sup>5</sup> Feferman, of course, famously rejects both impredicative set theoretic foundations as well as category theoretic foundations. So he might be willing to make this concession without seeing it as giving up very much.

## Chapter 2

# Categories for the Neologist

Abstraction principles provide implicit definitions of mathematical objects. In this paper, an abstraction principle defining categories is proposed. It is unsatisfiable and inconsistent in the expected ways. Two restricted versions of the principle which are consistent are presented.

### 2.1 Abstractionism

An abstraction principle is a sentence with the following syntactic form:

$$\forall A \forall B (f(A) = f(B) \iff E_f(A, B))$$

Where  $A$  and  $B$  are of the same logical type,  $f$  is a term-forming operator and  $E_f$  is a 2-place predicate defining an equivalence relation on entities of the type of  $A$  and  $B$ . Such sentences are taken to define the mapping represented by the operator  $f$  and, via this, the objects that occur in its range.

Being more careful, we observe that in set-based models for an abstraction principle, the operator  $f$  will be modeled by a function from entities of the appropriate type to the domain of the model. Objects in the range of this function are identified exactly when they are images of entities that are related by the equivalence relation represented by the two-place predicate  $E_f$ . Thus, the abstraction principle gives (at least something very like) identity conditions for the objects occurring in the range of the mapping

picked out by  $f$ .<sup>1</sup> These identity conditions, in turn, provide an *implicit definition* of these objects – an implicit definition that, importantly, relies for its intelligibility only on the intelligibility of the vocabulary used in forming  $E_f$ . Thus, if we can produce an abstraction principle wherein the vocabulary occurring in  $E_f$  has a privileged epistemic status – if, for example, there is reason to believe that we can come to know what sentences in that vocabulary *mean* purely in terms of the meaning of the components of such sentences, and we have reason to believe we can come to know the meaning of the basic components of these sentences in some highly secure way – then the fact that there is such an implicit definition can help to explain the privileged epistemic status of our knowledge regarding the objects defined by that abstraction operator.<sup>2</sup>

One may well worry whether such grand plans are likely to reach fruition. The surprising answer is a resounding yes! Two well-known examples of abstraction principles will help make this point: Hume’s Principle (**HP**), and Frege’s infamous Basic Law V (**BLV**):

$$\mathbf{HP}: \forall X \forall Y (\#(X) = \#(Y) \iff X \text{ bijects with } Y)^3$$

$$\mathbf{BLV}: \forall X \forall Y (\S(X) = \S(Y) \iff \forall z (Xz \iff Yz))$$

Let  $T_{\mathbf{HP}}$  be the theory that takes **HP** as its lone non-logical axiom, working in a background logic containing at least second-order logic with the full comprehension scheme. Let  $\text{PA}^2$  be second order Peano arithmetic. What is known as Frege’s Theorem (see, e.g. [22, 23]) tells us there is a translation of (the language of)  $\text{PA}^2$  into (the language of)  $T_{\mathbf{HP}}$  such that the translation of every axiom of  $\text{PA}^2$  is a theorem of  $T_{\mathbf{HP}}$ . What is known as Boolos’ Theorem (see, e.g. [24]) gives the converse result: there is a translation of  $T_{\mathbf{HP}}$  into the language of  $\text{PA}^2$  such that the translation of each  $T_{\mathbf{HP}}$ -theorem is a theorem of  $\text{PA}^2$ . Further, it can be shown that the composition of these translations maps each sentence  $\phi$  to a sentence equivalent to  $\phi$ . So **HP** serves as an implicit definition of numbers (more carefully: of Peano-numbers, or Peano-cardinals) that relies for its intelligibility only on the vocabulary of second-order logic.

---

<sup>1</sup> A good exposition of the neologicist programme that emphasizes the precise nature of the identity conditions they provide is found in [19].

<sup>2</sup> Further discussions of abstractionism, the status of abstraction principles as implicit definitions, and the role of second order logic can be found in [20] and [21].

<sup>3</sup> Note that a 2-place “bijects with” predicate can be encoded entirely in second-order logic.

**BLV**, on the other hand, is well-known to be inconsistent. With a bit of tinkering, one can easily see that in a standard model of **BLV** in a domain  $D$ , the function  $\S$  would be an injection from the set of all subsets of the domain of the model into the domain of the model. Such an injection is of course impossible, by Cantor’s theorem. This is a typical problem with abstraction principles (we will encounter it again below): if the equivalence relation  $E_f$  is too *fine grained*, no function  $f$  can satisfy every instance of the biconditional. For **BLV**, however, there is a partial workaround for this problem. In place of **BLV**, one can instead use principles (typically referred to in the literature (e.g. [25]) as “NewV”-type principles) of a slight different form that are consistent and which give rise to relatively powerful versions of set theory. In general, given a second-order (but purely logical) predicate “**Small**” holding of (intuitively) of those concepts under which only a small fragment of the domain of quantification falls, a “NewV”-type principle looks like

$$\mathbf{NewV}: \forall X \forall Y (\S(X) = \S(Y) \iff [\neg(\mathbf{Small}(X) \vee \mathbf{Small}(Y)) \vee \forall z (Xz \iff Yz)])$$

So while **BLV** itself will not work to found our knowledge of set theory, NewV-type principles serve as implicit definitions of (sufficiently small) sets, and rely for their intelligibility only the vocabulary of second-order logic.

These successes motivate further investigation of abstraction principles despite the need to qualify the success of **BLV**. Indeed, **HP** *seems to give* a much more natural definition of number than  $\text{PA}^2$  does, and to do so in a vocabulary (that of second-order logic) that it is not unreasonable to suppose we have some type of privileged epistemic access to. Further, it is a remarkable *mathematical* fact that this one natural-feeling axiom can provide, all by itself, a theory bi-interpretable with  $\text{PA}^2$ .

One needn’t have logicist inclinations to find projects like these interesting. From a metaphysical and mathematical point of view, abstraction principles “work” by proposing identity conditions for abstract objects. In the cases that are of interest one can prove – using (usually second-order) logic and the specified identity conditions – that the objects implicitly defined by an abstraction principle have all the properties they are “supposed to” have. When this is possible, we have reason to believe the identity conditions proposed specify something like the essential features of the objects in question.

It is thus a generally worthwhile project to produce abstraction principles that implicitly define objects that can be proved to have all the features the objects in some specified family of mathematical objects are supposed to have. That is, given a type  $t$  of mathematical object, it is philosophically valuable to produce an abstraction principle  $A_t$  for which both of the following hold:

- The objects implicitly defined by  $A_t$  can be shown, in the theory generated by  $A_t$ , to have all the features objects of type  $t$  are supposed to have, and
- For either *all* or at least *a substantial portion* of the objects  $x$  of type  $t$ , there corresponds an object  $x_A$  among those objects implicitly defined by  $A_t$  that exhibits all the features of  $x$ .

Thus, for example, the objects implicitly defined by **HP** have, by Frege and Boolos' theorems, exactly the features we expect Peano-numbers to have, satisfying the first condition. Further, one sees (e.g. by a “bootstrapping” argument like that in [26, pp. 133ff]) that models of **HP** always contain infinitely many numbers, satisfying the second condition. We thus have reason to believe that **HP** captures something like the essential features of the number concept.

One might still wonder, after doing this, whether the objects implicitly defined by  $A_t$  are actually *the* objects of type  $t$ . This, unfortunately, is an issue that has plagued logicism from the beginning – the so-called Casesar Problem. It would take us rather far afield to address this issue, so we won't. Instead, I will point out the following: whether the objects implicitly defined by  $A_t$  are actually *the* objects of type  $t$  or not, the very fact that  $A_t$  provides us with identity conditions for objects that provably *behave just like* the objects of type  $t$  – and that this can be done entirely in the vocabulary of second-order logic – is a significant philosophical and mathematical insight. It suggests, again, that the identity conditions pointed out mark *something like* the essential features of the objects in question.

In the remainder of the paper, I seek and propose an abstraction principle, the objects of which behave (in standard set-based models) just like categories. In doing so, I will work, as is standard in the neologicist literature, in a set theoretic metatheory. This dependence<sup>4</sup> on set theory in the metatheory is a weakness of neologicism in its

---

<sup>4</sup> [27] contains a very explicit explanation of this dependence. It is treated in detail in [28].

current state – a ladder that should eventually be kicked away. The philosophical value of finding such a principle is analogous to the philosophical value of the relationship between number and equinumerosity established by Frege and Boolos’ theorem: finding such a principle gives us reason to believe we’ve found something like the essential features of categories.

## 2.2 Definitions and Important Examples

Recall that in the standard definition, a category  $\mathcal{C}$  consists of the following:

- A collection of objects  $\text{Ob } \mathcal{C}$ ;
- For each  $(A, B) \in \text{Ob } \mathcal{C} \times \text{Ob } \mathcal{C}$  a collection  $\text{Mor}(A, B)$  called the morphisms from  $A$  to  $B$ ; and
- An associative operation that assigns to each  $(\phi, \psi) \in \text{Mor}(B, C) \times \text{Mor}(A, B)$  a morphism  $\phi \circ \psi \in \text{Mor}(A, C)$  called the composite of  $\phi$  and  $\psi$ .

We further require for each  $A \in \text{Ob } \mathcal{C}$  a morphism  $1_A \in \text{Mor}(A, A)$  so that for any object  $B$  and  $\phi \in \text{Mor}(A, B)$ ,  $\phi \circ 1_A = \phi$ , and for any object  $C$  and  $\psi \in \text{Mor}(C, A)$ ,  $1_A \circ \psi = \psi$ .

Typical examples of categories include the category with sets for objects and functions for morphisms; categories with structures of various sorts (e.g. groups, rings, Hopf algebras, etc.) for objects and structure-preserving mappings among such for morphisms; and categories with spaces of some sort (topological, smooth,  $C^k$ ,  $C^\omega$ , etc.) for objects and sufficiently well-behaved maps for morphisms. Another class of categories important for our purposes is the class of poset categories.

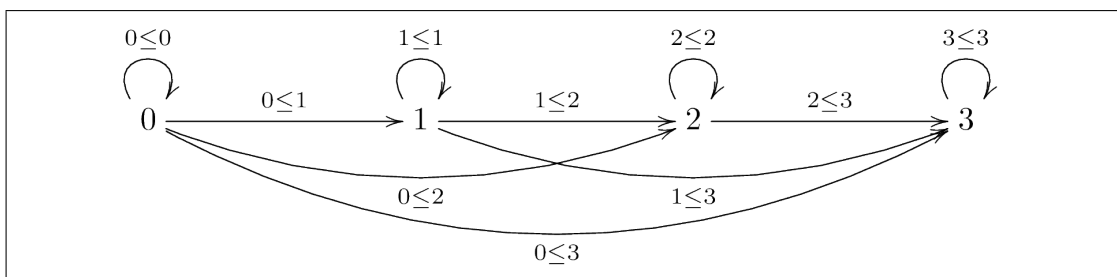
### 2.2.1 Poset Categories

A poset is a set  $P$  together with a reflexive, transitive, antisymmetric relation  $\leq$  on  $P$ . Given a poset  $(P, \leq)$ , we define a category  $\mathcal{P}$  as follows:

- $\text{Ob } \mathcal{P} = P$ ;
- The collection of morphisms between objects  $a$  and  $b$  is empty when  $a \not\leq b$ , and has one element (which we call  $\lceil a \leq b \rceil$ ) when  $a \leq b$ .

- $\lceil b \leq c \rceil \circ \lceil a \leq b \rceil = \lceil a \leq c \rceil$

The cardinality of the set of morphisms in the poset category  $\mathcal{A}$  associated to an ordinal  $\lambda$  will be important to us later. (For  $\lambda = 4 = \{0, 1, 2, 3\}$ , a picture of  $\mathcal{A}$  is presented in Figure 2.1.) We count the morphisms in such a category by adding up the number of morphisms leaving each object as we range from the first object to the last – that is, for an arbitrary ordinal  $\lambda$  the cardinality of the set of morphisms in  $\mathcal{A}$  is  $\sum_{\gamma < \lambda} |\lambda - \gamma|$ . If  $\lambda$  is infinite and we assume choice (which we do throughout this paper), then this is simply  $|\lambda|$ .



**Figure 2.1:** The poset category associated to the ordinal 4

### 2.2.2 When are Categories the Same?

We now turn to determining the appropriate category-theoretic identity conditions to impose on the objects defined above. This task is worth pausing to consider: in many mathematical practices, there is a readily apparent bifurcation of identity conditions: on the one hand, the foundational role played by set theory demands we define objects in such a way that many non-identical but isomorphic objects exist; on the other, most mathematical practices are actually only interested in studying the isomorphism-invariant features of the objects it considers. In practice, in fact, the objects quantified over by many sentences in many mathematical theories seem to either *be* or at least *correspond to* isomorphism-classes of objects given by the foundational constructions of that theory. For example, there is the (in)famous *Atlas of Finite Groups* [29]. Notice the title of this work is not “*Atlas of isomorphism-classes of finite groups.*” The neo-logicist can make sense of this omission in a particularly parsimonious way: abstraction principles allow us to produce, corresponding to what the set-theoretic foundationalist

would call, e.g. the isomorphism class of finite simple groups of order two, an individual object naturally labeled *the* finite simple group of order two.

We should thus examine category-theoretic practice to determine what the appropriate identity conditions are on categories for the working category-theorist. On inspection, one finds there are two candidate relations among categories to consider: isomorphism and equivalence. Categories  $\mathcal{C}$  and  $\mathcal{D}$  are *isomorphic* when there is an invertible structure-preserving map between them – that is, when there is an object-to-object and morphism-to-morphism map  $\mathcal{F} : \mathcal{C} \rightarrow \mathcal{D}$  that preserves all categorial structure and for which the inverse map  $\mathcal{F}^{-1} : \mathcal{D} \rightarrow \mathcal{C}$  also preserves all categorial structure. On the other hand, category  $\mathcal{C}$  is *equivalent* to category  $\mathcal{D}$  when there are structure-preserving maps  $\mathcal{F} : \mathcal{C} \rightarrow \mathcal{D}$  and  $\mathcal{G} : \mathcal{D} \rightarrow \mathcal{C}$  and morphisms  $i_x : \mathcal{G} \circ \mathcal{F}(x) \rightarrow x$  for each object  $x$  of  $\mathcal{C}$  such that each  $i_x$  is invertible in  $\mathcal{C}$ . Equivalence is a strictly *looser* relation than isomorphism – categories that are isomorphic are equivalent, since we can take the maps  $i_x$  to be the identity map in each case.

In practice, the notion of equivalence suffices for almost all standard applications of category theory. Despite this, it is more philosophically appropriate for the abstractionist to – at least at this early stage – use the former, stricter notion for her project, since it is occasionally important in practice that there exist non-isomorphic, equivalent categories. It would be philosophically worthwhile, however, to produce an abstraction principle that implicitly defines objects corresponding to categories identified only up to equivalence, rather than up to isomorphism. This project is likely to be pursued in future work.

### 2.3 BLC

The initial challenge: any formalization of the definition of a category (in particular via an abstraction principle) seems to demand language allowing separate quantification over objects and over morphisms. A language with more than one type of first-order variable for example would seem to allow such flexibility by letting us specify, using distinct quantifiers, the assumptions that must be made regarding the behavior of the objects and the behavior of the morphisms.

But for the neologist it is not a trivial matter to adopt a new language. As pointed



out above, an important part of the success of the abstractionist programme relies on the vocabulary occurring in the equivalence relations having a privileged epistemic status. Such privileging demands careful argumentation to justify, and cannot easily be transferred from one language to the next.

We should thus seek, if possible, to build the theory of categorial abstraction without leaving “standard” second order logic. It turns out that this is possible because no category-theoretic information is lost if we allow the role of objects to be played by identity morphisms. This allows us to work in a standard language for second-order logic containing only one type of first-order quantifier.<sup>5</sup> For completeness, we next present an explicit morphism-only definition of categories.

### 2.3.1 Categories Without Objects

With reference only to morphisms, we define a category to be a triple consisting of a collection  $M$ , a designated subcollection  $\text{Comp} \subseteq M \times M$  and a mapping  $\circ : \text{Comp} \rightarrow M$  all subject to the following:<sup>6</sup>

(CWO1) If  $(a, b)$  and  $(b, c)$  are both in  $\text{Comp}$ , then  $(\circ(a, b), c)$  and  $(a, \circ(b, c))$  are in  $\text{Comp}$  as well and  $\circ(\circ(a, b), c) = \circ(a, \circ(b, c))$

(CWO2) If  $a \in M$  then there are elements  $1_{Da}$  and  $1_{Ca} \in M$  that satisfy the following:

- (R1)  $(a, 1_{Da})$  and  $(1_{Da}, 1_{Da})$  are both in  $\text{Comp}$ ;
- (R2) If  $(x, 1_{Da}) \in \text{Comp}$  then  $\circ(x, 1_{Da}) = x$ ; and
- (R3) If  $(1_{Da}, x) \in \text{Comp}$  then  $\circ(1_{Da}, x) = x$ . Also,
- (L1)  $(1_{Ca}, a)$  and  $(1_{Ca}, 1_{Ca})$  are both in  $\text{Comp}$ ;
- (L2) If  $(1_{Ca}, x) \in \text{Comp}$  then  $\circ(1_{Ca}, x) = x$ ; and
- (L3) If  $(x, 1_{Ca}) \in \text{Comp}$  then  $\circ(x, 1_{Ca}) = x$ .

The collection  $M$  is to be thought of as the collection of morphisms of some category,  $\text{Comp}$  as specifying which morphisms compose, and  $\circ$  as the composition itself. Each

<sup>5</sup> Note that this idea dates to the origins of category theory in [18].

<sup>6</sup> The words “collection” and “mapping” are meant to be more general than the more standard “set”, “function” – no restriction in size is supposed at this point in the essay.

$a \in M$  is (intuitively) a morphism, so each  $a$  intuitively has a domain (where it comes from) and a codomain (where it goes to). In terms only of morphisms and composition, this means that each  $a \in M$  admits a left-composition-identity and a right-composition-identity. CWO2 guarantees the existence of an element  $1_{Da}$  that plays the “right-composition-identity for  $a$ ”-role (thus, intuitively, the role of the identity morphism at the domain of  $a$ ). Similarly,  $1_{Ca}$  plays the role of the left-composition-identity for  $a$  (thus, intuitively, of the identity morphism at the codomain of  $a$ ). Notice it is a consequence of CWO2 that  $\circ$  is a surjection. Thus, specifying  $\circ$  suffices to specify both  $\text{Comp}$  and  $M$ . So each category, from the morphism-only point of view, corresponds to a type of two-place mapping.

Of course, specifying that the category have some *particular* family  $M$  of morphisms – or, correspondingly, that it be identified with a *particular* mapping  $\circ$  – is not particularly faithful to category-theoretic practice. It would be akin, in group theoretic practice, to demanding that the group theorist tell you precisely what the elements of the Klein four group are before she tell you anything further about its features (are they the numbers 0, 1, 2, and 3? The ordered pairs (0,0), (0,1), (1,0) and (1,1)? Etc.). As we mentioned above, in category-theoretic practice, such objects are considered to be “the same” object when they are isomorphic (or equivalent, but we’ve agreed to use the stricter notion for the time being). Put more simply, the correspondence between categories (as that term is used in practice) and two-place functions is one-to-many.

### 2.3.2 The Abstraction Principle

Our goal is to produce an abstraction principle that implicitly defines objects corresponding to categories as the term “category” is used in category-theoretic practice. Given that we’ve seen the correspondence between categories and two-place mappings is one-many, we can expect that the abstraction principle will have the following features:

- The variables at the front-end of the abstraction principle should range over two-place functions,
- There will be two conditions such functions must satisfy in order that we say they correspond to the same category

- They must each satisfy conditions corresponding to (CWO1) and (CWO2), and
- They must be isomorphic.

So, to produce such an abstraction principle, we will need to build two predicates:

- A one-place predicate  $Ax$  such that  $Ax(f)$  holds exactly when  $f$  is a mapping that represents the composition in a category, and
- A two-place predicate  $Iso$  defining the equivalence relation separating the two-place functions into category isomorphism classes.

Once we do this, we expect that our abstraction principle will have the following general form:

$$(\forall f)(\forall g) \left[ \text{Cat}(f) = \text{Cat}(g) \longleftrightarrow \left( (\neg Ax(f) \wedge \neg Ax(g)) \vee \text{Iso}(f, g) \right) \right]$$

Where the variables  $f$  and  $g$  are of 2-place-function type;<sup>7</sup>

In the remainder of the section, we construct the predicates  $Ax$  and  $Iso$ . There are in fact three related predicates that we construct along the way:  $\mathbf{ax}$ ,  $Ax$ , and  $\text{Mdom}$ . Their intuitive meanings are presented in Figure 2.3.2. The reader may find it helpful to refer back to this chart in the course of reading the technical construction that follows.

Expression	Intuitive Meaning
$\mathbf{ax}(f, P)$	$f$ , restricted to the objects falling under $P$ , behaves like the composition of a category.
$Ax(f)$	There is a unique maximal concept $P$ such that $f$ satisfies $\mathbf{ax}(f, P)$ .
$\text{Mdom}(P, f)$	$P$ is the unique maximal concept such that $f$ satisfies $\mathbf{ax}(f, P)$ .

**Figure 2.2:** Some Useful Translations

### 2.3.3 Building $Ax$

The first step is to build the predicate  $Ax$ . We begin with a discussion of identity morphisms. (CWO2) tells us how these ought to behave. We focus on left-identities first; right identities will be similar.

<sup>7</sup> Note that this does not mean  $f$  and  $g$  range over 2-place *functions*, but that the syntactic type of these variables is that of 2-place functions. One can, of course, choose a model in which such variables range over a much broader (or narrower!) class of entities than mere 2-place functions.

As a naïve first guess, we could try to formalize “every morphism has a left-identity” by the sentence

$$(\forall x)(\exists y)[f(y, x) = x] \quad (2.1)$$

Intuitively, (2.1) guarantees every morphism  $x$  has a corresponding morphism  $y$  that is a left identity for it. Unfortunately,  $y$  being a left identity for  $x$  does not fully capture what it means for  $y$  to be the identity morphism at the codomain of  $x$ , as the following example demonstrates:<sup>8</sup>

Consider the “almost category”  $A$  with  $\text{Ob } A = \{\mathbb{R}^{\mathbb{N}}\}$  and  $\text{Mor}(\mathbb{R}^{\mathbb{N}}, \mathbb{R}^{\mathbb{N}}) = \{\pi_i\}_{i \in \mathbb{N}}$ , where

$$\pi_i(f)(n) = \begin{cases} f(n) & : n < i \\ 0 & : n \geq i \end{cases}$$

and with composition defined as usual.

Notice for any  $j > i$ ,  $\pi_j \circ \pi_i = \pi_i \circ \pi_j = \pi_i$ , so  $A$  models the axiom  $(\forall x)(\exists y)[\circ(y, x) = x]$  and its obvious right-hand counterpart. But  $A$  lacks a genuine identity morphism because no  $\pi_i$  is an identity morphism for *all* the  $\pi_j$ . So satisfying (1) is not the same as satisfying the conclusions of either (L2) or (L3).

A natural second candidate axiom is the following emendation of (2.1):

$$(\forall x)(\exists y)[f(y, x) = x \wedge (\forall z)(f(y, z) = z \vee f(z, y) = z)] \quad (2.2)$$

Sentence (2.2) does solve the particular problem presented by  $A$ , but it has new problems all its own.

Recalling that  $f$  is supposed to represent the composition in an arbitrary category, we can see that these problems stem from the following general issue: even when  $y$  is an identity morphism, we should not demand  $f(y, z)$  and  $f(z, y)$  be  $z$  for all  $z$ , but only for those  $z$  that  $y$  in fact composes with – in general there will be many morphisms  $z$  for which  $f(y, z)$  is simply undefined. Notice, in particular, that (L2) and (L3) each explicitly contain this as their antecedent. So, given category  $\mathcal{C}$  represented by mapping  $f$ , and given identity morphism  $y$  and arbitrary morphism  $z$ , we should have that *either*  $f(y, z) = z$ , *or*  $y$  and  $z$  don’t compose in  $\mathcal{C}$  (and similarly for  $f(z, y)$ ). For now we write

---

<sup>8</sup> Imposing (1) also does not guarantee identity morphisms be unique, but in elementary category theory identity morphisms are proved unique using associativity and the definition of identity morphisms, so it is natural to ignore *this* problem until after treating associativity.

$f(y, z) = \perp$  to mean “ $y$  and  $z$  don’t compose (in this order) in the category represented by  $f$ .”

Incorporating this, the assertion that identities exist seems captured by

$$(\forall x)(\exists y) \left[ f(y, x) = x \wedge (\forall z)(f(y, z) = z \vee f(y, z) = \perp) \wedge (\forall z)(f(z, y) = z \vee f(z, y) = \perp) \right] \quad (2.3)$$

A subtle problem remains: Notice (3) clearly gives us that if  $y$  is an identity morphism then either  $f(y, y) = y$  or  $f(y, y) = \perp$ . But the latter option is incompatible with  $y$  being an identity morphism, since (L1) specifies that identity-morphisms do in fact compose with themselves.

We *could* just adopt  $f(y, y) \neq \perp$  as an axiom to resolve this, but given we use  $\perp$  to mean “doesn’t compose” we really ought to have that for any  $x$ ,  $f(\perp, x) = \perp$ . That is, it is natural to take as an axiom the sentence

$$(\forall x)[f(x, \perp) = \perp \wedge f(\perp, x) = \perp],$$

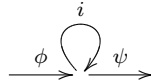
Which rules out a host of pathologies, among which are non-looping identity morphisms.

Since right identities will be analogous to left these considerations suffice to determine our first three axioms:

$$\begin{aligned} \text{(I)} \quad & (\forall x)[f(x, \perp) = \perp \wedge f(\perp, x) = \perp] \\ \text{(II)} \quad & (\forall x)(\exists y) \left[ f(y, x) = x \wedge (\forall z)(f(y, z) = z \vee f(y, z) = \perp) \right. \\ & \left. \wedge (\forall z)(f(z, y) = z \vee f(z, y) = \perp) \right] \\ \text{(III)} \quad & (\forall x)(\exists y) \left[ f(x, y) = x \wedge (\forall z)(f(y, z) = z \vee f(y, z) = \perp) \right. \\ & \left. \wedge (\forall z)(f(z, y) = z \vee f(z, y) = \perp) \right] \end{aligned}$$

Together these ensure that *identity* morphisms behave appropriately. A formal proof of part of this fact is given in Figure 2.3 on page 31. We next deal with (CWO1).

The conditions under which the composite of  $\phi$  and  $\psi$  is defined are captured by the diagram



Translating from diagram-ese to English, this says it makes sense to talk about the composite  $\psi \circ \phi$  if and only if there is an morphism  $i$  that is both the left identity for  $\psi$  and the right identity for  $\phi$ . Formalizing, we arrive at our penultimate axiom:

$$(IV) \quad (\forall x)(\forall y)[(\exists i)(\mathbf{Rid}_f(i, x) \wedge \mathbf{Lid}_f(i, y)) \longleftrightarrow f(y, x) \neq \perp]$$

Where  $\mathbf{Lid}_f(x, y)$  and  $\mathbf{Rid}_f(x, y)$  stand for the unquantified versions of axioms (II) and (III), respectively. Associativity is then what we expect it to be:

$$(V) \quad (\forall x)(\forall y)(\forall z)[f(f(x, y), z) = f(x, f(y, z))]$$

Finally, we define  $\mathbf{Lid}$  and  $\mathbf{Rid}$  in a  $\perp$ -free way by restricting everything to just the objects falling under a particular predicate  $P$ :

$$\begin{aligned} \mathbf{Lid}_{P,f}(a, b) &:= \left[ Pb \wedge f(b, a) = a \wedge (\forall z)(f(b, z) = z \vee \neg Pf(b, z)) \right. \\ &\quad \left. \wedge (\forall z)(f(z, b) = z \vee \neg Pf(z, b)) \right] \\ \mathbf{Rid}_{P,f}(a, b) &:= \left[ Pb \wedge f(a, b) = a \wedge (\forall z)(f(z, b) = z \vee \neg Pf(b, y)) \right. \\ &\quad \left. \wedge (\forall z)(f(b, z) = z \vee \neg Pf(b, z)) \right] \end{aligned}$$

Extending this trick, we can express all five axioms in a  $\perp$ -free way, giving us

$$\begin{aligned} (I) \quad &(\forall x)(\forall y)[\neg Py \longrightarrow (\neg Pf(x, y) \wedge \neg Pf(y, x))] \\ (II) \quad &(\forall x)\{Px \longrightarrow (\exists y)[\mathbf{Lid}_{P,f}(x, y)]\} \\ (III) \quad &(\forall x)\{Px \longrightarrow (\exists y)[\mathbf{Rid}_{P,f}(x, y)]\} \\ (IV) \quad &(\forall x)(\forall y)\{(Px \wedge Py) \longrightarrow [(\exists i)(\mathbf{Rid}_{P,f}(i, x) \wedge \mathbf{Lid}_{P,f}(i, y)) \longleftrightarrow Pf(y, x)]\} \\ (V) \quad &(\forall x)(\forall y)(\forall z)[(Pf(x, y) \wedge Pf(y, z)) \longrightarrow f(f(x, y), z) = f(x, f(y, z))] \end{aligned}$$

Define  $\mathbf{ax}(f, P)$  (intuitively,  $f$  defines a category over  $P$ ) to be the conjunction of the axioms. For “the” category defined by  $f$ , we then define

$$\mathbf{Ax}(f) := (\exists P)(\mathbf{ax}(f, P) \wedge (\forall Q)(\mathbf{ax}(f, Q) \longrightarrow (\forall x)(Qx \longrightarrow Px)))$$

That is, we define  $f$  to be a “category mapping” just when there is a unique maximal set on which  $f$  behaves like a category.

1	$\text{Lid}_{P,f}(a, b) \wedge Pa$	
2	$Pb$	Unpacking, 1
3	$f(b, a) = a$	Unpacking, 1
4	$(\forall z)(f(b, z) = z \vee \neg Pf(b, z))$	Unpacking, 1
5	$f(b, b) = b \vee \neg Pf(b, b)$	Inst. at $z = b$ , 4
6	$\neg Pf(b, b)$	Assump. for Reductio
7	$f(f(b, b), a) = f(b, f(b, a))$	Axiom V
8	$f(f(b, b), a) = f(b, a)$	Replacement, 3, 6
9	$f(f(b, b), a) = a$	Replacement, 3
10	$\neg Pf(b, b) \longrightarrow (\neg Pf(a, f(b, b)) \wedge \neg P(f(b, b), a))$	Axiom I
11	$\neg P(f(b, b), a)$	Unpacking, 10
12	$\neg Pa$	Replacement, 9, 11
13	$Pa$	Assumption, 1
14	$Pf(b, b)$	Reductio, 7–10
15	$f(b, b) = b$	Disj. Syllogism, 5, 14

**Figure 2.3:** Proof that Left Identities are Loops

### 2.3.4 Building Iso

We now building the predicate Iso that corresponds, intuitively, the the notion of isomorphism between functions that represent the same category.

To begin, we introduce the abbreviation  $\text{Inv}_{P,Q}(f, g)$  (intuitively read “ $f$  and  $g$  are inverses as functions between  $P$  and  $Q$ ”) for

$$(\forall x) \left( [Px \longrightarrow (Qf(x) \wedge g(f(x)) = x)] \wedge [Qx \longrightarrow (Pg(x) \wedge f(g(x)) = x)] \right).$$

Recall a functor is a composition-preserving map from morphisms to morphisms and objects to objects. In a category without objects, a functor is thus simply a map that preserves composition. If we abbreviate “ $h$  is a functor from the category  $(P, f)$  to the

category  $(Q, g)$ ” by  $\text{Fun}_{P,Q}^{f,g}(h)$ , then we have

$$\text{Fun}_{P,Q}^{f,g}(h) := (\forall x)(\forall y) \left( (Px \wedge Py) \longrightarrow [Qh(x) \wedge Qh(y) \wedge h(f(x, y)) = g(h(x), h(y))] \right)$$

To make things easier later, we also introduce the following abbreviation:

$$\text{Mdom}(P, f) := \text{ax}(P, f) \wedge (\forall Q)[\text{ax}(f, Q) \longrightarrow (\forall x)(Qx \longrightarrow Px)]$$

Intuitively we read  $\text{Mdom}(P, f)$  as “ $P$  is the maximal domain on which  $f$  behaves like a category.” The phrase “ $f$  and  $g$  define isomorphic categories,” abbreviated  $\text{Iso}(f, g)$ , is then defined by

$$\begin{aligned} (\forall P)(\forall Q) \left( [\text{Mdom}(P, f) \wedge \text{Mdom}(Q, g)] \longrightarrow \right. \\ \left. (\exists h)(\exists j)[\text{Fun}_{P,Q}^{f,g}(h) \wedge \text{Fun}_{Q,P}^{g,f}(j) \wedge \text{Inv}_{P,Q}(h, j)] \right) \end{aligned}$$

### 2.3.5 Defining Cat

Given all of the preceding technical work, we can now (re-)introduce the category abstraction principle, which looks just as we expected it would:

$$\mathbf{BLC} := (\forall f)(\forall g) \left[ \text{Cat}(f) = \text{Cat}(g) \longleftrightarrow \left( (\neg \text{Ax}(f) \wedge \neg \text{Ax}(g)) \vee \text{Iso}(f, g) \right) \right]$$

Of course, **BLC** cannot be modeled in any model of set theory as it stands – Iso is simply too fine-grained a way of discriminating among two-place functions. I give an argument modeled on the Burali-Forti paradox to demonstrate this.<sup>9</sup>

Suppose we are given a domain  $D$  with cardinality  $\kappa$  that might model **BLC**. Each ordinal  $\lambda$  with cardinality no more than  $\kappa$  defines a poset category with no more than  $\kappa$  morphisms (this follows from the combinatorial calculation above). For distinct ordinals these categories are non-isomorphic, so the cardinality of the class of categories produced by this abstraction operator (more carefully – of the class of distinct objects that must

---

<sup>9</sup> This demonstration that **BLC** is inconsistent is, of course, somewhat suspect, being only a demonstration that it simply has no models in *set theory*. Better would be to derive an actual contradiction from **BLC**. To do this, however, we would have to specify a system of axioms for second-order logic, and “unpack” the definitions – both of which would be rather large endeavors. Conveniently, since **BLC** was meant to capture the notion of a category given as a type of *mapping*, and since we are operating in a set-theoretic metatheory, the demonstration that no set theory can model it will suffice to convince the reader we should seek some emendation of the principle.



be in the range of the  $\text{Cat}$  operator when modeled in  $D$ ) is at least as large as the cardinality of the class  $S$  of ordinals with cardinality less than or equal to  $\kappa$ . But  $S$ , then, is itself an ordinal, and since  $S \notin S$ ,  $S$  has cardinality greater than  $\kappa$ .

## 2.4 “New” C

Since the abstraction principle we’ve produced treats categories like a type of function, the tricks the neologicists have used when constructing consistent “NewV”-type abstraction principles seem to not apply. So at first blush we might think we need some exciting new theory of function abstraction on hand in order to build a consistent version of **BLC**. But from a category-theoretic perspective, there are two other approaches worth trying first: we could look at what happens if we restrict the abstraction principle so all the categories it produces only have a “small” number of objects, or we could restrict so that between any two objects there is only a “small” number of morphisms. I show now that neither of these suffice alone to allow a consistent category-abstraction principle, even if “small” means “of cardinality one.”<sup>10</sup>

Regarding the first case, recall that a *monoid* is a set equipped with an associative law of composition that admits an identity. It is clear on inspection that there is a bijection between isomorphism classes of monoids and isomorphism classes of categories with one object. Thus, in any domain, there are as many isomorphism classes of monoids modelable in that domain as there are isomorphism classes of categories with one object. However, in a domain of cardinality  $\kappa$ , there are more than  $\kappa$  isomorphism classes of monoids. Thus even if we demand our categories have only a single object, **BLC** will still be unsatisfiable.

On the other hand, if we put no restriction on the objects at all and only demand the set of morphisms between any two objects be small, then even if small means “of cardinality no more than one” we still get too many categories. To see this, recall the poset category associated to an ordinal has, between any two objects, either one morphism or none. But, as above, there are too many ordinals on any given domain to introduce an object for each of them. Since the ordinals are among the posets, this suffices to show there are too many posets on a given domain to introduce an object for

<sup>10</sup> I remind the reader that the metatheory throughout is set-theoretic.

each of them.

So any restriction of this general type must restrict *both* the total number of objects and the total number of morphisms that occur between any two objects. In morphisms-only talk, that is, we must restrict both the number of identity and number of non-identity morphisms. A survey of the category theory literature, of course, reveals that this is not a surprising fact – the standard definition of a “small category” is a category with *both* a small collection of objects and a small collection of morphisms – that is, a category for which both the collection of objects and the collection of morphisms are actually *sets*. A similar condition that suffices to give a consistent version of **BLC** is that the total number of morphisms in the category be small, as we now demonstrate.

To this end, we introduce a defined predicate **SMD** that holds of functions. Intuitively,  $\text{SMD}(f)$  means  $f$  has a small maximal domain on which it behaves like a category. For various interpretations of “small,” then, we can introduce “NewC” principles of the form

$$(\forall f)(\forall g) \left[ \text{Cat}(f) = \text{Cat}(g) \longleftrightarrow \left( (\neg(\text{Ax}(f) \wedge \text{SMD}(f)) \vee \neg(\text{Ax}(g) \wedge \text{SMD}(g))) \vee \text{Iso}(f, g) \right) \right]$$

For appropriate “smallnesses”, this principle is satisfiable. I close the technical portion of the paper by demonstrating this by presenting two interpretations of “small” that allow us to easily see NewC, with the appropriate interpretation of “small”, can be modeled.

### 2.4.1 Small=Finite

Suppose “small” means finite, so  $\text{SMD}(f)$  means the maximal domain on which  $f$  behaves like a category is finite. Call the resulting principle **BLC<sub>Fin</sub>**. Then if the cardinality of the universe is  $\kappa$ , and we let  $\xi$  be the cardinality of the set of distinct objects in the range of the **Cat** operator, we have that

$$\xi = \sum_{i=0}^{\infty} \kappa^{i^2} \leq \aleph_0 \cdot \kappa$$

So, provided the universe is infinite, **BLC<sub>Fin</sub>** can be modeled.

### 2.4.2 Small=Exponentially Small

Recall the phrase “exponentially small” as defined by Fine:

“A cardinal  $\mathbf{c} \leq \mathbf{m}$  is said to be *exponentially small (relative to  $\mathbf{m}$ )* if  $\mathbf{m}^{\mathbf{c}} \leq \mathbf{m}$ ” [30]

Suppose “small” means exponentially small relative to the cardinality of the universe. Call the resulting abstraction principle  $\mathbf{BLC}_{ExpSml}$ . If the cardinality of the universe is  $\kappa \geq \aleph_0$ , and we let  $\xi_\lambda$  be the number of objects in the range of the Cat operator whose maximal domain has cardinality  $\lambda$ , we get (for infinite  $\lambda$ )

$$\begin{aligned} \xi_\lambda &\leq (\text{number of } \lambda\text{-sized subsets of the universe})(2^{\lambda^2}) \\ &\leq \kappa^\lambda 2^\lambda \\ &\leq (\kappa^\lambda)^2 \\ &\leq \kappa \end{aligned}$$

So, if we again let  $\xi$  be the cardinality of the set of objects in the range of the Cat operator, then

$$\begin{aligned} \xi &\leq \left( \sum_{\text{Exponentially Small Infinite } \lambda} \xi_\lambda \right) + \left( \sum_{\text{Finite } \lambda} \xi_\lambda \right) \\ &\leq \left( \sum_{\text{Exponentially Small Infinite } \lambda} \kappa \right) + \kappa \\ &\leq \kappa \end{aligned}$$

So  $\mathbf{BLC}_{ExpSml}$  can be modeled.

## 2.5 What Remains to Be Done

So we have an abstraction principle that implicitly defines objects that correspond to equivalence classes of two-place functions that represent the composition-function of a category. This should provide the basis for an abstractionist category theory. What remains is to prove analogs to Frege’s Theorem and Boolos’ Theorem. That is, given

some axiomatization  $A$  of the theory of categories, what we need is a bi-interpretability result regarding the theories based, respectively, on one of the NewC-type principles and on  $A$ . Unfortunately, there is no *standard* axiomatization of the theory of categories, so this is a rather difficult task.

Specifically, what would be needed in order to prove analogs to Frege’s and Boolos’ theorems is a *theory* in the sense of a deductively closed set of sentences, and further a theory that specifies precisely which categories exist and what (category-theoretically relevant) features they have. As [11] points out, however, “as usually presented, category theory ... *lacks substantive axioms of mathematical existence.*”<sup>11</sup> (emphasis in original)

One *could* attempt to use either the theory given in [2] or the one given in [32] (or modifications of these) to provide such a theory of existence. There are two problems with this, however. First, both of these theories have suffered some criticism. For example, regarding [2], Hellman points out that

“Doubts as to its consistency were raised early on ([33]), and, indeed, the stronger axioms of the theory are quite complex. Apparently, subsequent developments in topos theory overshadowed efforts to axiomatize a category of categories, but at the cost of suspending the articulation of existence axioms in the assertory sense, i.e., as credible truths outright rather than as merely part of an algebraico-structural definition. A detailed assessment of Lawvere’s axioms would examine this issue of credibility, which in turn rests on some prior, not merely structural, understanding of the primitives and intended interpretation, which does build in the notion of ‘category’ itself, and thereby presupposes the notion of ‘collection’ as well as that of ‘functor’. (Cf. [34].)” [11, p. 137]

On the other hand, regarding [32], Feferman has pointed out that

“it fails as an actual foundation of working category theory since we cannot speak within it of such categories as that of all small groups, all small topological spaces, and even the category of all small categories” [35, p. 8]

---

<sup>11</sup> This much is even granted by the otherwise-quite-critical [31].

Despite their seriousness, let us put aside these criticisms for the time. The second problem still remains: both [2] and [32] axiomatize *categories of* categories. As a result, to the extent that they give us theories of existence at all, they give theories of existence for categories *and functors*. One might hypothesize that the restriction of one of these theories (or a modification of it) to those claims it makes about categories alone would suffice as a theory of categories. But this, of course, is no more than an hypothesis, and a fairly grand one at that – I’d like to see the result before putting faith in it.

All this aside, some general details of how to translate “category-talk” into “NewC-talk” are nonetheless perfectly clear. To be as explicit as possible, let  $T_{\text{NewC}}$  be the second-order theory based on a NewC-type abstraction principle, and let  $T_A$  be a theory based on an appropriate axiomatization of category theory. Claims of the form “All morphisms in category  $X$  satisfying condition  $\phi$  have feature  $\psi$ ” occurring in  $T_A$  will be translated into  $T_{\text{NewC}}$  as

$$\forall f(X = \text{Cat}(f) \rightarrow ((a \in \text{Mdom}(f) \wedge T(\phi)(a)) \rightarrow T(\psi)(a)))$$

Where  $T$  is an appropriate translation from the language of  $T_A$  to the language of  $T_{\text{NewC}}$ .<sup>12</sup> A similar translation will work for claims of the form “there is a morphism in  $X$ ...”.

To make this procedure more concrete and to produce specific translations to and from  $T_{\text{NewC}}$ , we would need, again, to settle on a specific axiomatization of the theory of categories. It should be clear from the sketch just provided and from the examples given by Frege and Boolos’ Theorems roughly how such a procedure would go, were we equipped with a generally accepted axiomatization of category theory.

Finally, a last bit of work that remains is the determination of precisely *which* NewC-type abstraction principle is appropriate. Determining this will have to be done with an eye to the general size of categories in fact used in category-theoretic practice, a topic that is, at this point, the subject of further investigation (see e.g. [36]).

---

<sup>12</sup> In general, we should note, discussion of the behavior of *particular* arrows in a category will not be possible unless that arrow can be singled out as the unique morphism satisfying some condition  $\phi$ . But this is to be expected – we are providing a neologicist version of the theory of categories, not the theory of morphisms, and morphisms are only defined relative to the category they are in.

## Chapter 3

# Abstractionist Categories of Categories

If  $\mathcal{C}$  is a category whose objects are themselves categories, and  $\mathcal{C}$  has a rich enough structure, it is known that we can recover the internal structure of the *categories* in  $\mathcal{C}$  entirely in terms of the *arrows* in  $\mathcal{C}$ . In this sense, the internal structure of the categories in a rich enough category of categories is visible in the structure of the category of categories itself.

In this paper, we demonstrate that this result follows as a matter of logic – given one starts from the right definitions. This is demonstrated by first producing an abstraction principle whose abstracts are functors, and then actually recovering the internal structure of the individual categories that intuitively stand at the sources and targets of these functors by examining the way these functors interact. The technique used in this construction will be useful elsewhere, and involves providing an abstract corresponding not to every *object* of some given family, but to all the relevant *mappings* of some family of objects.

This construction should shed light, in particular, on questions about whether categories of categories qualify as *autonomous* mathematical objects – *if* admitting a neologicist definition qualifies a mathematical object as autonomous, *then* categories of categories are perfectly acceptable autonomous objects and thus, in particular, suitable for foundational purposes.

## 3.1 Introduction

One thing that stands out as a difference between a structure like a large category of small categories compared to a structure like a class of sets is the apparently greater amount of *internal* structure the objects inside the former have. In particular, it seems as if the objects (small categories) out of which the large category is formed have sufficient structure that one would need a fair amount of mathematics *already in place* in order to describe them. This very fact has occasionally led some writers (see, for example, [8], [10]) to label categories of categories as unsuitable for foundational purposes. As pointed out in [1], these objections centrally concern category theory’s *autonomy* – to understand what a category of categories *is*, it is claimed, you need to understand their internal structure and to understand this, in turn, an alternative mathematical theory is needed. So category theory is non-autonomous, thus unsuitable for foundational purposes.

In the technical construction that follows, we demonstrate that all internal structure categories are supposed to have can be recovered in terms of external features embodied by categories of categories, and – the novel part of this contribution – that this can be done as a matter of *logic*. We turn, in the remainder of the introduction, to briefly outlining the abstractionist programme.

### 3.1.1 Abstractionism

An abstraction principle is a sentence with the following syntactic form:

$$\forall A \forall B (\Phi(A) = \Phi(B) \iff E_f(A, B))$$

Where  $A$  and  $B$  are of the same logical type  $t$  and  $E_\Phi$  is a 2-place predicate that names an equivalence relation on entities of type  $t$ .  $\Phi$ , in turn, is what is usually called a  $t$ -functor. That is,  $\Phi$  is a syntactic object that is saturated by an entity of type  $t$  and that, when so saturated, becomes a name for an object. An example of such a (intuitively unnatural sounding) thing is given by the phrase “the number of”. This phrase is a syntactic object that is saturated by a concept and that, when so saturated, becomes a name of a number.

Of course, given our project, we will have need later of the word “functor” to mean something quite different from this. So we will adopt a word other than “functor” to

refer to the type of thing  $\Phi$  is. We will use “correspondence” to play this role, so  $\Phi$  will be called a “correspondence between entities of type  $t$  and objects” rather than a “ $t$ -functor” in what follows.

Now, given a language  $\mathcal{L}$  together with a consequence relation  $\models$  holding among the sentences of  $\mathcal{L}$ , and a particular abstraction principle  $A$  in the language of  $\mathcal{L}$ , we can consider the theory  $T_A = \{\phi \in \mathcal{L} : A \models \phi\}$ . Let  $M$  be a set-based model of  $T_A$ . Then in  $M$  the correspondence  $\Phi$  will be modeled by a function  $F_M$  from entities of type  $t$  to the domain of  $M$ . Further, since  $A$  is certainly a part of  $T_A$  and  $M$  models  $T_A$ ,  $M$  models  $A$ . Thus  $F_M$  maps two entities of type  $t$  to the same object of the domain exactly when the two entities satisfy the equivalence relation  $E_\Phi$ . The abstraction principle thus provides something very like identity conditions for the objects occurring in the range of  $F_M$ .<sup>1</sup> These identity conditions, in turn, provide an *implicit definition* of these objects – an implicit definition that, importantly, relies for its intelligibility only on the intelligibility of the vocabulary used in forming  $E_f$ . Thus, if we can produce an abstraction principle wherein the vocabulary occurring in  $E_f$  has a privileged epistemic status – if we can come to know, for example, *a priori*, or *analytically*, what sentences in that vocabulary *mean* – then the fact that there is such an implicit definition can help to explain the privileged epistemic status of our knowledge regarding the objects defined by that abstraction operator.<sup>2</sup>

Examining an example will help make this more clear. Consider for now the abstraction principle HP:

$$\text{HP} : \forall X \forall Y (\#(X) = \#(Y) \iff X \text{ bijects with } Y)^3$$

Let  $T_{\text{HP}}$  be the second-order theory that takes HP as its lone axiom. What is known as Frege’s Theorem (see, e.g. [22, 23]) tells us that there is a translation of the language of Second Order Peano Arithmetic ( $\text{PA}^2$ ) into the language of  $T_{\text{HP}}$  so that the translation of every  $\text{PA}^2$ -theorem is a  $T_{\text{HP}}$ -theorem. What is known as Boolos’ Theorem (see, e.g.

---

<sup>1</sup> A good exposition of the neologicist programme that emphasizes the precise nature of the identity conditions they provide is found in [19]. The remainder of this section is adapted from [37]

<sup>2</sup> Further discussions of abstractionism, the status of abstraction principles as implicit definitions, and the role of second order logic can be found in [20] and [21]

<sup>3</sup> Note that a 2-place “bijects with” predicate can be encoded entirely in second-order logic.



[24]) gives the converse result: there is a translation from the language of  $T_{\text{HP}}$  to the language of  $\text{PA}^2$  such that the translation of every  $T_{\text{HP}}$ -theorem is a  $\text{PA}^2$ -theorem. So HP serves as an implicit definition of numbers (more carefully: of Peano-numbers, or Peano-cardinals) that relies for its intelligibility only on the intelligibility of the vocabulary of second-order logic.

HP is the great success story for what we will call the *abstractionist programme*: the attempt to find abstraction principles from which the theorems in important branches of mathematics follow. Unfortunately, HP is, in many respects, an *unusual* abstraction principle. To see this, consider another abstraction principle, BLV that seems, at first glance, as acceptable as HP:

$$\text{BLV} : \forall X \forall Y (\S(X) = \S(Y) \iff \forall z (Xz \longleftrightarrow Yz))$$

BLV is well-known to be inconsistent. With a bit of tinkering, one can easily see that in a (standard) model of BLV in a domain  $D$ , the function  $\S$  would be an injection from  $2^D$  – the set of all subsets of the  $D$  – into  $D$ . Such an injection is of course impossible, by Cantor’s theorem. This is a typical problem with abstraction principles: if the equivalence relation  $E_{\Phi}$  is too *fine grained*, then in no model  $M$  can there be a function  $F_M$  that satisfies every instance of the abstraction principle’s biconditional.

Conveniently, for BLV there is a partial workaround for this problem: by appropriately restricting the size of the concepts  $X$  and  $Y$  for which we demand the biconditional hold, one can produce principles (typically referred to in the literature (e.g. [25]) as “NewV”-type principles) that are consistent and which give rise to relatively powerful versions of set theory. So NewV-type principles serve as implicit definitions of (sufficiently small) sets, and rely only on the intelligibility of the vocabulary of second-order logic for their own intelligibility.

These successes motivate further examination of abstraction principles. One needn’t have logicist inclinations to find such investigations interesting. From a metaphysical and mathematical point of view, abstraction principles “work” by proposing identity conditions for abstract objects. In the cases that are of interest one can prove – using higher-order logic and the specified identity conditions – that the objects implicitly defined by an abstraction principle have all the properties they are “supposed to” have.

When this is possible, we have reason to believe the identity conditions proposed specify something like the essential features of the objects in question. It is thus a generally worthwhile project to find, for various “natural families” of mathematical objects, an appropriately-constructed abstraction principle that realizes them as its abstracts.

Of course, this is rather difficult in general. One difficulty is that not all “natural families” of mathematical objects are well-characterized by identity conditions on the objects themselves. For some families of objects, the objects themselves are best characterized by instead examining the relations among the other members of the family. A good example of such “malcontent” objects are categories – one *can* produce an abstraction principle whose abstracts are categories (this is done in [37], which is briefly reviewed below). However, (set-based) models of such an abstraction principle contain, rather than a *category* of all small-enough categories (which would be natural), a *class* of all small-enough categories. Since the natural family within which categories occur is the category of categories, it would be better from a philosophical perspective to have an abstraction principle whose models contained a category of categories, rather than only a class of categories. In this note we propose such an abstraction principle. The abstraction methodology used for this purpose is in fact interesting on its own, so worth briefly commenting on.

Our starting point is the observation that, in a rich-enough (large) category of (small) categories, the internal structure of individual small categories can be read off of the external structure exhibited by the arrows in the large category. Intuitively, this means that even if we can’t “see” inside the small categories, we can still tell what’s “happening” there just by examining the way a given small category interacts with the other small categories. This provides the impetus for pursuing what I call a “relative” abstraction theory for categories: rather than finding an abstraction principle whose abstracts correspond to the categories *directly*, our goal is to find an abstraction principle whose abstracts correspond to functors, and then to recover the structure of individual categories by examining the way the functors interact.<sup>4</sup> Of course, there is nothing particularly special about categories that makes them amenable to such a programme of relative abstraction – many other mathematical objects are susceptible to

---

<sup>4</sup> Such a program is *relative* in the sense of Grothendieck’s *relative point of view* – to study object  $x$  in category  $\mathcal{C}$ , study the slice categories  $\mathcal{C} \downarrow x$  of maps to  $x$  and  $\mathcal{C} \uparrow x$  of maps from  $x$ .

similar treatment; relative abstraction for categories is simply one example of a possibly quite general method.

In the next section we ease into relative abstraction by looking at an easier case – mappings among sets. This provides us with the outlines of how we expect relative abstraction for categories to go. In Section 3, we then set about producing abstraction principles whose abstracts are (intuitively at least) functors. We then demonstrate that the identity functors in the range of these abstraction operators behave like categories, and then prove a few theorems about the categories of categories arising from the abstraction principles of the type produced.

## 3.2 Relative Abstraction Theory

In just a moment I am going to propose an abstraction principle. Whenever one does this, the lesson of BLV looms: one must ensure the equivalence relation being used is not too fine-grained, or inconsistency results. However, for the purposes of this introductory section, we will ignore inconsistency – whatever that means – and proceed naïvely. We will also be somewhat naïve in this example about the language in which our abstraction principle is formulated. I will specify only that the language has variables whose type is a triple consisting of a function and two properties, and that the abstraction operator is a correspondence from entities of this type to objects.

This section is meant as a propaedeutic to the functorial abstraction theory found in section three. Here is the intuitive setup: in many contexts it is useful to distinguish between functions that have the same domain and the same value on each element of that domain, but which have different codomains. Let’s reserve the word “map” for functions that we distinguish in this way. For example, it is as *maps* that we distinguish the identity function on the natural numbers from the inclusion of the natural numbers into the integers. For simplicity, let’s also focus only on the case of one-place maps that are defined on the entirety of their domain.

Let  $(f, P, Q)$  be a *variable* whose type is a triple consisting of a one-place function and two concepts. Employing several abbreviations that are sufficiently clear from context (and, again, ignoring issues that would arise from this principle’s likely inconsistency), the following abstraction principle (which we will call BLM) will have as its abstracts

objects that correspond to maps (in the technical sense of that word adopted above).

$$\forall(f, P, Q)\forall(g, R, S) \left[ m(f, P, Q) = m(g, R, S) \Leftrightarrow \left( \neg[P \xrightarrow{f} Q \vee R \xrightarrow{g} S] \vee \right. \right. \\ \left. \left. [P \xrightarrow{f} Q \wedge R \xrightarrow{g} S \wedge \forall x([Px \Leftrightarrow Rx] \wedge [Qx \Leftrightarrow Sx]) \wedge \forall y(Py \Rightarrow f(y) = g(y))] \right) \right]$$

In standard set-based models of BLM, the correspondence  $m$  will be modeled by a function  $M$  from triples  $(f, \mathcal{P}, \mathcal{Q})$  – consisting of a one-place total function  $f$  on the domain of the model and two subsets  $\mathcal{P}$  and  $\mathcal{Q}$  of the model – to elements of the domain of the model.<sup>5</sup> Given two such triples  $(f, \mathcal{P}, \mathcal{Q})$  and  $(g, \mathcal{R}, \mathcal{S})$ ,  $M$  will send them to the same element of the domain exactly when  $\mathcal{P} = \mathcal{R}$ ,  $\mathcal{Q} = \mathcal{S}$  and  $f$  agrees with  $g$  everywhere on  $\mathcal{P}$ . It seems natural, given this, to say the objects that occur in the range of  $M$  simply *are* mappings in our technical sense.

But there is a problem here: the objects that occur in the range of  $M$  are, in an important sense, *simples*. They do not “come with” domains or ranges or any obvious way to apply to objects. Adopting BLM demands only that the things denoted by the terms  $m(f, P, Q)$  be objects and that two such terms denote one and the same object just when certain other equalities hold. To justify calling these objects “maps,” then, we must first ensure that BLM alone (or, rather, BLM plus logic) ensures they exhibit the appropriate sorts of behaviors.

This task, in turn, seems at first glance to require defining another abstraction principle to capture sets, since part of what we want to ensure is that each map has a set as its domain and another as its range. But after a bit of poking about it becomes clear that BLM all by itself equips us with objects that can play this role – that is, with objects that behave just like sets. These set-surrogates are defined as follows:

Let’s suppose we have sufficient functional comprehension at our disposal so that  $\exists f \forall x (f(x) = x)$  is true. Then we can add a defined constant  $\text{id}$  to our language by simply translating, for any context  $\phi$ , the phrase “ $\phi(\text{id})$ ” as “ $\forall f (\forall z (f(z) = z) \rightarrow \phi(f))$ ”. We can then define the terms “Set” and “ $\in$ ” in the theory  $T$  that takes (a consistent version of) BLM as its lone axiom in the following way:

---

<sup>5</sup> Notice  $M$  may be a class-function and not necessarily a set-function.

1.  $\mathbf{Set}(x) := \exists P(x = m(\mathbf{id}, P, P))$
2.  $y \in x := \mathbf{Set}(x) \wedge \forall P(x = m(\mathbf{id}, P, P) \rightarrow Py)$

Using these definitions, we see that the theory of mapping abstraction contains within it a relative theory of set abstraction – we’ve recovered *sets* as the identity *maps*. One should, of course, ensure everything works at this point (and go back to take care of the pesky issue of inconsistency!). But since (a) it’s relatively clear what to do in these regards (e.g. check that for each mapping there is a set that serves as its domain, another that serves as its range, etc.) and (b) doing these things would take us too far afield, we instead leave it as an exercise for the reader.

And, on that note, we conclude our propaedeutic. In the next section we get down to the business of defining a relative abstraction theory for categories. It is worth noting that a category, as we will understand the term here, is simply a collection of arrows with an associative law of composition that admits left and right identities.<sup>6</sup> Functors, in turn, are then no more than maps between these that commute with the compositions. Taking a page from the construction just examined, a theory of functorial abstraction should associate objects that we will call functors to equivalence classes of one place total functions distinguished by what their domains and codomains are taken to be and how they treat the specified categorial structure in their domains. We expect to be able to use a *functorial* abstraction principle to recover individual *categories* by seeing them as *identity functors*. Thus, what follows will be a theory of *relative* categorial abstraction analogous to the theory of *relative* set abstraction just gestured towards. The major difference to be accounted for here is that categories have a great deal more “internal” structure than sets.

### 3.3 Generators and Functorial Abstraction

It is a theorem of elementary category theory that the poset category corresponding to the ordinal 2 is a generator for the category of all small categories. Informally, this means distinct parallel functors can always be distinguished by how they compose

---

<sup>6</sup> This relies on the well-known object-free definition of a category, where the word “object” is a defined term equivalent to the phrase “identity arrow.”

with maps from the ordinal category **2**. [4], contains the following formalization of this theorem:<sup>7</sup>

$$\forall \mathcal{A} \forall \mathcal{B} \forall f \forall g [\mathcal{A} \xrightarrow{f} \mathcal{B} \wedge \mathcal{A} \xrightarrow{g} \mathcal{B} \wedge \forall u (\mathbf{2} \xrightarrow{u} \mathcal{A} \Rightarrow fu = gu) \Rightarrow f = g]$$

Those of the right disposition will be struck immediately by how close this sentence is to having the syntactic form of an abstraction principle. Using the partial outline of a theory of mapping abstraction given above as a framework, we turn this almost-abstraction-principle into the following actual abstraction principle (it's written on several lines here so that it's easier to analyze later):

$$\text{(Line 1) } \forall \phi \forall \psi \left[ \mathcal{F}(\phi) = \mathcal{F}(\psi) \Leftrightarrow \right.$$

$$\text{(Line 2) } \left( \neg \left[ \mathbf{Fr1} \phi \vee \mathbf{Fr1} \psi \right] \vee \right.$$

$$\text{(Line 3) } \left[ \mathbf{Fr1} \phi \wedge \mathbf{Fr1} \psi \wedge \mathbf{Src} \langle \phi \rangle = \mathbf{Src} \langle \psi \rangle \wedge \mathbf{Tgt} \langle \phi \rangle = \mathbf{Tgt} \langle \psi \rangle \wedge \right.$$

$$\text{(Line 4) } \left. \left. \left. \exists T (\mathbf{2}T \wedge \forall \mu [(\mathbf{Fr1} \mu \wedge \mathbf{Src} \langle \mu \rangle = T \wedge \mathbf{Tgt} \langle \mu \rangle = \mathbf{Src} \langle \phi \rangle) \Rightarrow \mathbf{Fnc} \langle \phi \rangle \mathbf{Fnc} \langle \mu \rangle \equiv \mathbf{Fnc} \langle \psi \rangle \mathbf{Fnc} \langle \mu \rangle] \right) \right] \right]$$

We will call this abstraction principle **BLF**. Of course, **BLF** is useless to us as it stands since its right hand side (Lines 2-4) is mostly goofy-looking symbols without standard definitions. The next section will be spent providing definitions that will allow us to translate the entire right-hand side of **BLF** into a single sentence containing two free variables in a fourth-order language. However, before diving into the technicalities, an intuitive overview of **BLF** will help set the stage, beginning with Line 1.

Roughly, the variables  $\phi$  and  $\psi$  will range over quintuples of the form

⟨unary function, binary function, unary predicate, binary function,  
unary predicate⟩

---

<sup>7</sup> Lawvere's functions take their arguments from the left, so this is actually a slight syntactic modification of the sentence from his dissertation.

Let us say that variables of this form have type  $t$ . Since a category is, for us, no more than a binary function specifying how arrows from a specified family compose, we can think of entities of type  $t$  as actually being triples of the form

$$\langle \text{unary function, category, category} \rangle$$

The language we will define includes logical functions **Fnc**, **Src**, and **Tgt** defined so that if  $\phi$  is of type  $t$ , **Fnc** $\langle\phi\rangle$  is the unary function component of  $\phi$ , **Src** $\langle\phi\rangle$  is the first category in  $\phi$ , and **Tgt** $\langle\phi\rangle$  is the second category in  $\phi$ . Intuitively, **Fnc** stands for “function”, **Src** for “source”, and **Tgt** for “target”. The language will also contain a defined predicate **Fr1** such that if  $\phi$  has type  $t$ , then **Fr1**  $\phi$  holds if and only if the function **Fnc** $\langle\phi\rangle$  respects the categorial structure in the categories **Src** $\langle\phi\rangle$  and **Tgt** $\langle\phi\rangle$ . Thus **Fr1**  $\phi$  intuitively means that  $\phi$  behaves functorially. Last, there is a defined predicate **2** holding of those pairs of the form  $\langle \text{binary function, unary predicate} \rangle$  that intuitively represent a category isomorphic to the ordinal category **2** corresponding to the ordinal 2. Now, going through the remaining lines of BLF, we see that we will associate the same *functor* with two objects  $\phi$  and  $\psi$  of type  $t$  just when

**(Line 2)**  $\phi$  and  $\psi$  both fail to behave functorially; or

**(Line 3)**  $\phi$  and  $\psi$  *do* both behave functorially, and further have the same source and target, and

**(Line 4)** There is a category  $T$  isomorphic to the ordinal category **2** so that for all functors  $\mu$  with source  $T$  and whose target is the same as the source of  $\phi$  or  $\psi$  (and can hence compose with  $\phi$  or  $\psi$ ), the composition of the functor-part of  $\phi$  with the functor-part of  $\mu$  is equal everywhere on  $T$  to the composition of the functor-part of  $\psi$  with the functor-part of  $\mu$ .

With this intuitive picture in mind, we now turn to the technicalities.

### 3.4 Language and Definitions

We first deal with syntax. The language  $\mathcal{L}$  is an extension of the language of full *third-order* logic. It has, in particular, variables for and quantifiers ranging over entities of the following types:

- Object-type, for which we use the lowercase roman letters  $x, y, z$ .
- Predicate-type (unary suffices), for which we use the uppercase roman letters  $P, Q, R$ .
- Function-type (unary and binary), for which we use the lowercase roman letters  $f, g, h$ .
- A type consisting of pairs  $\langle \text{binary function, unary predicate} \rangle$ , for which we use the uppercase roman letters  $S$  and  $T$ . We will say the things these variables range over have the type **Cat**
- A type consisting of quintuples

$\langle \text{unary function, binary function, unary predicate, binary function, unary predicate} \rangle$ .

For these variables we will use lowercase greek letters  $\phi, \psi, \mu$  and  $\tau$ . We will say the things these variables range over have type **Fun**.

The reader who is concerned about the high order of this language will probably be able to tell, in the course of the construction that follows, that it is not strictly necessary. Most of what is done here could be carried out in a much less rich language, though it's unclear if the result of a reduction to as lean a language as possible would be intelligible.

We also assume the following logical operators:

- Two correspondences **Src** and **Tgt** between the type-**Fun** things and the type-**Cat** things, defined so that

$$\mathbf{Src}\langle\phi\rangle = \mathbf{Src}\langle f, g, P, h, R \rangle = \langle g, P \rangle$$

$$\mathbf{Tgt}\langle\phi\rangle = \mathbf{Tgt}\langle f, g, P, h, R \rangle = \langle h, R \rangle$$

- A correspondence **Fnc** between the type-**Fun** things and the things of function-type, defined so that

$$\mathbf{Fnc}\langle\phi\rangle = \mathbf{Fnc}\langle f, g, P, h, R \rangle = f$$



- A correspondence  $\mathbf{Cmp}$  between the type- $\mathbf{Cat}$  things and the things of function-type, defined so that

$$\mathbf{Cmp}\langle T \rangle = \mathbf{Cmp}\langle f, P \rangle = f$$

- A correspondence  $\mathbf{Dom}$  between the type- $\mathbf{Cat}$  things and the things of predicate-type, defined so that

$$\mathbf{Dom}\langle T \rangle = \mathbf{Dom}\langle f, P \rangle = P$$

- A composition operation  $\circ$  that maps pairs of functions to their composite in the obvious way.

Finally,  $\mathcal{L}$  contains a correspondence  $\mathcal{F}$  from entities of  $\mathbf{Fun}$ -type to objects.<sup>8</sup>

We now review a bit of the material in [37].

Let  $P$  be a concept variable and  $f$  a two-place function variable. Define the open sentences  $\mathbf{Rid}_{P,f}$  and  $\mathbf{Lid}_{P,f}$  by

$$\begin{aligned} \mathbf{Lid}_{P,f}(a, b) &:= \left[ Pb \wedge f(b, a) = a \wedge (\forall z) (f(b, z) = z \vee \neg Pf(b, z)) \right. \\ &\quad \left. \wedge (\forall z) (f(z, b) = z \vee \neg Pf(z, b)) \right] \\ \mathbf{Rid}_{P,f}(a, b) &:= \left[ Pb \wedge f(a, b) = a \wedge (\forall z) (f(z, b) = z \vee \neg Pf(b, y)) \right. \\ &\quad \left. \wedge (\forall z) (f(b, z) = z \vee \neg Pf(b, z)) \right] \end{aligned}$$

Intuitively,  $\mathbf{Lid}_{P,f}(a, b)$  (respectively,  $\mathbf{Rid}_{P,f}(a, b)$ ) says  $b$  behaves like a left (right)  $f$ -identity for  $a$  with respect to objects falling under  $P$ . In [37],  $\mathbf{Lid}$  and  $\mathbf{Rid}$  are used to help state the following axioms:

- (I)  $(\forall x)(\forall y) \left( \neg Py \Rightarrow [\neg Pf(x, y) \wedge \neg Pf(y, x)] \right)$
- (II)  $(\forall x) \left( Px \Rightarrow (\exists y) [\mathbf{Lid}_{P,f}(x, y)] \right)$
- (III)  $(\forall x) \left( Px \Rightarrow (\exists y) [\mathbf{Rid}_{P,f}(x, y)] \right)$
- (IV)  $(\forall x)(\forall y) \left( [Px \wedge Py] \Rightarrow [(\exists i) (\mathbf{Rid}_{P,f}(i, x) \wedge \mathbf{Lid}_{P,f}(i, y)) \longleftrightarrow Pf(y, x)] \right)$

---

<sup>8</sup> Note that the inclusion of this operator is what makes  $\mathcal{L}$  into a *fourth* order language.

$$(V) (\forall x)(\forall y)(\forall z) \left( [Pf(x, y) \wedge Pf(y, z)] \Rightarrow f(f(x, y), z) = f(x, f(y, z)) \right)$$

From a particular function  $g$  and concept  $S$  satisfying these axioms we recover a category  $\mathbb{S}$  whose arrows are all the objects falling under  $S$ . The composition in  $\mathbb{S}$  is given by leaving  $a \circ b$  undefined when  $g(a, b)$  is not in  $S$  and otherwise letting  $a \circ b = g(a, b)$ . Conversely, given a category  $\mathcal{C}$ , let  $g$  be any two-place total function on the universe satisfying  $g(a, b) = a \circ b$  when the latter is defined, and which sends other pairs to non- $\mathcal{C}$ -arrows.<sup>9</sup> A pair consisting of the concept that holds of all and only the arrows of  $\mathcal{C}$  and any such  $g$  will then satisfy axioms (I)-(V).

Define  $\mathbf{ax}(f, P)$  to be the conjunction of (I)-(V). Using  $\mathbf{ax}$ , we produce a defined  $\mathcal{L}$ -predicate  $\mathbf{Fr1}$  that holds of variables of type  $\mathbf{Fun}$ . Intuitively,  $\mathbf{Fr1} \phi$  says that  $\phi$  is *functorial*. This means, in particular, that  $\mathbf{Src}\langle\phi\rangle$  satisfies  $\mathbf{ax}$  and  $\mathbf{Tgt}\langle\phi\rangle$  satisfies  $\mathbf{ax}$ , and, in addition, that  $\mathbf{Fnc}\langle\phi\rangle$  respects the categorial structure in  $\mathbf{Src}\langle\phi\rangle$  and  $\mathbf{Tgt}\langle\phi\rangle$ . The following formula captures all these details:

$$\begin{aligned} \mathbf{Fr1} \phi =_{df} & \mathbf{ax}(\mathbf{Src}\langle\phi\rangle) \wedge \mathbf{ax}(\mathbf{Tgt}\langle\phi\rangle) \wedge \forall z(\mathbf{Dom}\langle\mathbf{Src}\langle\phi\rangle\rangle z \rightarrow \mathbf{Dom}\langle\mathbf{Tgt}\langle\phi\rangle\rangle z) \wedge \\ & \forall x \forall y [\mathbf{Dom}\langle\mathbf{Src}\langle\phi\rangle\rangle x \wedge \mathbf{Dom}\langle\mathbf{Tgt}\langle\phi\rangle\rangle y \rightarrow \\ & \mathbf{Fnc}\langle\phi\rangle(\mathbf{Cmp}\langle\mathbf{Src}\langle\phi\rangle\rangle(x, y)) = \mathbf{Cmp}\langle\mathbf{Tgt}\langle\phi\rangle\rangle(\mathbf{Fnc}\langle\phi\rangle(x), \mathbf{Fnc}\langle\phi\rangle(y))] \end{aligned}$$

Suppose  $T$  is of  $\mathbf{Cat}$ -type, and suppose  $\forall x(\neg \mathbf{Dom}\langle T \rangle x)$ . We will call such a  $T$  *empty* for obvious reasons. We have the following, worth noting at this time:

**Fact 1:** If  $\mathbf{Src}\langle\phi\rangle$  is empty and  $\mathbf{Tgt}\langle\phi\rangle$  is empty, then  $\mathbf{Fr1} \phi$ , no matter what  $\mathbf{Fnc}\langle\phi\rangle$  is.

We next define the predicate  $\mathbf{2}$ .  $\mathbf{2}$  is a predicate holding of things of  $\mathbf{Cat}$ -type, and  $\mathbf{2}T$  is meant to express that  $T$  is isomorphic to the ordinal category  $\mathbf{2}$ . This means that exactly three elements of the domain (call them  $a$ ,  $b$ , and  $c$ ) fall under  $\mathbf{Dom}\langle T \rangle$ , and with respect to these three elements the function  $\mathbf{Cmp}\langle T \rangle$  satisfies the conditions given in the table in Figure 3.1, where we use  $\perp$  as shorthand for the sentence expressing that the given object is not in  $\mathbf{Dom}\langle T \rangle$ . Together these ensure that the category  $T$  represents looks like the ordinal category  $\mathbf{2}$ , a picture of which can also be found in Figure 3.1

<sup>9</sup> For clarity, by “non- $\mathcal{C}$ -arrows” I mean non- $(\mathcal{C}$ -arrows); that is, things that fail to be  $\mathcal{C}$ -arrows, not merely arrows that are not among  $\mathcal{C}$ 's. Thus, for example, if  $\mathcal{C}$  is the category representing the ordinal 138, then among the objects qualifying as non- $\mathcal{C}$ -arrows are you and me and the set of all beluga whales, etc.

	$a$	$b$	$c$
$a$	$a$	$b$	$\perp$
$b$	$\perp$	$\perp$	$b$
$c$	$\perp$	$\perp$	$c$

**Figure 3.1:** Multiplication Table for and Picture of **2**

The last thing to define is the predicate  $\equiv$  that occurs written with infix notation at the end of BLF.  $\equiv$  is a four-place predicate that intuitively holds when (a) the composition of its first pair of inputs is functorial, (b) the composition of its second pair of inputs is functorial, and (c), the two compositions are in fact equal over the entire domain of the relevant categories. Formally, this is given by

$$\begin{aligned}
\equiv \langle \mathbf{Fnc}\langle\phi\rangle, \mathbf{Fnc}\langle\mu\rangle, \mathbf{Fnc}\langle\psi\rangle, \mathbf{Fnc}\langle\tau\rangle \rangle &=_{df} \mathbf{Fr1}\langle \mathbf{Fnc}\langle\phi\rangle \circ \mathbf{Fnc}\langle\mu\rangle, \mathbf{Src}\langle\mu\rangle, \mathbf{Tgt}\langle\phi\rangle \rangle \wedge \\
&\mathbf{Fr1}\langle \mathbf{Fnc}\langle\psi\rangle \circ \mathbf{Fnc}\langle\tau\rangle, \mathbf{Src}\langle\tau\rangle, \mathbf{Tgt}\langle\psi\rangle \rangle \wedge \\
&\forall x[\mathbf{Dom}\langle \mathbf{Src}\langle\mu\rangle \rangle x \rightarrow \\
&\quad \mathbf{Fnc}\langle\phi\rangle \circ \mathbf{Fnc}\langle\mu\rangle(x) = \mathbf{Fnc}\langle\psi\rangle \circ \mathbf{Fnc}\langle\tau\rangle(x)]
\end{aligned}$$

Using these definitions, one can methodically unfold the right hand side of BLF and reduce it to a sentence in  $\mathcal{L}$  containing only  $\phi$  and  $\psi$  free. For convenience, we restate BLF here without weird linebreaks:

$$\forall\phi\forall\psi \left[ \mathcal{F}(\phi) = \mathcal{F}(\psi) \Leftrightarrow \left( \neg \left[ \mathbf{Fr1}\phi \vee \mathbf{Fr1}\psi \right] \vee \left[ \mathbf{Fr1}\phi \wedge \mathbf{Fr1}\psi \wedge \mathbf{Src}\langle\phi\rangle = \mathbf{Src}\langle\psi\rangle \wedge \right. \right. \right. \\
\left. \left. \left. \mathbf{Tgt}\langle\phi\rangle = \mathbf{Tgt}\langle\psi\rangle \wedge \exists T(2T \wedge \forall\mu[(\mathbf{Fr1}\mu \wedge \mathbf{Src}\langle\mu\rangle = T \wedge \right. \right. \right. \\
\left. \left. \left. \mathbf{Tgt}\langle\mu\rangle = \mathbf{Src}\langle\phi\rangle) \Rightarrow \mathbf{Fnc}\langle\phi\rangle \mathbf{Fnc}\langle\mu\rangle \equiv \mathbf{Fnc}\langle\psi\rangle \mathbf{Fnc}\langle\mu\rangle] \right) \right] \right) \right]$$

In the next section, we examine consistency issues. It should be clear that BLF is both inconsistent and unsatisfiable, so we don't prove this – the interested reader can construct the proofs with little effort in any event.

### 3.5 Consistency and Size Issues

Recall that if  $P$  is a subconcept of  $Q$ ,  $|\{x : Px\}| = \kappa$  and  $|\{x : Qx\}| = \lambda$ , then  $P$  is called *exponentially small relative to  $Q$*  when  $\lambda^\kappa \leq \lambda$ . If the universe has cardinality  $c$ , then we say  $P$  is exponentially small relative to the universe when  $c^\kappa \leq c$ . A two-place “relative exponential smallness predicate” is expressible in third-order logic, see [30] for details. In the remainder, we let **Esm** be a defined predicate (expressible in our third-order language) that holds of exactly those concepts that are exponentially small relative to the universe. Further, we also define a 2-place predicate **Rsm**, that holds of a pair of concepts  $P$  and  $Q$  just when  $Q$  is exponentially small *relative to  $P$* .

Let  $\text{BLF}_E$  name the abstraction principle formed by restricting **BLF** to those quintuples  $\langle f, P, g, Q, h \rangle$  for which  $P$  and  $Q$  are both exponentially small relative to the universe. Being somewhat more explicit, the righthand side of  $\text{BLF}_E$  will begin

$$\begin{aligned} \dots \Leftrightarrow & \left( \neg \left[ \text{Fr1 } \phi \vee \text{Fr1 } \psi \right] \vee \neg \left[ \text{Esm} \langle \text{Dom} \langle \text{Src} \langle \phi \rangle \rangle \rangle \vee \text{Esm} \langle \text{Dom} \langle \text{Tgt} \langle \phi \rangle \rangle \rangle \right] \right) \\ & \vee \left[ \text{Fr1 } \phi \wedge \text{Fr1 } \psi \wedge \text{Esm} \langle \text{Dom} \langle \text{Src} \langle \phi \rangle \rangle \rangle \wedge \text{Esm} \langle \text{Dom} \langle \text{Tgt} \langle \phi \rangle \rangle \rangle \wedge \dots \right] \end{aligned}$$

$\text{BLF}_E$  is in fact consistent. This can be checked in a straightforwardly combinatorial way (see Appendix for details). Despite this,  $\text{BLF}_E$  is somewhat uninteresting on its own because too many objects lie in the range of its abstraction operator. That is, the concept holding of exactly the objects in the range of  $\text{BLF}_E$  is not exponentially small relative to the universe. Recall that our general plan had been to (a) produce an abstraction principle whose abstracts correspond to functors (which we’ve now done), then (b) to realize individual categories as the identity functors, and finally (c) discern the internal structure of the categories themselves by examining how they behave in the category of all such categories. This latter aim is frustrated by the fact that the collection of all functors – the natural object from which to build our category of categories – is too big for  $\text{BLF}_E$  to act on. So in some sense,  $\text{BLF}_E$  cannot provide us with an object corresponding to a category of all categories.

To rectify this we proceed as follows: suppose  $U$  is a concept that is exponentially small relative to the universe. We will add to  $\mathcal{L}$  a predicate **U** holding of all and only those objects falling under  $U$ . We then restrict **BLF** to those quintuples  $\langle f, P, g, Q, h \rangle$  for which  $P$  and  $Q$  are both exponentially small relative to  $U$ . If we call this new

abstraction principle  $\text{BLF}_U$ , its righthand side will begin

$$\begin{aligned} \dots \Leftrightarrow & \left( \neg \left[ \text{Fr1 } \phi \vee \text{Fr1 } \psi \right] \vee \neg \left[ \text{Rsm} \langle \text{U}, \text{Dom} \langle \text{Src} \langle \phi \rangle \rangle \rangle \vee \text{Rsm} \langle \text{U}, \text{Dom} \langle \text{Tgt} \langle \phi \rangle \rangle \rangle \right] \right) \\ & \vee \left[ \text{Fr1 } \phi \wedge \text{Fr1 } \psi \wedge \text{Rsm} \langle \text{U}, \text{Dom} \langle \text{Src} \langle \phi \rangle \rangle \rangle \wedge \text{Rsm} \langle \text{U}, \text{Dom} \langle \text{Tgt} \langle \phi \rangle \rangle \rangle \wedge \dots \right] \end{aligned}$$

The thing to take note of regarding  $\text{BLF}_U$  is the following: if we let  $\mathcal{F}_U$  be the abstraction operator in  $\text{BLF}_U$ , then the concept  $\mathbf{C}_U$  named by the predicate  $\exists \phi (x = \text{BLF}_U(\phi))$  is itself exponentially small (not relative to  $U$  but) relative to the universe.

Let the abstraction operator in  $\text{BLF}_E$  be called  $\mathcal{F}_E$ . Then the concept that holds of the objects in the range of  $\text{BLF}_U$  is small enough for  $\mathcal{F}_E$  to act non-trivially on. In particular, if  $\mathbf{C}_U = \mathcal{F}_E(\text{id}, \mathbf{C}_U, f, \mathbf{C}_U, f)$  where  $f$  is an arbitrary function representing the composition of the objects in the range of  $\mathcal{F}_U$  (more on this below), then  $\mathbf{C}_U$  is the identity functor (category) at (of) all the  $U$ -small categories and functors. In this way, functorial abstraction gives us, for each exponentially small concept  $U$ , a category of all  $U$ -small categories. That is, functorial abstraction is a relative theory of categorial abstraction that includes a relative theory of category-of-category abstraction.

In the next section, a number of technical constructions demonstrating the features of  $\mathbf{C}_U$  are given. The results are summarized at the start of section 7, so the reader uninterested in seeing these details can skip there.

### 3.6 Features of the Category of $U$ -small Categories

Let  $U$  be an exponentially small concept that contains a nonempty subconcept exponentially small relative to it. The theory we will be examining will take as axioms both  $\text{BLF}_U$  and  $\text{BLF}_E$ .

As just demonstrated,  $\text{BLF}_E$  admits an identity functor  $\mathbf{C}_U$  that we identify with the category of all  $U$ -small categories. To justify seeing  $\mathbf{C}_U$  as a category of categories, we need to determine that  $\mathbf{C}_U$  behaves as a category of categories ought. In particular, notice that just as it was with  $\text{BLM}$ ,  $\text{BLF}_U$  does not supply us with highly structured objects – the objects in the range of  $\mathcal{F}_U$  are not in an obvious way *required* by a theory that takes  $\text{BLF}_U$  as its only axiom to have any properties other than that they correspond to certain equivalence classes of functions. A category of categories on the other hand should have inside of it genuine *categories* with honest-to-god *functors* connecting them.

In the remainder of this section we demonstrate that the identity arrows in  $\mathbf{C}_U$  interact with the other arrows in  $\mathbf{C}_U$  in a way that allows us to *see* each identity functor *as* a category in its own right. To complete this task, I should also show the non-identity arrows “behave functorially” with respect to this derived internal categorial structure. However, this demonstration would essentially be a repetition of details that can be found in [32], so is not reproduced here.

A final introductory note to this section: to every abstract of  $\mathbf{BLF}_U$  there is a corresponding abstract of  $\mathbf{BLF}_E$ . There is thus a strict sense in which  $\mathbf{BLF}_U$  is superfluous. Nonetheless, it would be difficult to actually do any of the constructions that follow relying only on  $\mathbf{BLF}_E$ . Those with logicist leanings, however, should rest assured that the “ $\mathbf{BLF}_U$ -ladder” is (at least in principle) something we can kick away at the end.

### 3.6.1 Preliminaries

The idea is the following: In the case of  $\mathbf{BLM}$  we were able to use the structure imparted by  $\mathbf{BLM}$  to recover the “internal” structure of the sets we represented by identity mappings. Similarly, given a  $U$ -category  $C$  – that is, an identity arrow in the range of  $\mathbf{BLF}_U$  – we can use the functors in the category  $\mathbf{C}_U$  to discern the internal structure of  $C$ . Since most of the “action” is in the second disjunct of the right hand side of the principles  $\mathbf{BLF}_U$  and  $\mathbf{BLF}_E$  we restrict our attention to these portions. In order to not have to repeat myself too much, when I make claims that hold for both  $\mathcal{F}_U$  and  $\mathcal{F}_E$ , I will use  $\mathcal{F}$  and  $\mathbf{BLF}$ , unsubscripted, and leave it to the reader to supply the two versions of the statement that result from applying subscripts in the appropriate places.

The first thing to notice is that there is an obvious way to compose objects in the range of the  $\mathcal{F}$ -operator. Explicitly,

DEFINITION 1:

$$\mathcal{F}(\phi) \circ \mathcal{F}(\psi) =_{df} \mathcal{F}(\mathbf{Fnc}\langle\phi\rangle \circ \mathbf{Fnc}\langle\psi\rangle, \mathbf{Src}\langle\psi\rangle, \mathbf{Tgt}\langle\phi\rangle).$$

It is trivial to verify this is well-defined and associative.

Also, recall that objects of  $\mathbf{Fun}$ -type which fail to satisfy  $\mathbf{Fr1}$  are all mapped by the correspondence  $\mathcal{F}$  to the same object. We call this object “the bad object”, and we have

DEFINITION 2:

$\odot$  is the bad object.

There is no reason to suppose that the object  $\odot$  that plays the “bad object” role for  $\mathbf{BLF}_E$  is the same as the one that plays this role for  $\mathbf{BLF}_U$ . Nonetheless we use the name  $\odot$  to refer to either of them, and allow context to distinguish which one we are referring to.

Returning to composition, using  $\odot$ , we can now state

FACT 2:

Identities behave correctly: if  $x = \mathcal{F}(\mathbf{id}, P, f, P, f)$ , then for any  $y$ , if  $x \circ y \neq \odot$  then  $x \circ y = y$ , and if  $y \circ x \neq \odot$  then  $y \circ x = y$  (this is clear from the definitions).

Next, suppose  $2S$  and  $\mu = \langle f, S, T \rangle$ , where  $T$  is of  $\mathbf{Cat}$ -type. It is customary to call such a  $\mu$  a **2**-element of  $T$ . It is immediate from  $\mathbf{BLF}$  that parallel functors are either equal or disagree on being evaluated at (that is to say, composed with) some **2**-element. Speaking in the usual informal language of category theory:

FACT 3:

**2** is a generator.<sup>10</sup>

Now for some terminology: It will be important to distinguish between the abstracts in the range of  $\mathbf{BLF}_E$ , which we will call *functors* and for which we will use bold font characters (e.g. **A**, **B**, . . . , **1**, **2**, . . . , etc.), and the abstracts in the range of  $\mathbf{BLF}_U$ , which we will call *arrows* and for which we will use non-bolded upper-case characters.

Observe that if  $S$  is of  $\mathbf{Cat}$ -type and is empty, then there is no  $T$  and no  $\mu$  so that  $\mathbf{Fr1} \mu \wedge \mathbf{Src}\langle \mu \rangle = T \wedge \mathbf{Tgt}\langle \mu \rangle = S$ . It thus follows from  $\mathbf{BLF}$  directly that all quintuples  $\langle f, S, S \rangle$  where  $S$  is empty will be identified by either version of  $\mathbf{BLF}$ . But by **Fact 1**, this object is not  $\odot$ . We thus have

DEFINITION 3:

If  $S$  is of  $\mathbf{Cat}$ -type and empty, then  $\mathbf{0} =_{df} \mathcal{F}_E(\mathbf{id}, S, S)$  and  $\mathbf{0} =_{df} \mathcal{F}_U(\mathbf{id}, S, S)$ .

---

<sup>10</sup> Notice that this was essentially the starting point for the theory, so recovering it as a consequence is not unexpected.

It follows from Definition 3 that for each identity functor  $\mathbf{B} = \mathcal{F}_E(\text{id}, T, T)$ , there is exactly one functor  $!_{\mathbf{B}}$  that simultaneously satisfies both  $\mathbf{B} \circ !_{\mathbf{B}} \neq \odot$  and  $!_{\mathbf{B}} \circ \mathbf{0} \neq \odot$ . That is,

FACT 4:

$\mathbf{0}$  is initial among the functors and  $0$  is initial among the arrows.

Next, suppose  $f$  is a particular but arbitrary two-place total function for which there is exactly one element of the universe satisfying  $f(x, x) = x$ . Call this element  $\alpha$  and suppose  $\alpha$  falls under  $U$ . Also suppose for any  $x \neq \alpha$  the  $f$ -multiplication table for  $\alpha$  and  $x$  looks like Table 3.1 (repeating abbreviations from before).

	$\alpha$	$x$
$\alpha$	$\alpha$	$\perp$
$x$	$\perp$	$\perp$

**Table 3.1:** Multiplication Table for  $\mathbf{1}$

Let  $T$  be the pair consisting of the function  $f$  and the concept holding only of  $\alpha$ .

DEFINITION 4:

$\mathbf{1} = \mathcal{F}_E(\text{id}, T, T)$  and  $1 = \mathcal{F}_U(\text{id}, T, T)$ .

By an argument similar to the one above we can see

FACT 5:

$\mathbf{1}$  and  $1$  are terminal among the functors and among the arrows, respectively.

In a similar way, by writing out multiplication tables and encoding these in our logic, we can produce identity functors for the ordinals  $\mathbf{2}$ ,  $2$ , and  $\mathbf{3}$ , which we will need in the remainder.

Now, given  $x = \mathcal{F}_E(f, S, T)$ , notice that the functors  $\mathbf{A} = \mathcal{F}_E(\text{id}, S, T)$  and  $\mathbf{B} = \mathcal{F}_E(\text{id}, T, T)$  are well defined, and these are the unique functors that behave as left and right composition-identities for  $x$ . For this reason, we write  $x : \mathbf{A} \rightarrow \mathbf{B}$ . A similar set of definitions can obviously be stated for  $\mathcal{F}_U$ .

Using all of this, we now have



DEFINITION 5:

By “an *object of  $\mathbf{C}_U$* ,” we mean either (a) An identity arrow in the range of  $\mathbf{BLF}_U$  or (b) A functor  $f : \mathbf{1} \longrightarrow \mathbf{C}_U$  in the range of  $\mathbf{BLF}_E$ . These are in natural bijection, so we do not distinguish between them. We call the objects of  $\mathbf{C}_U$  *U-categories* to distinguish them from the objects (identity functors) in the range of  $\mathbf{BLF}_E$ , which we will call *E-categories*.

The bijection between identity arrows and functors  $\mathbf{1} \longrightarrow \mathbf{C}_U$  can be extended to a bijection between arrows in general and functors  $\mathbf{2} \longrightarrow \mathbf{C}_U$ . So we have

DEFINITION 6:

By “an *arrow of  $\mathbf{C}_U$* ” we will mean either (a) An abstract in the range of  $\mathbf{BLF}_U$  or (b) A functor  $f : \mathbf{2} \Rightarrow \mathbf{C}_U$  in the range of  $\mathbf{BLF}_E$ .

It is a consequence of the definitions of  $\mathbf{1}$  and  $\mathbf{2}$  that there are exactly two functors  $\mathbf{1} \longrightarrow \mathbf{2}$ . If  $f$  is an arrow of  $\mathbf{C}_U$ , then, we can compose  $f$  with these two functors to get a unique pair of objects  $s_f$  and  $t_f$  which we distinguish (when they are different) by demanding that  $f \circ s_f$  be defined and that  $t_f \circ f$  be defined. We call  $s_f$  the *source* of  $f$  and  $t_f$  the *target* of  $f$  and say that  $f$  goes *from  $s_f$  to  $t_f$* .<sup>11</sup>

DEFINITIONS 6-9:

Given  $C$  is an object of  $\mathbf{C}_U$  (that is, a *U-category*), we define a *C-object* to be an arrow from  $\mathbf{1}$  to  $C$  and a *C-morphism* to be an arrow from  $\mathbf{2}$  to  $C$ . A *constant arrow* is an arrow whose range is a single object; that is, an arrow that factors through an object. A *discrete U-category* is a *U-category*  $C$  for which every arrow  $\mathbf{2} \longrightarrow C$  is constant. Repeating much of the previous paragraph, we can define *source* and *target*, *to* and *from* for morphisms of a given *U-category*.

### 3.6.2 Theorems

The following theorems are easily proved:

---

<sup>11</sup> This method of distinguishing source from target depends on the conflation of the two definitions of objects. This is not the only way to accomplish this, but is the most intuitive, to my reckoning.

- Given morphisms  $a$  and  $b$  in the  $U$ -category  $C$ , there is a discrete  $E$ -category  $\{a, b\}$  whose objects are exactly  $a$  and  $b$ . There is also a discrete  $U$ -category  $E$ -isomorphic to  $\{a, b\}$ . For convenience, we identify these two isomorphic categories, so  $\{a, b\}$  will be considered a  $U$ -category.
- There are two non-constant epic arrows  $2 \longrightarrow \{a, b\}$ .<sup>12</sup> We identify these with the pairs  $(a, b)$  and  $(b, a)$ .

Using pairs we can easily define products of  $U$ -categories, in the usual way.

Similarly, we can easily build disjoint union categories: if  $S$  and  $T$  are of type  $\mathbf{Cat}$ , we can assume without loss of generality that  $\mathbf{Dom}\langle S \rangle$  and  $\mathbf{Dom}\langle T \rangle$  are disjoint (since they are small relative to  $U$ ). The disjoint union of  $S$  and  $T$  is then  $\mathcal{F}_U(\mathbf{id}, S \coprod T, S \coprod T)$  where  $S \coprod T$  is the pair consisting of the concept that holds of the disjoint union of  $\mathbf{Dom}\langle S \rangle$  and  $\mathbf{Dom}\langle T \rangle$ , and the two-place function that satisfies  $\mathbf{Cmp}\langle S \rangle$  when both its arguments lie inside  $S$ , and which satisfies  $\mathbf{Cmp}\langle T \rangle$  everywhere else. This paragraph and the previous allow us to conclude that the category of  $U$ -categories has finite products and finite coproducts.

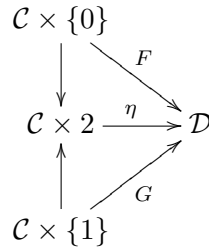
One can easily show the equalizer of any pair of parallel arrows exists: Suppose  $\Phi$  and  $\Psi$  are of  $\mathbf{Fun}$  type, and suppose  $\mathbf{Src}\langle \Phi \rangle = \mathbf{Src}\langle \Psi \rangle$ , and  $\mathbf{Tgt}\langle \Phi \rangle = \mathbf{Tgt}\langle \Psi \rangle$ . Let  $F = \mathcal{F}_U(\Phi) \neq \odot$  and  $G = \mathcal{F}_U(\Psi) \neq \odot$ . Finally, let  $A = \mathcal{F}_U(\mathbf{id}, \mathbf{Src}\langle \Phi \rangle, \mathbf{Src}\langle \Phi \rangle)$  and  $B = \mathcal{F}_U(\mathbf{id}, \mathbf{Tgt}\langle \Psi \rangle, \mathbf{Tgt}\langle \Psi \rangle)$ . It is standard to write this  $A \xrightarrow[F]{G} B$ . Then the equalizer of  $F$  and  $G$  is  $\mathcal{F}_U(\mathbf{id}, F_E G, F_E G)$ , where  $F_E G$  is the pair consisting of the concept holding of those elements of  $\mathbf{Src}\langle \Phi \rangle$  where  $\mathbf{Fnc}\langle \Phi \rangle = \mathbf{Fnc}\langle \Psi \rangle$  and the function  $\mathbf{Cmp}\langle \mathbf{Tgt}\langle \Phi \rangle \rangle$ .

Coequalizers take work but can also be shown to exist. Rather than provide all the necessary details here, I refer the reader to the excellent presentation in [38]. It is clear that a similar class of constructions to these can be carried out.

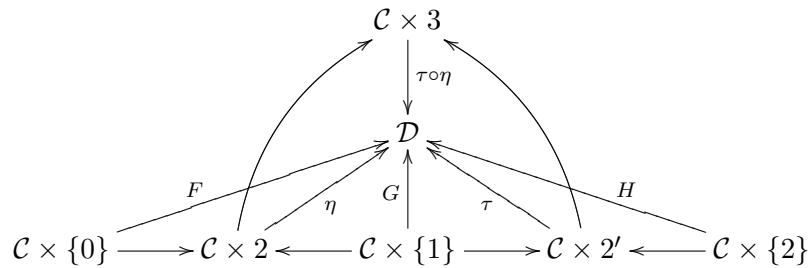
As in Lawvere's thesis, given arrows  $F, G : C \longrightarrow D$ , we identify natural transformations  $\eta : F \longrightarrow G$  with arrows  $C \times 2 \longrightarrow D$ . We can thus define exponential objects: given two  $U$ -categories  $C$  and  $D$ , let  $D^C = \mathcal{F}^*(\mathbf{id}, d^c, d^c)$  where  $d^c$  is a pair consisting of the concept  $P_{d^c}$  holding of those arrows  $\eta$  for which there are arrows  $F, G : C \longrightarrow D$

<sup>12</sup> Recall that an arrow  $a$  is *epic* if it is right-cancelable, that is, if whenever  $b \circ a = c \circ a$  we can conclude that  $b = c$ .

that make the following diagram commute, and we explain the \* notation below:



and a function that maps a pair of such arrows to the arrow  $\tau \circ \eta$  that makes the following diagram commute when such exists, and maps to objects not falling under  $P_{dc}$  otherwise:<sup>13</sup>



A lengthy but routine verification then establishes that this object is indeed an exponential object for  $\mathcal{C}$  and  $\mathcal{D}$ . However, at this point we may leave the  $E$ -category  $\mathbf{C}_U$  – exponential objects need not be  $\mathbf{U}$ -categories, as they may simply be too big. However, they will certainly be  $E$ -categories. Thus,  $\mathcal{F}^*$  is meant to pick out whichever of the  $\mathcal{F}$ 's is needed here.

### 3.7 Philosophical Conclusion

Altogether, the category of  $U$ -small categories can now be seen to be finitely complete and cocomplete, and exponential objects – despite potentially being too large – are available, though not all of them are properly part of the category of  $U$ -small categories. These features demonstrate abstractionist categories of categories have some intrinsic interest. This construction has raised more questions than it has answered, though. The three that most interest me are the following:

---

<sup>13</sup> In this diagram,  $2'$  is actually just  $2$ , but where we now identify it as  $1 \longrightarrow 2$  instead of  $0 \longrightarrow 1$ . This makes sense of the two arrows  $\mathcal{C} \times 2 \longrightarrow \mathcal{C} \times 3$ .

- (1) Can we prove the existence of nontrivial topoi?
- (2) How do we produce enriched categories or (more specifically), 2-categories?
- (3) The identity conditions we've imposed on functors are a bit severe. Can they be relaxed?

Expanding on (3) a bit, the question to examine is whether we can recover essentially the same theory using a “nicer” version of  $\mathbf{Fr1}$ .

A great deal of development remains to be done before we can answer these. A broader project worth pursuing is to philosophically and technically examine a general theory of relative abstraction along the lines hinted at in this paper.

As promised, however, at this point we put aside technicalities and examine the philosophical significance of the above derivation of the internal structure of particular  $U$ -categories in terms of the “external” structure of the  $E$ -category of all  $U$ -categories. As pointed out before, the fact that categories have a rich internal structure has been grounds for rejecting categories of categories for foundational purposes in the past. The above demonstration that all the internal structure of a (small) category can be recovered as a matter of logic from only features of the (large) category of all (small) categories should put pressure on this view, since it suggests a blurring of the internal structure/external structure distinction on which the objection relies.<sup>14</sup>

Nonetheless, a persistent objector could continue arguing as follows: The above construction only satisfied me on the condition that it provided an account of the internal structure of particular  $U$ -categories in terms of the structure of the  $E$ -category of all  $U$ -categories. Thus, in order to understand the foundational significance of the construction provided (and thus the foundational significance of category theory in general), a prior knowledge of this internal structure is necessary. So categories of categories are not *properly* autonomous, hence cannot serve as an autonomous foundation for mathematics.

I have two responses to this.

---

<sup>14</sup> I am by no means claiming to be original here. Both this result and this interpretation of it are common in the category-theory literature and “folklore”. I should further point out that the *heavy* reliance of this approach on higher-order logic and full second-order comprehension mean that this result in no way puts pressure on issues regarding the autonomy of a *first-order* theory of a category of categories.

- In the first place, the fact that I was *looking for* an external characterization of the internal structure of a category does not demonstrate the characterization I found *depends on* knowledge of both structures in order to be coherent.
- In the second place, all one needs is *a* category-theory independent way of understanding the notion of a category (such as that given in [39]) in order to recognize that a particular category one has been presented with can be seen in such a way that its identity morphisms intuitively behave like categories themselves. This is exactly what the above construction allows.

Thus, at least *this* objection to a category of categories as an autonomous mathematical object, and hence as acceptable foundational purposes seems inadmissible. This in no way obviates, e.g. the objections to the autonomy of a first-order category theory found in [11]. In particular, Hellman points out that in a category of categories “somehow we need to make sense of talk of structures *satisfying* the axioms of category theory, i.e. being categories.” This seems correct. In fact, far from obviating such points, the approach on hand is an explicit *instance* of them: the first and second conjuncts of the predicate  $\text{Fr1}$  demand that  $\text{Tgt}\langle\phi\rangle$  and  $\text{Src}\langle\phi\rangle$  satisfy  $\text{ax}$  – that is, that they be *categories*. That we can make this demand at all depends, in turn, on our use of higher-order logic in the statement of this abstraction principle. Thus, we should perhaps add the following caveat: a category of categories is as acceptable for foundational purposes *as anything else that relies on higher-order logics*. For the abstractionist, presumably, this isn’t giving up much.

Aside from what this construction says about foundational matters – interesting though this may be – there is further philosophical merit to the construction provided. As remarked above, one need not be a neologist to find the production of abstraction principles valuable. When the identity conditions laid down by an abstraction principle suffice to determine all the features some given type of object is supposed to have, one has reason to believe that those identity conditions specify something like the essential features of the objects in question. We’ve now seen that  $\text{BLF}_U$  does exactly this for the category of  $U$ -small categories. Putting aside all neologist claims about abstraction principles, this result retains significant *epistemic* significance: it gives us reason to believe that the identity conditions  $\text{BLF}_U$  captures tell us something essential about

categories.

The remarkable thing about this is the essential *simplicity* of what  $\mathbf{BLF}_U$  imposes. Underneath the technicalities, what it “says” is the defining feature of a category of categories is that  $\mathbf{2}$  is a generator. That using “ $\mathbf{2}$  is a generator” as an abstraction principle is, in turn, capable of producing (together with higher-order logic) *so much* category theory suggests that  $\mathbf{2}$  being a generator is not only essential to what it means to be a category of categories, but also to category *theory writ large*.<sup>15</sup>

Such justification for the theory produced here is clearly needed. After all, a second-order Hilbert-style axiomatization of categories of categories would seem to be a much easier approach to providing defining conditions on categories of categories than the route taken here.<sup>16</sup> When one examines such an axiomatization – for example, one could take a version of either the theory given in [2] or the one in [32] – the central role being played by “ $\mathbf{2}$  is a generator” is entirely obscured. One can see the same thing at play when comparing the theory of arithmetic provided by  $\mathbf{HP}$  with, say, second-order Peano Arithmetic. The latter obscures any relevant role being played by “concepts in bijection get the same number”. It is a substantial *epistemic* insight into the nature of arithmetic as a mathematical and scientific practice to establish – as Frege and Boolos’

---

<sup>15</sup> These facts are suggestive of an analogy worth looking into more – any one-element set, after all, is a generator for a category of sets in precisely the same way that  $\mathbf{2}$  is a generator for a category of categories. Perhaps a relative set abstraction theory based on “1 is a generator” would have some intrinsic interest. More speculatively, it might be worth seeing what the the relative abstraction theory based on “ $\lambda$  is a generator” might lead us to for arbitrary ordinals  $\lambda$ .

<sup>16</sup> It is worth pausing here to point out that there is an important sense in which the Hilbertian-structuralist approach to the foundations of mathematics and the neologist approach can be seen as answering *different questions*. These differences are more easily seen in the case of arithmetic. The Hilbertian-structuralist would propose – *a la* Dedekind’s famous treatment in [40] – their axioms as solutions to the question “How must things *be* in order that they behave as numbers?” or, using the standard jargon, “How can we define what it is to be a natural numbers structure?”. The neologist reconstruction of arithmetic, on the other hand answers the question “What are the essential features of numbers?” and answers it by providing identity conditions in terms of these essential features. As it turns out, isolating these essential feature still leaves open the actual identity of the objects in question – this is the problem known in the literature as the Caesar Problem (see, e.g. [41], [19], [42], [43], or [44]).

The fact that the questions being asked by the devotees of the two approaches are different precedes and in some sense explains the epistemic difference being discussed here. We should be unsurprised, that is, to *not* find in an axiomatization of a category of categories like Lawvere’s, anything that tells us which features of categories of categories describe their “essential role” in our mathematical and scientific practices. Lawvere’s axioms were simply not aimed at providing this sort of information, but rather at describing how things must *be* in order that they behave as categories – no matter what those things might actually be.

theorems do – that in fact “concepts in bijection get the same number” provides us with the whole of arithmetic. Similarly, it is a substantial epistemic insight into the nature of category theory as a scientific and mathematical practice to observe that “**2** is a generator” gets us at least a substantial and important fragment of that discipline.

## Appendix: Proof that $\text{BLF}_E$ Can be Modeled

The following is a sketch of a proof that  $\text{BLF}_E$  can be modeled. We begin by counting, for a set  $S$  of cardinality  $\mathbf{c}$  and an arbitrary cardinal  $\aleph_0 \leq \kappa \leq \mathbf{c}$ , the cardinality of

$$\{(A, B, f, g, h) : A \subseteq S, B \subseteq S, |A| \leq \kappa, |B| \leq \kappa, f : A \times A \Rightarrow S, g : B \times B \Rightarrow S, h : A \Rightarrow B\}$$

We proceed in stages.

Stage 1: Observe there are  $\mathbf{c}^\kappa$  choices for  $A$  and for  $B$ . Thus, there are  $(\mathbf{c}^\kappa)^2$  pairs  $(A, B)$ .

Stage 2: Next, notice that for a given  $A$ , there are no more than  $\mathbf{c}^{\kappa^2}$  options for  $f$ . Thus, there are no more than  $\mathbf{c}^{\kappa^2} \cdot (\mathbf{c}^\kappa)^2$  triples  $(A, B, f)$ .

Stage 3: Similarly, for a given  $B$ , there are no more than  $\mathbf{c}^{\kappa^2}$  options for  $g$ . Thus, there are no more than  $(\mathbf{c}^{\kappa^2})^2 \cdot (\mathbf{c}^\kappa)^2$  quadruples  $(A, B, f, g)$ .

Stage 4: Finally, for a given  $A$  and  $B$ , there are no more than  $\kappa^\kappa$  options for  $h$ . Thus, there are no more than  $\kappa^\kappa \cdot (\mathbf{c}^{\kappa^2})^2 \cdot (\mathbf{c}^\kappa)^2$  quintuples  $(A, B, f, g, h)$ .

In any model of  $\text{BLF}_E$  in a domain of cardinality  $\mathbf{c}$ , the abstracts it produces correspond to equivalence classes formed from a subset of the quintuples with this form, so the cardinal number just computed is an upper bound on the cardinality of the collection of abstracts in the range of  $\text{BLF}_E$ .

Finally, supposing  $\aleph_0 \leq \kappa \leq \mathbf{c}$  and  $\mathbf{c}^\kappa \leq \mathbf{c}$ , it is clear that this upper bound is no more than  $\mathbf{c}$ . Thus,  $\text{BLF}_E$  produces no more than  $\mathbf{c}$  abstracts, so it can be modeled.

## Chapter 4

# The Semantics of Social Constructivism

In a 2010 paper, Jill Dieterle criticized the view in Cole’s 2009 paper for being unable to account for the atemporality of mathematical existents. Cole’s 2013 paper addresses this objection, providing a modification of his 2009 paper allowing for atemporal mathematicalialia. An unusual consequence of Cole’s account is that at least some existential claims about mathematicalialia used to be false but now have always been true.

By examining the semantics of such claims, we demonstrate that social constructivism is in fact, despite Cole’s attempts to rectify matters, incompatible with atemporal mathematicalialia. In the course of examining these semantic details, however, an alternative hybrid view of fictionalism and social constructivism emerges. Those tempted by social constructivism, while perhaps disappointed by the negative results of the paper, may be encouraged by how much of their view can be recovered in this alternative account.

### 4.1 The Cole-Dieterle Debate

Consider the following well-known puzzle:

“The freedom and authority that mathematicians feel they enjoy to creatively postulate mathematical entities do not accord well with realist or



Platonist interpretations of mathematical theories. Yet, the intellectual ease with which mathematicians ontologically commit themselves to mathematical entities does not accord well with fictionalist or modal nominalist interpretations of mathematical theories.” [5, p. 593]

It is worth explaining a few details of Cole’s presentation of this problem, as it underlies much of the appeal social constructivist claims have in general.

Cole understands mathematical platonism to directly challenge the authority mathematical practice suggests the mathematician has to postulate new mathematical entities. Roughly, as Cole puts it, if mathematical entities exist but are confined to other ‘realms,’ then “mathematicians’ authority – particularly to postulate new mathematical entities – is legitimate only if [the mathematicians] are epistemologically in tune with these realms.” That is, in the world as it appears to the platonist, mathematicians do not have the authority to introduce new entities and endorse new existential claims at will. Instead, a mathematician is authorized to make only those claims that agree with the facts about which mathematical entities actually populate the ‘other realms.’ So, unless she is in sufficient contact with these realms to have a (at least partial) roster of their constituents, the mathematician is without authority to “endorse existential pure mathematical statements” with the levels of freedom and authority mathematical phenomenology would lead her to expect.

Fictionalism and modal nominalism face a different type of problem. For these sorts of theories, it is no problem to postulate new mathematical entities, but this freedom comes at the cost of changing how such claims are to be interpreted. Of particular importance for this essay is that fictionalism is characterized by understanding mathematical existence claims to be literally false. Such an interpretation, however, does not appear consistent with the way these claims are made by practicing mathematicians.

Social constructivism can be seen as an alternative to these traditional accounts. Mathematicalia, on this account, are “pure constitutive social constructs” that are “constituted by mathematical activities.” Cole defines these phrases as follows:

- “[A]n item is a constitutive social construct if and only if it exists in virtue of a group of individuals having granted [it] a normative role in certain of their activities.” [5, p. 597]

- *Pure* constitutive social constructs are those social constructs which are not pre-existing objects now being assigned a new function, but those whose very presence relies on social construction.

Cole thus supposes, in particular, that mathematical objects are objects whose existence both depends on and supervenes on mathematicians' doing mathematics.

[6] presents criticisms specifically aimed at Cole's 2009 account, but which apply more generally to any attempt to provide a social-constructive metaphysical account of mathematics. Of the criticisms presented there, the one that will occupy us she calls "The Contingency Thesis:"

"if one claims that  $X$  is socially constructed, then one must believe that  $X$  is not inevitable. If  $X$  is the product of social forces, then  $X$  exists, or exists in the way that it does, only contingently. At least on first inspection, this seems exactly wrong in the case of mathematics." [6, p. 323]

In particular, social constructs seem highly *temporally* contingent. For example, it seems social constructivists (and Cole in particular) must deny mathematical objects existed prior to the presence of a group of individuals granting mathematical objects a role in certain of their activities. But if this is true, it seems quantifying over mathematical objects in descriptions of the very distant past may be illegitimate since the entities being quantified over literally did not exist at such times. This puts much of our very best science (e.g. cosmology, evolutionary biology, etc.) at risk of incoherence in the following way:

Suppose we ask "why is the Sun not a blue dwarf?" Presumably, our best answer would be something like "when the Sun was forming, the cloud it accreted from had a density between  $x$  and  $y$  and a mass of available material between  $w$  and  $z$ . This caused the development to proceed in such and such a way." This answer appears to contain terms meant to refer to numbers with respect to a time when, if numbers are social constructs, there would be no referent for such terms. At least *prima facie*, such claims seem at risk of being rendered incoherent by any account of the metaphysics of mathematical objects (like Cole's) which holds

- (i) Mathematical claims must be literally true in order to have their intended meaning, and

(ii) Truths about mathematical existence claims supervene on mathematical practices.

In his 2013 paper, Cole responds to Dieterle’s criticism by claiming social construction has powers not only of ontological creativity, but also of modal creativity. That is, Cole claims that not only can social action lead to the production of new *things*, it can also lead to new *necessities*. This modal creativity is introduced via what Cole names temporal and modal profiles:

“Label the collection of times at which some facet of reality exists or its atemporality that facet’s *temporal profile* and the collection of possible worlds in which it exists or its amodality its *modal profile*.” [7, p. 20]

Cole’s claim is that we can “*Declare and collectively recognize* that institutional facets of reality have whatever modal profile best serves our purposes.” (ibid.) So, when social activity results in the existence of a new item (e.g. the number two), the responsible community might, in the creative process “*Declare and collectively recognize*” that it not only now exists, but also existed at other times and in other worlds (e.g. for two, it is declared to have existed in all worlds at all times).

Now, Cole takes the trouble in his 2013 paper to provide us with a specific construction (socially constructed temporal and modal profiles) that allows his account to avoid Dieterle’s criticism. However, any account of the metaphysics of mathematical holding (i) and (ii) above will be subject to Dieterle’s criticism and will have to, in some way, explain how existential claims concerning mathematical can be atemporal truths despite mathematical themselves depending on mathematical practices. So, while the criticism that follows poses a problem for any social-constructivist solution to the problems posed by Dieterle, the emphasis will be on the details of Cole’s particular construction.

## 4.2 Regimenting Cole’s Discussion

It is useful to begin by analyzing the linguistic structure of Cole’s proposal. When doing mathematics, it seems we use a language in which all claims of the form “at time  $t$ , there was a prime less than three” are truths. We will call this language the *object-language* (abbreviated  $\mathcal{OL}$ ) to distinguish it from the *meta-language*  $\mathcal{ML}$  in which Cole theorizes

about claims within  $\mathcal{OL}$  and, in particular, about this type of temporal claim. My worry centers on the semantics of  $\mathcal{ML}$ , and, in particular, about the content of certain temporal claims within  $\mathcal{ML}$  that are central to the success of Cole’s account.

First, an observation: a large amount of mathematics will have to be expressible in  $\mathcal{ML}$ . For example, Cole’s discussion of temporal and modal profiles is presumably happening in  $\mathcal{ML}$ ; given minimal assumptions about what our temporal discourse commits us to,  $\mathcal{ML}$  will thus have to have at least enough expressive power to formulate claims about sets of rational numbers. From here, of course, an enormous volume of mathematics becomes encodable within  $\mathcal{ML}$ .

If Cole is to be taken as doing more than painting a veneer of social-constructivism on top of a non-social-constructive core theory, though, he must be read as claiming the referents of all mathematical language – whether occurring in  $\mathcal{OL}$  or in  $\mathcal{ML}$  – are social constructs. That is, the referents of  $\mathcal{ML}$ -mathematical language must be “pure constitutive social constructs” just as he claims the referents of  $\mathcal{OL}$ -mathematical language are, else his solution to the problem of the temporality of his mathematicalialia implicitly assumes atemporal mathematicalialia already exist. In particular, the claim (made in  $\mathcal{ML}$ ) that we can “*Declare and collectively recognize* that institutional facets of reality have whatever modal profile best serves our purposes” must apply to the referents of  $\mathcal{ML}$ -mathematical language.

The next observation to make is that a declaration or collective recognition of the temporal profile of a given institutional facet of reality must happen *at some particular time*. Prior to that time, then, that facet of reality was without the specified temporal profile. Further, we do as a matter of fact know there were times prior to the declaration or collective recognition of any given temporal profile. Thus  $\mathcal{ML}$  must contain true sentences of the form “this facet of reality has such and such temporal profile, but it has not always had that temporal profile.” Choosing an instance, if mathematicalialia are social constructs, then it is currently the case the following  $\mathcal{ML}$ -sentence expresses a truth:

“It has always been the case that there is a prime less than three, but it has not always been the case that it has always  
 (3)  
 been the case that there is a prime less than three.”

### 4.3 The Analysis of Temporal Sentences

It is reasonable to wonder whether sentence (1) is consistent and (if it is) whether it can be made sense of. In this section I examine two languages –  $\mathcal{L}^P$  (the language of Priorean Temporal Logic) and an expressive expansion  $\mathcal{L}_{SU}^P$  of  $\mathcal{L}^P$  – that share a common setting for their semantics. The goal of the examination is to try to make sense of sentences like (1) in a way consistent with the role they play in Cole’s account. We will see that neither language suffices for this. In  $\mathcal{L}^P$ , sentences with the same logical form as (1) are always false, while in  $\mathcal{L}_{SU}^P$ , such sentences can be interpreted in a satisfiable way only if we do not take them literally. Given these failures, we then turn to the more complex language of Priorean Doubly-Temporal Logic,  $\mathcal{L}^{2P}$ .

By the end of section 5, we will see  $\mathcal{L}^{2P}$  is also insufficient to allow for mathematicalities that are both atemporal and socially constructed. The appendix sketches how to generalize this result for the languages  $\mathcal{L}^{\alpha P}$  for arbitrary ordinal  $\alpha$ , seeming to provide a substantial roadblock to social constructivism (or Julian Cole’s version of it, at least). Despite these negative results,  $\mathcal{L}^{2P}$  does allow for a version of social constructivism that, while not allowing for *genuinely* atemporal mathematicalities, nonetheless can make sense of the fact that mathematicalities *appear* atemporal.

#### 4.3.1 Syntax

Despite the ontological thrust of Cole’s account, examining existential claims as propositions is sufficient to expose its problems. The simplification achieved by using propositional rather than predicate temporal logic makes the few difficulties this raises worth suffering.

With that in mind, let  $\mathcal{L}$  be a standard language for propositional logic. The alphabet of  $\mathcal{L}^P$  differs from the alphabet of  $\mathcal{L}$  only by the addition of two one-place sentential operators  $G$  and  $H$ . When translating  $\mathcal{L}^P$ -formulas into natural language (NL),  $G\phi$  is read “ $\phi$  is always **going** to be true;”  $H\phi$  as “ $\phi$  always **has** been true.” Grammatically,  $\mathcal{L}^P$  is the closure of  $\mathcal{L}$  after the addition of these operators.

In the remainder of this essay, we frequently refer back to sentence (1). To make this easier, in every language introduced, we let  $Q$  be the proposition corresponding to the sentence “there is a prime less than 3.” Thus, in  $\mathcal{L}^P$ , (1) is most naturally translated

$\mathcal{M}, t \Vdash P$	iff	$v_t(P) = T$
$\mathcal{M}, t \Vdash \neg\psi$	iff	$\mathcal{M}, t \not\Vdash \psi$
$\mathcal{M}, t \Vdash \phi \wedge \psi$	iff	$\mathcal{M}, t \Vdash \phi$ and $\mathcal{M}, t \Vdash \psi$
$\mathcal{M}, t \Vdash G\phi$	iff	$\mathcal{M}, s \Vdash \phi$ for all $t < s$
$\mathcal{M}, t \Vdash H\phi$	iff	$\mathcal{M}, s \Vdash \phi$ for all $s < t$

**Figure 4.1:**  $\mathcal{L}^P$  Semantics

as  $HQ \wedge \neg HHQ$ .

Notice if we read the conditional via the definition  $a \supset b := \neg(a \wedge \neg b)$ , then our translation of (1) is logically equivalent to  $\neg(HQ \supset HHQ)$ . This conditional formulation of (1) is worth spending a moment examining. Recall that S4 modal logic is characterized axiomatically by the inclusion of “ $\Box\phi \supset \Box\Box\phi$ ” among the axioms and semantically by the demand that the accessibility relation be transitive. Regarding temporal logic as a branch of modal logic, we see that we are going to face a rather difficult task:  $HQ \supset HHQ$  is a sentence we anticipate will be entailed by the transitivity of the ordering of time and, absent *extremely* compelling evidence, the assumption that times are ordered transitively is something we are unlikely to give up. The remainder of the section will make more explicit the full extent of this difficulty.

### 4.3.2 Semantics

An  $\mathcal{L}^P$ -model is a triple  $\langle \mathcal{T}, <, v \rangle$ , where

- $\langle \mathcal{T}, < \rangle$  is a poset (partially ordered set – i.e.  $<$  is an irreflexive and transitive relation on  $\mathcal{T}$ ) representing the “times” or “moments” together with their ordering and
- $v$  is a function from  $\mathcal{T}$  to the set of functions from **Prop** (the set of propositions) to the two-element set  $\{T, F\}$ .

For  $t \in \mathcal{T}$   $v(t)$  is written  $v_t$ . Thus at  $t \in \mathcal{T}$ , the truth value of proposition  $P$  under valuation  $v$  is written  $v_t(P)$ . In symbols we have

$$\begin{aligned}
v : \mathcal{T} &\longrightarrow \{f : \mathbf{Prop} \longrightarrow \{T, F\}\} \\
t &\longmapsto (\mathbf{Prop} \ni P \longmapsto v_t(P) \in \{T, F\})
\end{aligned}$$

Given an  $\mathcal{L}^P$ -model  $\mathcal{M} = \langle \mathcal{T}, <, v \rangle$  and  $t \in \mathcal{T}$ , we define truth in  $\mathcal{M}$  at  $t$  in the expected way (see Figure 4.1). Truth in a model generally, validity in a given poset, validity in a class of posets, and validity in general are also defined as expected.

### 4.3.3 An Example

Let  $\phi$  be any  $\mathcal{L}^P$ -sentence whatsoever. Consider the  $\mathcal{L}^P$ -sentence  $H\phi \supset HH\phi := \neg(H\phi \wedge \neg HH\phi)$ .

We can determine the class of posets in which this sentence is valid as follows: Let  $\mathcal{M} = \langle \mathcal{T}, <, v \rangle$  be an  $\mathcal{L}^P$ -model.  $\neg(H\phi \wedge \neg HH\phi)$  is valid in  $\mathcal{M}$  if and only if for every  $t \in \mathcal{T}$  we have that  $\mathcal{M}, t \Vdash \neg(H\phi \wedge \neg HH\phi)$ . Using the semantic rules given in Figure 4.1, we see

$$\begin{aligned}
\mathcal{M}, t \Vdash \neg(H\phi \wedge \neg HH\phi) &\text{ iff } \mathcal{M}, t \nVdash H\phi \wedge \neg HH\phi \\
&\text{ iff } \mathcal{M}, t \nVdash H\phi \text{ or } \mathcal{M}, t \nVdash \neg HH\phi \\
&\text{ iff } \mathcal{M}, t \Vdash HH\phi \text{ or for some } s < t \mathcal{M}, s \nVdash \phi
\end{aligned}$$

Next, notice  $\mathcal{M}, t \Vdash HH\phi$  iff for all  $s < t$ ,  $\mathcal{M}, s \Vdash H\phi$ , which in turn is true iff for all  $u < s$ ,  $\mathcal{M}, u \Vdash \phi$ . In sum,  $\mathcal{M}, t \Vdash \neg(H\phi \wedge \neg HH\phi)$  iff either

- (a) For some  $s < t$ ,  $\mathcal{M}, s \nVdash \phi$ , or
- (b) For all  $u$  and all  $s$  with  $u < s < t$ ,  $\mathcal{M}, u \Vdash \phi$ .

But if (a) is false then for every time  $s < t$ ,  $\mathcal{M}, s \Vdash \phi$ . In particular, this makes (b) true. We conclude that for every  $\mathcal{L}^P$ -sentence  $\phi$ , the sentence  $\neg(H\phi \wedge \neg HH\phi)$  is valid in every  $\mathcal{L}^P$ -model.

#### 4.3.4 $\mathcal{L}^P$ as a Solution

$\mathcal{L}^P$  together with the semantic theory in Figure 4.1 is not an appropriate setting for interpreting NL-sentences like (1) – at least, not if we think (as Cole seems to) that such sentences at least *might* be true. To see this, observe that by the example just examined,  $\neg(HQ \wedge \neg HHQ)$  is true in every  $\mathcal{L}^P$ -model. On the other hand, (1) is most naturally translated as  $HQ \wedge \neg HHQ$ . So (1), when translated into  $\mathcal{L}^P$  in the most natural way, is unsatisfiable.

One could argue this is exactly the expected result. After all, (1) seemed to border on incoherence; the analysis demonstrating unsatisfiability only means the incoherence is more-than-bordered-on. There seems to be some intuitive content in (1), however, and a charitable interpretation of Cole should try a bit harder to squeeze this content out before simply abandoning his project. In particular, it is worth spending a moment seeing if we can suss some meaning out of (1) without abandoning  $\mathcal{L}^P$  models as the setting for our semantic theory.

#### 4.3.5 “Since...” and “Until...”

To solve the problem we are facing, it seems we will need to change one or more of the following:

- (a) How we interpret NL-sentences like (1),
- (b) The target language of our interpretation, or
- (c) The semantic theory attached to the target language of the interpretation.

The plan in this section is to focus our attention on (a). Loosely, the NL-sentence “there has always been a prime less than 3” will be interpreted not as “ $HQ$ ,” but as (a formal version of) “since (such and such),  $Q$ .” To accommodate this, we have to change both (a) and (b), but will not need to modify (c) at all. Since this type of solution depends on interpreting NL-sentences containing the word “always” as shorthand for longer sentences that reveal hidden conditions of the form “since (such and such)” on which the original sentence was supposed to rely, we call it a “hidden variables” solution. As we will see, the problem with a hidden variables solution is not satisfiability but



$\mathcal{M}, t \Vdash S(\phi, \psi)$  iff there is an  $s < t$  so that  $\mathcal{M}, s \Vdash \phi$  and  $\mathcal{M}, u \Vdash \psi$  for all  $s < u < t$ .  
 $\mathcal{M}, t \Vdash U(\phi, \psi)$  iff there is an  $t < s$  so that  $\mathcal{M}, s \Vdash \phi$  and  $\mathcal{M}, u \Vdash \psi$  for all  $t < u < s$ .

**Figure 4.2:**  $\mathcal{L}_{SU}^P$  Semantics

fidelity. That is, even if  $\mathcal{L}_{SU}^P$  is a viable solution (which we will cast into doubt without ruling out completely), it demands we read claims like (1) as declaring something other than the genuine atemporality of mathematicalialia.

### $\mathcal{L}_{SU}^P$

A hidden variables solution is well-motivated by observing that for many NL-sentences  $\phi$ , the claim “it has always been the case that  $\phi$ ” is more naturally interpreted using a “since...” clause than by taking the “always” literally. For example, the sentence “I have always loved baseball” is not well translated as “ $H(I \text{ love baseball})$ .” Rather, “I have always loved baseball” is more plausibly rendered “since I can remember, I have loved baseball.” This motivates interpreting some sentences containing the NL-temporal quantifier “always” as shorthand for claims with hidden “since...” or “until...” clauses.

To accommodate “since...” and “until...” sentences in the formal theory we must extend  $\mathcal{L}^P$ . We do this by adding operators  $S$  and  $U$  (for since and until) to the alphabet of  $\mathcal{L}^P$ . Examining their NL-counterparts, we see the  $S$  and  $U$  operators will be of an unusual sort – they take not one but *two* sentences as arguments. Because of this, the grammar of the extended language is a bit complicated to specify. However, specifying appropriate syntactic rules can be done (and has been done, see [45] for the first such exposition) and done in roughly the way one would expect. We call the extended language  $\mathcal{L}_{SU}^P$ . Semantically,  $\mathcal{L}_{SU}^P$  sentences are interpreted in  $\mathcal{L}^P$ -models, with the novel features being translated as per the two clauses in Figure 4.2. We read  $U(\phi, \psi)$  as “until  $\phi$  happens  $\psi$  will be the case” and  $S(\phi, \psi)$  as “since  $\phi$  happened  $\psi$  has been the case.”

### 4.3.6 $\mathcal{L}_{SU}^P$ as a Solution

Before examining  $\mathcal{L}_{SU}^P$  as a solution to the problem of interpreting NL-sentences like (1), we state a theorem that will be needed later. It is easy to prove, so its proof is omitted.

THEOREM:

If  $\mathcal{M}, t \Vdash S(\phi, \psi)$ , and  $s < t$  is such that  $\mathcal{M}, s \Vdash \phi$  and  $\mathcal{M}, u \Vdash \psi$  for all  $s < u < t$ , then if  $s < t' < t$ ,  $\mathcal{M}, t' \Vdash S(\phi, \psi)$  as well.

The first difficulty in implementing a translation of NL-sentences into  $\mathcal{L}_{SU}^P$  is determining just *which* NL-occurrences of the word “always” should be translated into  $\mathcal{L}_{SU}^P$  using  $S$  or  $U$  rather than  $H$  or  $G$ . This is a serious but not particularly interesting difficulty, being as it is a case of ordinary NL-ambiguity. Spelling out the details just a little, there are easily fabricated NL-sentences containing the word “always,” for which it is unclear whether they should be translated using  $H$  or using a “since...” clause.<sup>1</sup>

Putting this aside, a further difficulty arises concerning the interpretation of NL-sentences that *should* be interpreted using  $S$  or  $U$ . In these cases, the interpretation demands the “always” clause be rewritten to take account of the *correct* hidden conditions on which the  $S$  or  $U$  operators should depend. In general it will be very difficult to know what these conditions should be.

This can be seen even in the example already examined. There are natural conditions we might propose adding to the sentence “there has always been a prime less than three” when we interpret it as containing a hidden “since...” clause. For example, if we assume mathematical entities are social constructs, then we could argue this sentence is shorthand for “since two and three were first declared to exist, there has been a prime less than three.” This, in turn, is naturally translated into  $\mathcal{L}_{SU}^P$  as  $S(\phi, Q)$ , with  $\phi$  being the proposition representing the sentence “two and three have been declared to exist.” Things get more complicated when we examine the NL-sentence

“it has always been the case that it has always been the case  
that there is a prime less than three.” (4)

---

<sup>1</sup> This is especially true once we adopt the practice of treating mathematical entities like social constructs.

If we follow the lead above and translate “there has always been a prime less than three” as “ $S(\phi, Q)$ ,” then it seems (2) should be translated as “it has always been the case that  $S(\phi, Q)$ .” But the occurrence of “always” in this sentence is no more naturally translated as  $H$  than the other. So (2) seems best translated as “ $S[\psi, S(\phi, Q)]$ ” for an appropriate sentence  $\psi$ . The next few paragraphs show just how difficult finding such a  $\psi$  is.

Suppose “it has always been the case that there is a prime less than three” means something like “since  $\phi$  happened, there has been a prime less than three.”<sup>2</sup> Then it seems (2) must mean something like “since  $\phi$  happened it has been the case that since  $\phi$  happened there has been a prime less than 3.” Thus, the natural candidate for  $\psi$  is either  $\phi$  or something logically equivalent to  $\phi$ .

But we turned to  $\mathcal{L}_{SU}^P$  to allow us to translate those NL-sentences like (1) that matter for Cole’s account into  $\mathcal{L}_{SU}^P$  as  $\mathcal{L}_{SU}^P$ -sentences that are at least *satisfiable*. If our translation scheme demands  $\psi$  be logically equivalent to  $\phi$ , however, then when we translate (1) into  $\mathcal{L}_{SU}^P$ , the resulting sentence is *unsatisfiable*.

To see this, notice  $S(\phi, Q) \wedge \neg S(\psi, S(\phi, Q))$ , is satisfiable if and only if there is a model  $\mathcal{M}$  and time  $t$  for which

$$\begin{aligned} \mathcal{M}, t \Vdash S(\phi, Q), \text{ and} \\ \mathcal{M}, t \not\Vdash S[\psi, S(\phi, Q)] \end{aligned}$$

From these we conclude

- (1) There is a time  $s$  so that
  - (a)  $s < t$ ,
  - (b)  $\mathcal{M}, s \Vdash \phi$ , and
  - (c)  $\mathcal{M}, u \Vdash Q$  for all  $s < u < t$ , and

- (2) For all  $v$ , if

---

<sup>2</sup> It has been suggested to me several times that perhaps we could combine methods here: why not interpret (2) as  $S(\phi, HQ)$ ? The problem with this approach is that, when one works out the semantic details,  $\mathcal{M}, t \Vdash S(\phi, HQ)$  if and only if  $\mathcal{M}, t \Vdash \neg H\neg\phi \wedge HQ$ . But then a model of (1) would still require an  $\mathcal{L}^P$ -model modelling the sentence  $HQ \wedge \neg HHQ$ , which we’ve already seen is impossible.

- (a)  $v < t$  and
- (b)  $\mathcal{M}, w \Vdash S(\phi, Q)$  for all  $v < w < t$ , then
- (c)  $\mathcal{M}, v \not\models \psi$ .

But these rule out the possibility that  $\phi$  is logically equivalent to  $\psi$ : The theorem proved above gives that if  $s < w < t$ , then  $\mathcal{M}, w \Vdash S(\phi, Q)$ . However, if  $\phi$  is logically equivalent to  $\psi$ , (1b) guarantees the negation of (2c) when  $v = s$ . Thus, not only must we provide a general strategy for finding the correct conditions  $\phi$  under which an NL-sentence containing “always” should be interpreted “ $S(\phi, \psi)$ ” or “ $U(\phi, \psi)$ ” rather than “ $H(\psi)$ ” or “ $G(\psi)$ ,” we must also contend with sentences like (1), where the most natural interpretation would have the outside and inside occurrences “always” subject to the same condition  $\phi$ , but whose satisfiability depends on us interpreting them otherwise.

But solving these problems would still not allow  $\mathcal{L}^{2P}$  to deal with a more serious challenge: the hidden variables account is explicitly contrary to the stated purpose of Cole’s social constructivism. Cole’s aim is not to find an acceptable way to *translate away* claims about the atemporality of mathematicalialia, but rather to find ways to genuinely make sense of atemporal mathematicalialia in a socially-constructive setting. Cole introduces temporal profiles precisely to allow him to account for the atemporality of mathematicalialia. It was interpreting the complex NL-sentences that arose from the introduction of temporal profiles that led us to consider  $\mathcal{L}_{SU}^P$ . So, even if the problems plaguing the implementation of a translation of natural language into  $\mathcal{L}_{SU}^P$  could be overcome, the hidden variables account does not offer a genuine option for taking seriously claims concerning the atemporality of mathematicalialia.

## 4.4 Two-Dimensional Solutions

One might complain at this point that neither solution examined has taken the full (NL-)syntax of our exemplar sentences to heart. Sentence (1) and, more importantly, all of Cole’s own examples can be read as containing elements of the general form “**At**  $t_1$ ,  $\phi$  was true **of**  $t_2$ ...” Such sentences are indexed both by a temporal *perspective* (the time **at** which  $\phi$  is being asserted) and a temporal *location* (the time **of** which  $\phi$ ’s truth is asserted). Cole provides a helpful example to make clear that the distinction between temporal perspective and temporal location is not an entirely artificial construction:

Major League Baseball’s regulations permit teams to place a certain number of players on a 15-day disabled list. League regulations also permit placing a player on the disabled list *retroactively*. So, as Cole points out, while

“on the Friday after his injury a player might not be officially on the 15 day disabled list, yet on the following Monday he might be retroactively placed on this list going all the way back to the previous Monday. Hence, who appears on the official disabled lists for a given day can vary over time.” [7, p. 20]

So, if we ask questions about the disabled list with reference only to the particular temporal *location* ‘the Friday after player X’s injury’, the answers we get will themselves vary with our temporal *perspective*.

The remainder of the section will proceed in a somewhat backward manner. First, I build the appropriate setting for a “doubly-temporal” semantics that can deal with both temporal locations and temporal perspectives. After this, I produce syntax that allows us to take advantage of some of these new semantic resources. We then see these resources allow us to produce a (unique) translation of (1) that (a) takes account of its full NL-syntax and (b) is satisfiable.

However, in section 5 we show that while the translation we produce of (1) is satisfiable in the technical sense of admitting *some* model, it nonetheless cannot be true *in the real world*. In the course of examining this, however, we do find a way to use  $\mathcal{L}^{2P}$ -semantics to produce a hybrid of social constructivism and fictionalism that allows one to explain how mathematicalialia can be social constructs that *appear* atemporal while not actually *being* atemporal.

#### 4.4.1 Semantics

Introducing a second temporal variable to the semantics demands we change everything, but in mostly unsurprising ways. Models remain triples  $\langle \mathcal{T}, <, v \rangle$ , but the domain of  $v$ , rather than being  $\mathcal{T}$ , is now  $\mathcal{T} \times \mathcal{T}$ . Valuations thus map pairs of times  $\langle t_1, t_2 \rangle$  to functions from **Prop** to  $\{T, F\}$ . In symbols,

$$v : \mathcal{T} \times \mathcal{T} \longrightarrow \{f : \mathbf{Prop} \longrightarrow \{T, F\}\}$$

$$\langle t_1, t_2 \rangle \longmapsto (\mathbf{Prop} \ni P \longmapsto v_{t_1, t_2}(P) \in \{T, F\})$$

If we maintain  $\mathcal{L}^P$  for our language despite the semantic change, then when we turn to defining truth, validity, etc., it seems we must make a choice about which temporal variable  $H$  (and  $G$ , for that matter) will be evaluated with respect to. We could either say

- (a)  $\mathcal{M}, t_1, t_2 \Vdash H\phi$  iff  $\mathcal{M}, s, t_2 \Vdash \phi$  for all  $s < t_1$  or
- (b)  $\mathcal{M}, t_1, t_2 \Vdash H\phi$  iff  $\mathcal{M}, t_1, s \Vdash \phi$  for all  $s < t_2$ .

Now,  $H$  was intended to capture (or at least carefully model) the meaning of the NL-phrase “it has always been the case that...”. On examining (a) and (b), though, it is unclear which best captures this meaning.

To help clarify the matter, we adopt the convention of reading the semantic sequent “ $\mathcal{M}, t_1, t_2 \Vdash \phi$ ” in the following way: the first temporal variable will specify with respect to which temporal perspective  $\phi$  is being viewed; the second temporal variable will specify at which temporal location  $\phi$ ’s truth is being asserted. Thus, “ $\mathcal{M}, t_1, t_2 \Vdash \phi$ ” should be interpreted as “in model  $\mathcal{M}$  and from the point of view of  $t_1$ ,  $\phi$  is true as a statement about  $t_2$ ,” or, briefly, “in  $\mathcal{M}$  from  $t_1$ ,  $\phi$  is true of  $t_2$ .”

Given this reading, changing the temporal variable we interpret  $H$  with respect to gives it very different meanings. Interpretation (a) reads “ $H\phi$ ” as “from the perspective of any prior time,  $\phi$  is true as a statement about the temporal location under consideration.” (b), on the other hand, reads “ $H\phi$ ” as “from the temporal perspective under consideration,  $\phi$  is true about any prior time.”

Neither of these interpretations is obviously “the right one.” Rather, both seem to capture more precise fragments of the full meaning of the NL-phrase “it has always been the case that...” than were available in the languages previously examined. So it seems we ought not choose between (a) and (b), but rather should expand our language to put both possibilities at our disposal. To this end, in the remainder we let  $H$  stand for the  $\mathcal{L}^P$  interpretation of “always,”  $\mathcal{H}$  for  $\mathcal{L}^{2P}$ -interpretation (a) and  $\mathfrak{H}$  for  $\mathcal{L}^{2P}$  interpretation (b).

$$\begin{array}{ll}
\mathcal{M}, t_1, t_2 \Vdash \mathcal{G}\phi & \text{iff } \mathcal{M}, s, t_2 \Vdash \phi \text{ for all } t_1 < s \\
\mathcal{M}, t_1, t_2 \Vdash \mathcal{H}\phi & \text{iff } \mathcal{M}, s, t_2 \Vdash \phi \text{ for all } s < t_1 \\
\mathcal{M}, t_1, t_2 \Vdash \mathfrak{G}\phi & \text{iff } \mathcal{M}, t_1, s \Vdash \phi \text{ for all } t_2 < s \\
\mathcal{M}, t_1, t_2 \Vdash \mathfrak{H}\phi & \text{iff } \mathcal{M}, t_1, s \Vdash \phi \text{ for all } s < t_2
\end{array}$$

**Figure 4.3:**  $\mathcal{L}^{2P}$ -Semantics

To conclude the construction, let the language  $\mathcal{L}^{2P}$  of Priorean Doubly Temporal Logic be the closure of  $\mathcal{L}$  under the addition of the four operators  $\mathcal{H}$ ,  $\mathcal{G}$ ,  $\mathfrak{H}$ , and  $\mathfrak{G}$ . The novel features of  $\mathcal{L}^{2P}$ -semantics are given in Figure 4.3. The non-novel features are treated analogously to  $\mathcal{L}^P$ .

#### 4.4.2 An Example

Consider the sentence  $\mathfrak{H}\phi \wedge \neg\mathcal{H}\mathfrak{H}\phi$ . We translate this from  $\mathcal{L}^{2P}$  into NL in stages as follows:

- (1)  $\mathfrak{H}\phi$  reads “from this perspective,  $\phi$  is true of all prior times.”
- (2)  $\neg\mathcal{H}\mathfrak{H}\phi$  reads “there is a previous perspective where  $\mathfrak{H}\phi$  fails to be true.”
- (3) Combining the two, we arrive at a sentence best paraphrased “from this perspective,  $\phi$  is true of all prior times, but there is a previous time at whose perspective  $\phi$  would have been false for at least one yet earlier time.

Thus the language  $\mathcal{L}^{2P}$  allows expressive power sufficient to describe very nuanced interactions between truth and time. In particular, notice if  $\mathcal{M}, t, s \Vdash \neg\mathcal{H}\mathfrak{H}\phi$ , then there is a time  $u < t$  for which  $\mathcal{M}, u, s \Vdash \neg\mathfrak{H}\phi$ . From this we conclude there is a time  $v < s$  for which  $\mathcal{M}, u, v \Vdash \neg\phi$ . It is worth noting that this is compatible with it nonetheless being the case that  $\mathcal{M}, t, s \Vdash \mathfrak{H}\phi$ . This will be carefully demonstrated in a moment.

#### 4.4.3 The $\mathcal{L}^{2P}$ -Solution

In  $\mathcal{L}^P$ , NL-sentences of the form “ $\phi$  has always been the case” were interpreted in the only available way; namely as  $H\phi$ . In  $\mathcal{L}^{2P}$ , such NL-sentences can be interpreted as

Quantifier	Language	Truth Conditions
$H$	$\mathcal{L}$	$\mathcal{M}, t \Vdash H\phi$ iff $\mathcal{M}, s \Vdash \phi$ for all $s < t$
$\mathcal{H}$	$\mathcal{L}^{2P}$	$\mathcal{M}, t_1, t_2 \Vdash \mathcal{H}\phi$ iff $\mathcal{M}, s, t_2 \Vdash \phi$ for all $s < t_1$
$\mathfrak{H}$	$\mathcal{L}^{2P}$	$\mathcal{M}, t_1, t_2 \Vdash \mathfrak{H}\phi$ iff $\mathcal{M}, t_1, s \Vdash \phi$ for all $s < t_2$

**Figure 4.4:** Guide to Important Temporal Quantifiers

either  $\mathcal{H}Q$  or  $\mathfrak{H}Q$ . (Figure 4.4 summarizes the meanings of these three operators – the reader may find it helpful to refer back to this chart in the remainder.) The presence of these options allows for an acceptable interpretation of (1). In particular,  $\mathfrak{H}Q \wedge \neg\mathcal{H}\mathfrak{H}Q$  is the unique plausible interpretation of (1) that is also satisfiable.

### Satisfiability

Choose an  $\mathcal{L}^{2P}$ -model  $M = \langle \mathbb{R}, <, v \rangle$ , where  $\langle \mathbb{R}, < \rangle$  is the real numbers with their usual ordering, and  $v$  satisfies

$$v_{1,s}(Q) = T \text{ for all } s < 1, \text{ and}$$

$$v_{0,s}(Q) = F \text{ for all } s.$$

Observe that since  $v_{1,s}(Q) = T$  for all  $s < 1$ , we have that  $M, 1, s \Vdash Q$  for all  $s < 1$ . Thus, reading from the appropriate line in Figure 4.3, we see that  $M, 1, 1 \Vdash \mathfrak{H}Q$ . Also, from  $v_{0,0}(Q) = F$  we conclude  $M, 0, 0 \not\Vdash Q$ . Thus  $M, 0, 1 \not\Vdash \mathfrak{H}Q$ . But this gives us that  $M, 1, 1 \not\Vdash \mathcal{H}\mathfrak{H}Q$ . Combining these two results, we see  $M, 1, 1 \Vdash \mathfrak{H}Q \wedge \neg\mathcal{H}\mathfrak{H}Q$ , so  $\mathfrak{H}Q \wedge \neg\mathcal{H}\mathfrak{H}Q$  is satisfiable.

### Uniqueness

From a technical perspective, there are eight potential translations of sentence (1) into  $\mathcal{L}^{2P}$ :

- |  |  |
|--|--|
| (i) $\mathcal{H}Q \wedge \neg\mathcal{H}\mathcal{H}Q$    | (iii) $\mathcal{H}Q \wedge \neg\mathfrak{H}\mathcal{H}Q$ |
| (ii) $\mathfrak{H}Q \wedge \neg\mathcal{H}\mathfrak{H}Q$ | (iv) $\mathcal{H}Q \wedge \neg\mathcal{H}\mathfrak{H}Q$  |



- |  |   |
|--|---|
| (v) $\mathcal{H}Q \wedge \neg\mathfrak{H}\mathfrak{H}Q$  | (vii) $\mathfrak{H}Q \wedge \neg\mathfrak{H}\mathcal{H}Q$   |
| (vi) $\mathfrak{H}Q \wedge \neg\mathcal{H}\mathfrak{H}Q$ | (viii) $\mathfrak{H}Q \wedge \neg\mathfrak{H}\mathfrak{H}Q$ |

It is immediate from the semantic rules that (i) and (viii) are not satisfiable; the argument to establish this is precisely parallel to the argument demonstrating  $H\phi \wedge \neg HH\phi$  is not satisfiable in any  $\mathcal{L}^P$ -model. The remaining six sentences are satisfiable; models demonstrating this are not difficult to construct. For plausibility, however, we must not only demonstrate satisfiability but compatibility with the intended meaning of (1) in the context of Cole’s social constructivism.

Throughout the remainder, assume mathematicalalia are social constructs. Let  $t_0$  be any time prior to social construction of the number two, and let  $\mathcal{R}$  be the real world considered as an  $\mathcal{L}^{2P}$ -model. Recall “ $\mathcal{R}, u, v \Vdash Q$ ” means “in the real world, from the perspective of  $u$ ,  $Q$  is true of  $v$ .” Since  $t_0$  is a time prior to the social construction of the number two, we see that  $\mathcal{R}, t_0, s \not\Vdash Q$  for any  $s$  whatsoever.

Next, let  $t_1$  be Friday, June sixth, 2014 at 10:59pm. Observe that

- (a) At  $t_1$ ,  $Q$  was true of  $t_1$ . Since the number two is introduced with an unbounded temporal profile,
- (b) At  $t_1$ ,  $Q$  has always been true.

(b) can plausibly be translated into  $\mathcal{L}^{2P}$  in only two ways: as  $\mathcal{H}Q$  or as  $\mathfrak{H}Q$ . However, if we interpret (b) as  $\mathcal{H}Q$ , then we are left affirming  $\mathcal{R}, t_1, t_1 \Vdash \mathcal{H}Q$ , and so (reading from the corresponding row in Figure 4.3), we see  $\mathcal{R}, s, t_1 \Vdash Q$  for all  $s < t_1$ . But we just saw  $\mathcal{R}, t_0, s \not\Vdash Q$  for any  $s$ . Since  $t_0 < t_1$ , we would thus have both that  $\mathcal{R}, t_0, t_1 \not\Vdash Q$  and  $\mathcal{R}, t_0, t_1 \Vdash Q$ . Thus the interpretation of the first conjunct of (1) as  $\mathfrak{H}Q$  is the only  $\mathcal{L}^{2P}$  option that allows (1) to be compatible with Cole’s social constructivism. This rules out options (iii)-(v). Notice this argument shows it is in general the case that, on pain of contradiction, the only sort of atemporality a social constructivist can attribute to mathematicalalia is of the “ $\mathfrak{H}$ ” variety. This will be an important detail later.

Finally, the second conjunct of (1) is meant as a denial of the at-all-times verity of the first conjunct. Thus, the inside occurrence of “always” in the second conjunct of (1) should be interpreted in the same way the occurrence of “always” in the first conjunct

of (1) is. Only (vi) has the requisite form, so we conclude it is the unique appropriate translation of (1) into  $\mathcal{L}^{2P}$ .

## 4.5 The Real World as an $\mathcal{L}^{2P}$ -model

A moment ago, we very casually made reference to “the real world as an  $\mathcal{L}^{2P}$ -model.” We will now try to make sense of what exactly this phrase means. To do this, we need to specify what features of the world determine the truth of the semantic sequent “ $\mathcal{R}, t_1, t_2 \Vdash \phi$ .”

There are two broad approaches one could take to this problem. First, one could suppose it in fact takes two times to specify which atomic propositions are true in the real world – that is, that propositional truth is underspecified by considering only a temporal perspective or only a temporal location. In this case, the fact about the world that makes “ $\mathcal{R}, t_1, t_2 \Vdash \phi$ ” be a truth is simply the fact that  $\phi$  is true at the temporal location  $t_2$  from the perspective of  $t_1$ . I show below that this method of solving the problem of the atemporality of socially constructed mathematical objects will not suffice – it neither captures the phenomenology of mathematics nor allows it to play the role in causal explanation that we want it to.

The second option is to suppose truth takes only one time to specify. In this case, the correct temporal model of the world is an  $\mathcal{L}^P$ -model. This has as a consequence that mathematical objects are either not actually atemporal or are not social constructs.

I call the first option the complicated world option, and the second the simple world option. In the remainder of this section I expand on these options. While we will see that neither allows  $\mathfrak{H}Q \wedge \neg\mathcal{H}\mathfrak{H}Q$  to be *true* in the real world, in the second case we can interpret  $\mathcal{L}^{2P}$ -semantics in a way that allows us to recover more of what the social-constructivist is presumably after than we can in the first. The last section of the paper explores this option more.

### 4.5.1 A Complicated World

Assume truth takes two times to specify. We observed already if  $\mathcal{R}, t, s \Vdash \neg\mathcal{H}\mathfrak{H}Q$ , then there is a time  $u < t$  and a time  $v < s$  for which  $\mathcal{R}, u, v \Vdash \neg Q$ . Thus, in order to admit  $\mathfrak{H}Q \wedge \neg\mathcal{H}\mathfrak{H}Q$  as not just satisfiable but *true*, we must admit a pair of times  $u, v$ ,

for which it is the case that from the perspective of  $u$ , the statement “there is a prime less than three” is false as a statement about the world at  $v$ .<sup>3</sup> Thus, pursuing the complicated world option demands we accept there are temporal perspectives in which arithmetic “truths” are at best temporarily true.

In a complicated world, then, mathematicians, while free to assert mathematical truths as holding at all temporal locations, would not be free to simultaneously hold that these same “truths” would remain true from other temporal perspectives. But this is exactly the opposite of what the phenomenology of doing mathematics leads us to believe. It is a feature of our experience as users of mathematics that we are entirely free not only to assert mathematical truths, but to assert them as true both from any temporal perspective and about any temporal location. For example, there is nothing wrong with the mathematical claims of the astrophysicist who, when explaining her work to a popular audience, says “what I’ve discovered is that if one were present in the solar system several billion years ago and found that the density of the cloud of gas accumulating there was between  $x$  and  $y$ , then one would be able to infer a star would form nearby and that 10 billion years later it would be a red giant.” But notice this claim relies on changing both temporal *perspectives* (to a time several billion years ago) and temporal *locations* (to a time several billion years hence), something we very casually expect the mathematics to withstand. Exactly this type of move is not generally justified if we admit as true claims of the form  $\neg\mathcal{H}\mathfrak{H}Q$ .

Even putting this aside, the complicated world option still has problems. In particular, it prevents Cole’s account from being a response to a crucial part of Dieterle’s objection. As mentioned above, Dieterle called social constructivism to task for its inability to legitimate the use of mathematics in descriptions of, for example, the very early universe. This is again at stake: as we saw above, the only type of  $\mathcal{L}^{2P}$ -atemporality we can allow mathematicalities is of the  $\mathfrak{H}$ -variety, but many scientific claims are at least *prima facie* of a form incompatible with this. For example, returning to our astrophysical stalking horse, suppose we ask “why is our sun not a blue dwarf?” Again, the likely answer takes the form “when the Sun was forming, the cloud it accreted from had a

---

<sup>3</sup> The astute reader will at this point wonder why we should care about what the truth is about the world at time  $v$  from any perspective other than  $v$ . What seems relevant is what the natural diagonal  $\mathcal{L}^P$ -model hidden inside the real-world  $\mathcal{L}^{2P}$ -model gives us. This is an important point which we will return to shortly.

density between  $x$  and  $y$  and a mass of available material between  $w$  and  $z$ . This caused the development to proceed in such and such a way.” Notice that the fact appealed to in this explanation is not the density of the cloud *from our current temporal perspective*, but the density of the cloud *at that time*. That is, we do not attribute causal force to ways things are, at that time, from our current perspective, but to ways things are, at that time, from the perspective of that time.<sup>4</sup>

In more detail: let  $v$  be today,  $w$  some time early in the development of the solar system, and let  $d(t)$  give the density of matter in the solar neighborhood as a function of temporal location. Then we would not consider  $\mathcal{R}, v, w \Vdash x < d(w) < y$  a determinative factor in whether the sun is a blue dwarf today. What determined that the Sun did not become a blue dwarf was that  $\mathcal{R}, w, w \Vdash x < d(w) < y$ . But exactly this latter fact is unavailable on the complicated-world interpretation of Cole’s account – we cannot infer features of the ordering of the real numbers (or whether there were real numbers) from the perspective of a time in the early universe because the only information we have about the truth of  $Q$  at other times involves the truth of  $Q$  at other temporal *locations* from our current temporal perspective. Formally this can be recognized from the fact that even if  $w < v$ , we cannot infer  $\mathcal{R}, w, w \Vdash \phi$  from  $\mathcal{R}, v, v \Vdash \mathfrak{H}\phi$  or from  $\mathcal{R}, v, v \Vdash \mathfrak{H}\phi$ .

#### 4.5.2 A Simple World

If we suppose truth requires only one time to specify, then mathematicalialia cannot be both atemporal and social constructs. This was the lesson of the third section of the essay and was the reason we moved to  $\mathcal{L}^{2P}$ . So in this case, Cole simply cannot be both a social constructivist and admit that mathematicalialia actually are atemporal.<sup>5</sup> But this is an unsatisfactory stopping point because (a) the extra structure in  $\mathcal{L}^{2P}$ -models feels like it’s tracking *something*, even if that something isn’t truth simpliciter, and (b) it seems social constructivism should at the very least be able to make sense of why it is that mathematicalialia *appear to be* atemporal. By appeal to an interesting technical construction, we will be able to resolve both of these issues at once.

---

<sup>4</sup> Thanks to an anonymous referee for helping clarify this passage.

<sup>5</sup> The reader whose instinctive response to this is to reach for another temporal dimension may wish to read the appendix at this time.

Given an  $\mathcal{L}^{2P}$ -model  $\mathcal{M} = \langle \mathcal{T}, <, v \rangle$ , fix a particular time  $\tau$  and consider the structure  $\mathcal{M}_\tau := \langle \mathcal{T}, <, v_\tau \rangle$  where  $v_\tau(t) = v_{\tau,t}$ .  $\mathcal{M}_\tau$  is an  $\mathcal{L}^P$ -model that is, intuitively, “stacked up” at the time  $\tau$ . For each  $\tau \in T$ , this construction gives an  $\mathcal{L}^P$ -model  $\mathcal{M}_\tau$ , that represents the model of the world available at  $\tau$ . It is reasonable to interpret the phrase “ $\mathcal{M}_\tau, t \Vdash \phi$ ” as meaning “at  $\tau$ ,  $\phi$  appeared to be true at  $t$ .”

### Examples

Example 1. Suppose  $\mathcal{R}_t, t \Vdash HQ$ . Then, for all  $s < t$ ,  $\mathcal{R}_t, s \Vdash Q$ . We conclude that  $\mathcal{R}, t, s \Vdash Q$  for all  $s < t$ . That is,  $\mathcal{R}, t, t \Vdash \mathfrak{H}Q$ .

Example 2. Suppose  $\mathcal{R}, t, t \Vdash \mathfrak{H}\mathfrak{H}Q$ . Then, for all  $s < t$  and for all  $u < s$ ,  $\mathcal{R}, s, u \Vdash Q$ . But this just means that for all  $u < s < t$ ,  $\mathcal{R}_s, u \Vdash HQ$ .

In addition to all the  $\mathcal{M}_\tau$ , there is also a natural “diagonal”  $\mathcal{L}^P$  model  $\mathcal{M}_\Delta$  that we can associate to  $\mathcal{M}$ .  $\mathcal{M}_\Delta := \langle \mathcal{T}, <, v_\Delta \rangle$ , where  $v_\Delta(t) = v_{t,t}$ . Notice this definition gives  $\mathcal{M}_\Delta, t \Vdash \phi$  if and only if  $\phi$  is true of  $t$  at  $t$ .

### 4.5.3

Let’s recall the situation: we are assuming that mathematical entities are social constructs. We are also assuming that the structure of propositional truth is such that it takes only one time to fully specify it. The world is thus well-modeled by an  $\mathcal{L}^P$ -model we will call  $\mathcal{W}$ . We’ve seen that such a model will not validate both social constructivism and the claim that mathematical entities are atemporal. So *if* we are to be social constructivists, *then* we must deny that mathematical entities are atemporal.

Recall also that social constructivism (as Cole’s brand of it makes explicit) commits one to the belief that certain normative commitments – for example the normative commitments we have involving the correct use of number-terms – suffice to make it the case that certain entities do in fact exist. Cole’s response to Dieterle made explicit that a social constructivist must also be committed to these same normative commitments not only entailing that certain objects *do* exist, but also that they have existed at other times. As we’ve now seen, such claims seem untenable from a semantic perspective.

The technical construction above showed we can read *part of* the structure of  $\mathcal{L}^{2P}$ -models as giving a description of what appears to be true from a given perspective. But

Model	Description
$\mathcal{W}$	The Real World as an $\mathcal{L}^P$ -model
$\mathcal{R}$	The Real World as an $\mathcal{L}^{2P}$ -model, semantics read <i>normatively</i>
$\mathcal{R}_t$	The correct $\mathcal{L}^P$ -model of the world given the normative commitments in play at time $t$
$\mathcal{R}_\Delta$	The “diagonal” model derived from $\mathcal{R}$ . We’ve seen that if $\mathcal{R}_\Delta, t \Vdash Q$ , then $\mathcal{W}, t \Vdash Q$ .

**Figure 4.5:** Key to Translation

the reason it sometimes *appears* as if numbers exist is because at such times one has the sort of normative commitments that the social constructivist claims suffice to make it the case that numbers exist. Thus, perhaps a more useful perspective to adopt is to see the fragments of  $\mathcal{L}^{2P}$  semantics we characterized as giving a description of what *appeared to be true* as giving descriptions of what one *ought* to say is true, given the normative roles one is allowing various facets of reality to play in one’s life.

Adopting this perspective amounts to reading “ $\mathcal{M}, s, t \Vdash \phi$ ” as “in  $\mathcal{M}$ , given the normative commitments in play at  $s$ , one ought to claim  $\phi$  is true of time  $t$ .” Of course, it is entirely consistent with this reading of the sequent that at time  $t$ ,  $\phi$  might fail to be true. But this need not involve us in a contradiction, because we are only committed to endorsing  $\phi$  as true about  $t$  so long as the norms in play at  $s$  remain in play.

Adopting this normative reading of “ $\mathcal{M}, s, t \Vdash \phi$ ” has a further and rather unexpected result. Recall we showed above the sentence “ $\mathfrak{H}Q \wedge \neg\mathcal{H}\mathfrak{H}Q$ ” was the unique appropriate translation of (1) into  $\mathcal{L}^{2P}$ . This sentence can also be independently arrived at simply by thinking about what needs to be true in a *correct*  $\mathcal{L}^{2P}$ -model of the world (with the normative reading of its semantics), given we accept that (a) social constructivism about mathematicalities is *true*, so that (b) if, at a given time  $t$  we *ought* to say that, e.g. a prime less than three exists at that time, then a prime less than three *actually does* exist at  $t$ . In terms of our newly-adopted normative reading of  $\mathcal{L}^{2P}$  semantics, (b) allows us to pass from  $\mathcal{R}_\Delta, t \Vdash Q$  to  $\mathcal{W}, t \Vdash Q$ .

To arrive at “ $\mathfrak{H}Q \wedge \neg\mathcal{H}\mathfrak{H}Q$ ”, we argue as follows (the reader may find the Key to Translation given in Figure 4.5 helpful while reading this argument) Observe that since we are assuming that social constructivism is *true*, we must have that, if  $t$  is the current

time, then  $\mathcal{W}, t \not\vdash HQ$  – social constructivism is incompatible (in  $\mathcal{L}^P$ -models) with atemporal mathematicalialia. Notice if  $\mathcal{R}_\Delta, t \vdash HQ$ , then since this entails  $\mathcal{R}_\Delta, s \vdash Q$  for all  $s < t$ , which in turn entails that  $\mathcal{W}, t \vdash Q$  for all  $s < t$ , we get  $\mathcal{W}, t \vdash HQ$ .

But example 2 just showed that  $\mathcal{R}_\Delta, t \vdash HQ$  (and, thus, given the previous paragraph,  $\mathcal{W}, t \vdash HQ$ ) follows from the assumption that  $\mathcal{R}, t, t \vdash \mathcal{H}\mathfrak{H}Q$ . So to maintain our social constructivism, we must have that  $\mathcal{R}, t, t \vdash \neg\mathcal{H}\mathfrak{H}Q$ .

On the other hand, there is no denying that we recognize rules for the correct use of number-terms well beyond the range of times for which there existed a society present to construct the numbers they refer to. That is, we *do in fact* grant mathematicalialia a normative role in our theorizing about, for example, past times. Among the things this commits us to is the claim that, given the norms we *currently* grant mathematicalialia, one *ought to* endorse the existence of a prime less than three at all past times. That is, that if  $t$  is the current time, then  $\mathcal{R}_t, t \vdash HQ$ . But Example 1 showed that a consequence of this is that  $\mathcal{R}, t, t \vdash \mathfrak{H}Q$ . Together with the conclusion of the previous paragraph, then, we get  $\mathcal{R}, t, t \vdash \mathfrak{H}Q \wedge \neg\mathcal{H}\mathfrak{H}Q$

Thus, simply from (a) and (b), the constructions in the previous two paragraphs lead us to naturally expect that if  $\mathcal{R}$  is the *correct*  $\mathcal{L}^{2P}$ -model of the world, and we read its semantics *normatively*, then  $\mathcal{R}, t, t \vdash \mathfrak{H}Q \wedge \neg\mathcal{H}\mathfrak{H}Q$ . On the other hand, this sentence seemed, given the details, the only natural interpretation of (1) – this confluence of evidence for the correctness of the sequent “ $\mathcal{R}, t, t \vdash \mathfrak{H}Q \wedge \neg\mathcal{H}\mathfrak{H}Q$ ” is really quite remarkable and satisfying.

## 4.6 Fictionalism

So if we assume mathematicalialia are social constructs, then we must abandon their atemporality. We said before that this seems to put much of our very best science at risk. We should examine this last claim more closely.

The worry cannot *merely* be that we want scientific theories to not entail falsehoods, for the simple reason that scientific theories giving rise to falsehoods is the norm. Even our very best scientific theories give us, even in the ranges where they are maximally accurate, only approximate truths. From the semantic perspective, though, approximate truths are just (a fancy type of) falsehoods. So the idea that our scientific theories

ought not entail falsehoods is not quite accurate – we want to ensure they do not entail *contradictions*. Since (classically, at least) anything follows from a contradiction, a theory that entails contradictions is much worse than simply *wrong* – it’s *useless*. What we saw above, however, was that the assumption that mathematicalalia are both socially constructed and atemporal led to genuine contradictions, and the assumption that mathematicalalia are atemporal does seem to be needed in many of our best scientific theories.

But it is a matter of fact that (as we just pointed out), as a society we recognize rules for the correct use of number-terms well beyond the range of times for which there existed a society present to construct them. So what are we to do if we are convinced (along with Cole) that this recognition commits us to (a) admitting that numbers exist, and (b) making sense of their existence in a way compatible with the atemporality we seem to grant them.

The solution is as follows:  $\mathcal{R}_\tau$  provides a model of the sentences we ought to endorse, given the normative commitments in play at  $\tau$ . Since our best scientific theories need to assume mathematicalalia are atemporal, we infer that  $\mathcal{R}_\tau$  captures the structure of the world under the hypothesis that atemporal (and, *eo ipso* non-socially constructed) mathematicalalia exist. If social constructivism is true, this hypothesis is literally *false*. Thus, at least some existential claims endorsed by some of the  $\mathcal{R}_\tau$  are literally false. But if these hypotheses are literally false, one would expect their negations to also be entailed by an accurate account of the world like  $\mathcal{R}_\tau$ . This, of course, would then mean that  $\mathcal{R}_\tau$  entailed a contradiction and was thus trivial.

The above technical construction showed that we can get around this by supposing that we are using distinct fragments of the total  $\mathcal{L}^{2P}$ -semantics governing the meaning of our terms when we make the claims “mathematicalalia are social constructs” and “mathematicalalia are atemporal.” In particular, the  $\mathcal{R}_\tau$  make sense of the latter claim and  $\mathcal{R}_\Delta$  makes sense of the former.  $\mathcal{L}^{2P}$  thus gives us the resources to affirm that mathematicalalia are social constructs while nonetheless allowing us to have non-trivial scientific theories which behave as if they aren’t. Collapse is avoided by our not affirming *contradictions*: in  $\mathcal{R}_\tau$ , this means we allow mathematicalalia that are atemporal but not social constructs. In  $\mathcal{R}_\Delta$ , on the other hand, mathematicalalia are socially-constructed but not atemporal. In neither “dimension” do we get a contradiction, and since the



semantic theory does not allow inferences across dimensions, collapse is avoided. Thus, on the semantic theory offered, despite their rather Moore-ian paradoxicality we can affirm as true all of the following:

- (i) There is a prime less than three,
- (ii) According to our best evidence, a prime less than three has always existed, but
- (iii) A prime less than three has not always existed.

In some sense this puts us on the side of the fictionalist –  $Q$  is a prototypical mathematical existence statement;  $\mathcal{R}_\Delta, t \Vdash \neg Q$  entails  $\mathcal{W}, t \Vdash \neg Q$ , and the latter sequent simply means that  $Q$  is literally false at time  $t$ . Yet despite this, our scientific theories seem to endorse claims that have as consequences that  $\mathcal{R}_\tau, t \Vdash Q$ . On the account provided this can be the case without risk of triviality.

If you're comfortable with such an account and would like to label it a variety of social constructivism, I won't stand in your way. But there is more than a hint of fictionalism here as well.

## Appendix: What If We Had More Time (Dimensions)?

Given Cole was willing to suppose we could collectively declare objects to have whatever temporal profile suited our purposes, I see no reason to suppose he would be unwilling to grant social-constructivism doubly-modal creativity. Thus, Cole could respond to all of the above by claiming that we can “*Declare and collectively recognize* that institutional facets of reality have whatever [doubly-]modal profile best serves our purposes.”

My response to this is roughly as follows: modifying all the above arguments, we could demonstrate that to making of the account that results from allowing socially-constructed doubly-temporal profiles demands moving to triply temporal logic. Cole could respond to this by proposing we adopt socially-constructed triply-temporal profiles, and the back and forth could go on *ad infinitum*.

Being more explicit, Cole could propose that social-constructivism's modal creativity continues to the second temporal variable. Thus, the community responsible for

$\mathcal{M}, f \Vdash P$	iff	$v_f(P) = T$
$\mathcal{M}, f \Vdash \neg\psi$	iff	$\mathcal{M}, f \not\Vdash \psi$
$\mathcal{M}, f \Vdash \phi \wedge \psi$	iff	$\mathcal{M}, f \Vdash \phi$ and $\mathcal{M}, f \Vdash \psi$
$\mathcal{M}, f \Vdash G_n\phi$	iff	$\mathcal{M}, g \Vdash \phi$ whenever $f(n) < g(n)$ but $g(x) = f(x)$ for $x \neq n$
$\mathcal{M}, f \Vdash H_n\phi$	iff	$\mathcal{M}, g \Vdash \phi$ whenever $g(n) < f(n)$ but $g(x) = f(x)$ for $x \neq n$

**Figure 4.6:**  $\mathcal{L}^{\infty P}$  Semantics

introducing a social construct could declare and collectively recognize that it had whatever doubly-temporal profile suited their needs. Nonetheless, we would have to embrace the following:

It is now the case that for all  $t_1, t_2$ ,  $\mathcal{R}, t_1, t_2 \Vdash Q$ , but it has not always been the case that for all  $t_1, t_2$ ,  $\mathcal{R}, t_1, t_2 \Vdash Q$ . (5)

(3) sounds at least confusing and possibly contradictory, so we need to introduce a third temporal variable to account for the semantics of (3). The arguments from before will carry over, *mutatis mutandis*, to establish that Cole must choose again between fictionalism and non-atemporal mathematicalalia. We could then repeat this process to force a move to quadruply-temporal logic, etc.

Let's skip to the end of this process and see what the results are. We introduce the language  $\mathcal{L}^{\infty P}$  of infinitely temporal logic for this purpose.

#### 4.6.1 Syntax

Alphabetically,  $\mathcal{L}^{\infty P}$  differs from a base language  $\mathcal{L}$  of propositional logic by the addition of a countable infinity of temporal modal operators  $H_1, G_1, H_2, G_2$ , etc. Grammatically,  $\mathcal{L}^{\infty P}$  is the closure of  $\mathcal{L}$  after the addition of all these operators. Natural language equivalents of most of these operators are too complex to be worth spelling out.

#### 4.6.2 Semantics

An  $\mathcal{L}^{\infty P}$ -model is a triple  $\langle \mathcal{T}, <, V \rangle$ , where

- $\langle \mathcal{T}, < \rangle$  is a poset (partially ordered set – i.e.  $<$  is an irreflexive and transitive

relation on  $\mathcal{T}$ ) representing the “times” or “moments” together with their ordering and

- $V$  is a function from  $\mathcal{T}^{\mathbb{N}} = \{f : \mathbb{N} \rightarrow T\}$  to the set of functions from **Prop** (the set of propositions) to the two-element set  $\{T, F\}$ .

For  $f \in \mathcal{T}^{\mathbb{N}}$ ,  $V(f)$  is written  $V_f$ . In symbols we have

$$\begin{aligned} V : \mathcal{T}^{\mathbb{N}} &\longrightarrow \{g : \mathbf{Prop} \longrightarrow \{T, F\}\} \\ f &\longmapsto (\mathbf{Prop} \ni P \longmapsto V_f(P) \in \{T, F\}) \end{aligned}$$

To specify the notions of truth, validity, etc., we use the semantic details given in Figure 4.6. Notice the final two clauses in this semantic theory are actually clause-schemes, one instance for each  $n \in \mathbb{N}$ .

### 4.6.3 $\mathcal{L}^{\infty P}$ as a Solution

The observation to make here is exactly the same as the one we made when analyzing  $\mathcal{L}^P$  or  $\mathcal{L}^{2P}$  as solutions. Suppose the modal creativity of social constructivism allows for us to declare and collectively recognize that for absolutely any function  $f : \mathbb{N} \rightarrow \mathcal{T}$ ,  $\mathcal{R}, f \Vdash Q$ . Then, it seems the sentence

$$\begin{aligned} &\text{It is now the case that for all } f : \mathbb{N} \rightarrow \mathcal{T}, \mathcal{R}, f \Vdash Q, \text{ but it} \\ &\text{has not always been the case that for all such } f, \mathcal{R}, f \Vdash Q. \end{aligned} \tag{6}$$

demands we adopt yet another temporal perspective from which to analyze the semantics of Cole’s account. We would thus be forced to adopt the language  $\mathcal{L}^{(\omega+1)P}$ , whose semantic theory concerned valuations  $V : \omega + 1 \rightarrow \mathcal{T}$ . Again, mutatis mutandis, the above arguments can be given, and again Cole will be forced to abandon one of the core elements of his theory or ascend to a yet more complex temporal language in which to state it.

# Conclusion

A category need be seen as no more than a way of combining two objects to make a third that satisfies some axioms, just as a set need be seen as no more than a way of collecting or selecting things that satisfies some axioms. Ways of combining two things to make a third can, in turn, be perfectly well-understood as two-place functions. An abstraction principle defining objects that correspond exactly to equivalence classes of two-place functions that represent the same category can then be given.

The object that has garnered the most attention as a potential sticking point in discussion of category theory, however, is the *category of categories*. One might worry that, even if categories themselves are acceptable, that categories of categories embody too much complexity to be the type of thing that can be understood on their own. But just as we can provide an abstractionist theory of categories, we can provide an abstractionist theory of categories of categories. Further, one can in fact combine the insights of Chapters 1 and 3: a category of categories is just a way of combining two things to make a third in which one of the “combinees” behaves as **2** ought, and in which this combinee is a generator. The phrases “behaves as **2** ought” and “is a generator” in turn, can be encoded in pure logic, as shown in Chapter 3.

Thus, whether one thought the sticking point was with categories themselves or with categories of categories, all is well in the land of category theory. Of course, the last three-quarters of a century of mathematical practice has borne out much the same conclusion for the set-theoretic perspective. Category theory and set theory provide very different perspectives on mathematics. Both perspectives are perfectly acceptable ways to look at mathematics, so we are left with a pluralism of mathematical perspectives. This, in turn, seems to suggest that at least to some extent we can *choose* what the referents of our mathematical terms are.

As I pointed out in the introduction, this should not lead us to conclude the referents of our mathematical terms – the mathematicalalia, if you will – are themselves the kind of thing that depend on social action. The sort of referential pluralism just discussed *does in fact* leave the door open for a certain type of relativism about the nature of mathematical objects, but it's a relativism of a rather benign sort.

Nonetheless, despite the fact that social constructivism is not a *consequence of* foundational pluralism, it might tempt one to think favorably of social constructivism, and even to propose it as an ontological theory regarding mathematical entities. The problem, we have seen, is that social constructivism involves one in temporal anomalies of the most pernicious sort. The atemporality of mathematical truths is simply something that is too costly to our science for us to abandon, but it seems to drive the social constructivist to rather alarming conclusions – conclusions like “currently, mathematics has always been true, but it is not the case that mathematics has always had the feature of having always been true.” Making sense of such claims, in turn, demands rather nuanced semantic analysis and – even after this – cannot be maintained concurrently with endorsement of the way mathematical practice in fact treats mathematical truth.

Thus, despite the presence of category theory giving rise to a referential pluralism, we must not be tempted to adopt a social constructivism about mathematicalalia.

# References

- [1] Øystein Linnebo and Richard Pettigrew. Category theory as an autonomous foundation. *Philosophia Mathematica*, 2011.
- [2] F. William Lawvere. The category of categories as a foundation for mathematics. In *Proceedings of the conference on categorical algebra*, pages 1–20. Springer, 1966.
- [3] Gottlob Frege. *Basic Laws of Arithmetic*. Oxford University Press, Oxford, UK, 2013.
- [4] F. W. Lawvere. Functorial semantics of algebraic theories. *Proceedings of the National Academy of Sciences of the United States of America*, 50(5):869–872, Nov 1963.
- [5] Julian Cole. Creativity, Freedom, and Authority: A New Perspective on the Metaphysics of Mathematics. *Australasian Journal of Philosophy*, 87:589–608, 2009.
- [6] JM Dieterle. Social Construction in the Philosophy of Mathematics: A Critical Evaluation of Julian Cole’s Theory. *Philosophia Mathematica*, 18(3):311–328, 2010.
- [7] Julian Cole. Towards an Institutional Account of the Objectivity, Necessity, and Atemporality of Mathematics. *Philosophia Mathematica*, 21(1):9–36, 2013.
- [8] Solomon Feferman. *Categorical Foundations and Foundations of Category Theory*, pages 149–169. Logic, Foundations of Mathematics, and Computability Theory. Springer Netherlands, 1977.
- [9] Elaine Landry. The genetic versus the axiomatic method: Responding to Feferman 1977. *The Review of Symbolic Logic*, 6(01):24–51, 2013.

- [10] Geoffrey Hellman and John L. Bell. Pluralism and the foundations of mathematics. *Scientific Pluralism*, (19):64–79, 2006.
- [11] Geoffrey Hellman. Does category theory provide a framework for mathematical structuralism? *Philosophia Mathematica*, 11(2):129–157, 2003.
- [12] Geoffrey Hellman. *What is Categorical Structuralism?*, pages 151–161. The Age of Alternative Logics. Springer, 2006.
- [13] Steve Awodey. An answer to Hellman’s question: ‘Does category theory provide a framework for mathematical structuralism?’. *Philosophia Mathematica*, 12(1):54–64, 2004.
- [14] C. McLarty. Exploring categorical structuralism. *Philosophia Mathematica*, 12(1):37–53, 2004.
- [15] Eli Hirsch. *Quantifier Variance and Realism: Essays in Metaontology*. Oxford University Press, 2010.
- [16] Colin McLarty. Categorical foundations and mathematical practice. *Philosophia Mathematica*, 20(1):111–113, 2012.
- [17] Geoffrey Hellman. *Mathematics without Numbers: Towards a Modal-Structural Interpretation*. Clarendon Press, 1989.
- [18] Samuel Eilenberg and Saunders MacLane. General theory of natural equivalences. *Transactions of the American Mathematical Society*, 58(2):231–294, 1945.
- [19] Fraser MacBride. Speaking with shadows: A study of neologicism. *The British journal for the philosophy of science*, 54(1):103–163, 2003.
- [20] Bob Hale and Crispin Wright. *Implicit definition and the a priori*, pages 286–319. New essays on the a priori. Oxford: Oxford University Press, 2000.
- [21] Richard G. Heck. On the consistency of second-order contextual definitions. *Noûs*, pages 491–494, 1992.

- [22] Crispin Wright. *On the philosophical significance of Frege's theorem*, pages 201–244. *Language, thought, and logic, Essays in honour of Michael Dummett*. Oxford University Press, Oxford and New York, 1997.
- [23] Richard G. Heck. *Frege's Theorem*. Oxford University Press, New York, New York, 2011.
- [24] George Boolos. Frege's theorem and the Peano postulates. *The Bulletin of Symbolic Logic*, 1(3):317–326, Sep. 1995.
- [25] Roy T. Cook. Iteration one more time. *Notre Dame Journal of Formal Logic*, 44(2):63–92, 2003.
- [26] Kevin C. Klement. Neo-logicism and Russell's logicism. *Russell*, 32(127):159, 2012.
- [27] Stewart Shapiro. Frege meets Dedekind: A neologist treatment of real analysis. *Notre Dame Journal of Formal Logic*, 41(4):335–364, 2000.
- [28] Philip Ebert and Stewart Shapiro. The good, the bad and the ugly. *Synthese*, 170(3):415–441, 2009.
- [29] J. H. Conway, R. T. Curtis, S. P. Norton, R. A. Parker, R. A. Wilson, and J. G. Thackray. *Atlas of finite groups : maximal subgroups and ordinary characters for simple groups*. Clarendon, Oxford, 2009. ID: 849449912.
- [30] Kit Fine. *The limits of abstraction*. Oxford University Press, New York, New York, 2002.
- [31] Steve Awodey. An answer to Hellman's question: 'does category theory provide a framework for mathematical structuralism?'. *Philosophia Mathematica*, 12(1):54–64, 2004.
- [32] Colin McLarty. Axiomatizing a category of categories. *The Journal of symbolic logic*, 56(04):1243–1260, 1991.
- [33] John Rolfe Isbell. Review of Lawvere: 1966. *Mathematical Reviews*, 34(7332):1354–1355, 1967.



- [34] J. L. Bell. Category theory and the foundations of mathematics. *British Journal for the Philosophy of Science*, 32(4):349–358, 1981.
- [35] Solomon Feferman. Foundations of unlimited category theory: What remains to be done. *Review of Symbolic Logic*, 6(1):6–15, 2013.
- [36] John T. Baldwin. *How big should the monster model be?*, pages 31–50. Logic Without Borders. Walter de Gruyter, Berlin, Germany, 2015.
- [37] Shay Logan. Categories for the neologicist, 2015.
- [38] Marek A. Bednarczyk and Andrzej M. Borzyszkowski. Concrete (co) constructions in the category of small categories, 1998.
- [39] Shay Logan. Category theory is a contentful theory. *Philosophia Mathematica*, 23(1):110–115, 2015.
- [40] Richard Dedekind. *Essays on the theory of numbers*. Dover Publications, New York, 1963.
- [41] Richard Heck. *The Julius Caesar Objection*, pages 273–308. Language, Thought, and Logic: Essays in Honour of Michael Dummett. Oxford University Press, 1997.
- [42] Matthias Schirn. Fregean abstraction, referential indeterminacy and the logical foundations of arithmetic. *Erkenntnis*, 59(2):203–232, 2003.
- [43] Roy T. Cook and Philip A. Ebert. Abstraction and identity. *Dialectica*, 59(2):121–139, 2005.
- [44] Joongol Kim. A strengthening of the caesar problem. *Erkenntnis*, 75(1):123–136, 2011.
- [45] Hans Kamp. Tense logic and the theory of linear order. 1968.