

# **About What There Is**

## **An Introduction to Contemporary Metaphysics**

T. Roy  
Fall, 2009

## Contents

<i>Preface</i> .....	p. iii
I. Introduction to Metaphysics.....	p. 1
(1) An Overall Picture of the Metaphysical Project .....	p. 2
II. Quine's Method for Metaphysics.....	p. 24
(2) Plato's Beard: A Problem About Method .....	p. 25
(3) Russell's Way Out: A Plausible Canonical Notation.....	p. 44
(4) Quine's First Thesis: How the Notation Commits.....	p. 73
(5) Quine's Second Thesis: Sufficiency and Application.....	p. 96
III. Realism and Truth.....	p. 118
(6) Putnam's Anti-Realism.....	p. 119
IV. Problems of Metaphysics	
tbd	
Appendix: Validity and Soundness.....	p. 144
<i>Further Reading</i> .....	p. 158
<i>Assignment Schedule</i> .....	p. 161

## Preface

Strangely, many metaphysics texts do not really introduce what contemporary metaphysicians do. A result is that huge chunks of contemporary philosophy may seem mysterious or bizarre, even to students with an undergraduate degree in philosophy. This might happen for different reasons.

First, it is my conviction that a text for upper-level philosophy is properly justified as a “window” into standard original works. A text should not replace original works, but rather provide a “pathway” through them. An upper-level text should supply background so that original works are accessible, provide context to integrate them, and enter into a student’s conversation with the original works themselves. In this latter role, a textbook might itself count as original. I think many texts for upper-level metaphysics do not adequately serve these ends. Thus, e.g., many do not make direct contact with the literature. One might think that, even so, a text could count as a legitimate introduction to the subject matter. Unfortunately, many metaphysics texts do not do even this.

A metaphysics course that takes up questions about god, mind and body, freedom and determinism, space and time, etc. takes up interesting and important metaphysical questions. But such a course may overlap with standard introductions to philosophy, and with more advanced courses in philosophy of religion, philosophy of mind, philosophy of science, etc. Thus, contrary to fact, metaphysics may seem to have no subject matter of its own. And such a course may never raise those questions about reality, truth, abstract objects, events, and the like, which dominate so much of contemporary (specifically) metaphysical discussion. Even worse, a course devoted to questions about god, mind and body, freedom and determinism, etc. may leave the more specifically metaphysical questions strange and unmotivated.

My aim is to remedy this situation. I think a narrow focus on metaphysical method, and some specifically metaphysical questions, not only introduces metaphysics proper, but also illuminates metaphysical discussion more generally, and even philosophical discussion beyond the borders of metaphysics. I approach the task in four sections. The short first section develops an overall picture of the metaphysical project. The second section takes up metaphysical method and especially W. V. O. Quine’s classic article, “On What There Is.”<sup>1</sup> These first two sections raise many important and interesting metaphysical questions. The latter two focus on a few metaphysical questions more directly. The third section stands between the second and the fourth, insofar as it takes up metaphysical issues which matter for understanding the metaphysical method. The last section takes up some metaphysical problems more for their own sake. In this

---

<sup>1</sup> Quine, “On What There Is,” *The Review of Metaphysics* 2 (1948): 21-28. Reprint in many places including Quine, *From a Logical Point of View* 2nd ed. (Cambridge: Harvard University Press, 1980).

section, I offer a perspective on what there is which, I hope, is original and interesting in its own right. I make no claim to comprehensiveness or breadth. Thinking about these few questions should put us in a position to take up questions beyond those directly addressed. And *this*, I think, is what an introduction to metaphysics should do.

This text aims both high and low. On the one hand, it's a significant task to make contact with contemporary metaphysics (e.g., Quine's slogan, "to be is to be the value of a variable" as discussed in chapter 4, and the consequences of extensionality as discussed in chapter 5). So working through these issues requires a certain degree of philosophical sophistication. On the other hand, no particular understanding, beyond a familiarity with validity and soundness that might be obtained from an introductory course on critical thinking, is assumed (and even those notions are discussed in an appendix). It is likely that the reader will benefit from background in formal logic or philosophy of language. But every effort is made to supply whatever particular content is required. Philosophical background, especially in logic and philosophy of language, should ease the way into this text. But, correspondingly, working through this text should ease the way into logic and philosophy of language. So it's not obvious that one order is better than another. If there were a *standard* order, later courses could presuppose content from earlier ones. But there is no such order, and none is presupposed.

I have, in the past, organized metaphysics courses around Keith Campbell's text, *Metaphysics: An Introduction* and, in broad outline, this book reflects the first and third sections of his.<sup>2</sup> Naturally, I draw on many different sources; I am especially indebted to my teacher, Michael Jubien. Unfortunately, *About What There Is* is not yet complete. Thanks to comments and discussion from my colleague Matthew Davidson, along with students, especially Richard Jensen, Meggan Coté, Dan Bridges, Donovan Rinker, Robb Vitt, and Sean Korb, in past versions of PHI 380 at CSUSB, it is better than it was. Your sufferings should make it better still. Perhaps it seems unfair to have to work through a text in this state. It shouldn't. My upper-division course on metaphysics has always taken up topics in metaphysics proper. The aim of this text is to aid this project. Insofar as Campbell's text (and others) remain available, what you have can only make things better. Note that page references in the text prefaced with a lowercase 'e' are to the essential readings reprinted and separately bound.

I find the material to be fascinating. I very much hope that you will find it to be so, as well!

T.R.

---

<sup>2</sup> Campbell, *Metaphysics: An Introduction*, (Encino: Dickenson Publishing Company, 1976).

# Part I

## Introduction to Metaphysics

## Chapter One

### An Overall Picture of the Metaphysical Project

Metaphysics is the study of certain very general and basic questions about what the world is like. Central to contemporary metaphysics are questions about *what there is*. These will be our focus. In their contemporary dress, such metaphysical questions arise when one tries to “get behind” ordinary claims and theories and ask what does, or could, make them true. The aim of this chapter is to motivate that project, and to evoke an overall impression of it. I begin with a general characterization, move to an extended example, and make a few applications from the case.

#### *I. General Characterization*

Metaphysics is, and is not, like ordinary science. Let’s assume that there is an “external” world, and that ordinary science uses experiment to develop more-or-less well-supported theories about what it is like. Then metaphysics is like science insofar as it involves theorizing about the world. But metaphysicians do not typically perform experiments on the world. So metaphysics is unlike science regarding the data to which its theories are responsible. Let’s take up each point in turn.

(A) I suggest that contemporary metaphysics is like science insofar as it involves theorizing about the world. It has not always been so—even today, not all philosophers would agree. We can appreciate what is at issue when we see something of the alternatives. It is possible to differ both with respect to the *method* and the *object* of metaphysics. Some philosophers see metaphysics not so much as *theorizing* about the world, as *demonstrating* sure and indubitable truths about it. And some philosophers see metaphysics not so much as an investigation of the *world*—conceived as an object which exists independently of us, as an investigation of our *experience* of it, or of the way it *appears* to us. I don’t intend to argue about what metaphysics is or should be. The suggestion that metaphysics involves theorizing about the world is offered as a theory (!) about what happens when metaphysical questions arise. As such, it is to be evaluated in the context of responses to metaphysical questions and, in particular, may be evaluated relative to the extended example that follows. If necessary, we may think of “contemporary metaphysics” as one among many approaches to the metaphysical questions. Even so, it may help to clarify what is at issue, if I say a bit more about the alternatives.

First, there is a “classical” tradition, stretching from Plato in the fifth century BC at least through Descartes, Spinoza, and Leibniz in the seventeenth century, on which metaphysicians do not *theorize* about the world, but use reason, apart from observation and experiment, to *demonstrate* truths about it. On this view, metaphysics is immune from the error and approximation associated with ordinary science. As it turns out, philosophers like Descartes, Spinoza, and Leibniz don’t agree about the supposedly sure

results of pure reason, and there are reasons to doubt the conclusions of each. Perhaps, then, contemporary metaphysicians are simply *discouraged* about the prospects for a metaphysics of this sort. Naturally, lack of agreement does nothing to *show* that the classical project is misconceived. For all I have said, Leibniz, say, may be right. Evaluating the work of a particular philosopher requires particular and detailed argument. But there is a general theoretical worry as well: Insofar as reason has no input from the world, it is not clear how or why it is responsible to the world—and especially how or why it is able to determine *what there is*. Suppose, e.g., I form the concept of a “hingledopper”—where part of what it is to be a hingledopper is to be big and slow. It follows, from reason alone, that all hingledoppers are big and slow. Perhaps I enjoy thinking about hingledoppers, am afraid of meeting one, and so forth. But can I know whether there are any? For this, observation seems required. And, similarly, it may be thought that “pure” reason is inevitably limited to its own objects and cannot break through to conclusions about what exists in the outside world. For this, observation or experiment seem required and, if they are required, we are back to the vagaries of ordinary science.<sup>1</sup> As above, even if this negative conclusion is denied, there plausibly remains a place for observation and theory; and we might thus see contemporary metaphysicians as having simply abandoned the classical project in favor of a different one—we may see contemporary metaphysicians as having adopted one mode, among others, of approaching the questions of metaphysics.

Second, as one may think that reason apart from observation and experience cannot “break through” to conclusions about the world, so one may think that reason *with* observation and experience is incapable of reaching beyond experience to legitimate conclusions about the “external” world. Legitimate conclusions are limited to experience itself. This negative judgment has seemed to some, including ancient skeptics and twentieth century “logical empiricists,” a reason to reject all metaphysics as absurd. But there is another tradition, stretching from Immanuel Kant in the eighteenth century through modern-day “continental philosophy” and “anti-realism” which, given this “lemon,” makes lemonade. On their view, the only “world” that matters is the world of experience—and we should be happy to see metaphysics as having *it* as its object. This tradition is hardly monolithic. Kant thinks of metaphysics as demonstrating truths about the world of experience. Modern-day anti-realists seem to think more in terms of theorizing about it. We will think carefully about anti-realism in part III. For now, observe that there is a general theoretical difficulty here as well: It is very difficult to make sense of “experience” without an experiencer, and of “appearance” without something that appears. That is, these views “cry out” for at least some theorizing about

---

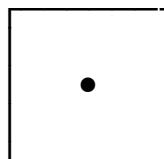
<sup>1</sup> Students familiar with Hume will recognize this as a variant of his point about matters of fact and relations of ideas. Thus *An Enquiry Concerning Human Understanding* concludes, “If we take in our hand any volume; of divinity or school metaphysics, for instance; let us ask, *Does it contain any abstract reasoning concerning quantity or number?* No. *Does it contain any experimental reasoning concerning matter of fact and existence?* No. Commit it then to the flames: For it can contain nothing but sophistry and illusion” p. 114 in the Steinberg edition (Indianapolis: Hackett Publishing Company, 1977). Hume allows that one might sensibly investigate relations among ideas, or investigate matters of fact, but resists using considerations of one sort to reach conclusions about the other.

the “external” world—for something of what I have called, “contemporary metaphysics.” The anti-realist thinks theorizing about the external world has problems of its own. We will get to these in due time. I propose, however, that we at least begin with the assumption that our theorizing has as its object an independently existing “external” world. Surely this is the consensus of common sense and ordinary science and, as we shall see, it is natural for our metaphysics to begin there. If it should turn out that the assumption is somehow misguided, the results of our study will not be void—though they will require reinterpretation!

(B) Say metaphysics is like science insofar as it involves theorizing about the world. This leaves open what the questions metaphysics tries to answer are, and what the data to which metaphysical theories are responsible is. I suggest that, in their contemporary dress, metaphysical questions arise when one tries to “get behind” ordinary claims and theories and ask what does, or could, make them true. Correspondingly, ordinary claims and theories themselves count as part of the “data” to which metaphysical theories are responsible. Let me explain.

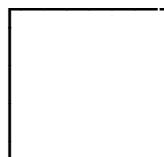
It’s natural to think that *the world* makes a statement true or false. Or, better, given that the *meaning* of some statement is fixed, its truth and falsity is a matter of how things are. So, e.g., consider the following statement,

There is a dot in this box:



And compare it with,

There is a dot in this box:



The first says that there is a dot in the upper box, and the second that there is a dot in the lower box. The first is true and the second is false. What is the difference? Intuitively, the statements say something about the way things are, and what they say is true if and only if (iff) things are, in fact, that way. The first is true because things are the way it represents them to be, and the second is false because things are not the way it represents them to be. What the first says *corresponds* to reality, but what the second says does not; so the first is true and the second is not. Similarly, “The earth is round” is true because things are the way the statement represents them to be, and “The earth is flat” is false



because things aren't the way it represents them to be.<sup>2</sup> And, more generally, it is natural to think that an arbitrary statement is *true* if and only if reality is as the statement represents it to be—if and only if what it says corresponds to reality—and *false* if and only if it does not.

Very generally, then, my suggestion is that we get to the questions of metaphysics when we consider various statements and ask *what it is* that makes them true or false. We will sharpen the issue if we get a bit more specific about truth. Restrict attention to statements in the simple subject-predicate form, e.g.,

- (1) Bill is happy.

We'll see (1) as consisting of a *subject* term, 'Bill', and a *predicate* term, 'is happy'. The subject term picks out an object, and the predicate term says something about it. In English, proper names ('Bill', 'Mt. Whitney'), demonstratives ('this', 'that'), pronouns ('it', 'he'), and definite descriptions ('the tallest woman', 'the president') are used to pick out objects and so may serve as subject terms. Verbs and verb phrases are predicate terms. Notice: 'the tallest woman' may serve to pick out an object, as 'the tallest woman is bald', but it may also be used to say something about an object, as 'Hillary is the tallest woman.' The point of the subject/predicate distinction is not about the particular form of the words used, but rather about the *role* they play. As we are understanding subject/predicate statements, there is a single subject term which picks out or *refers* to a thing, and a single predicate term which *applies* to a thing or not, depending on how the thing is. In the one case, then, 'the tallest woman' is a subject term, and in the other case, it is part of the predicate. On this basis, the following may seem obvious.

- T1** A subject-predicate statement is *true* if and only if the subject term refers to some object, and the predicate term applies to the object to which the subject term refers.

Suppose **T1** is right—that it accounts for the truth of arbitrary (simple) subject-predicate statements. Its application to (1) is straightforward. The subject term, 'Bill' picks out a person, in this case one who used to be the president of the United States, and the statement is true just in case the predicate term applies to him—just in case that person happens to be happy. But consider,

- (2) Seven is a prime.  
 (3) Frodo is a hobbit.  
 (4) *Fanfare for the Common Man* is brassy.

---

<sup>2</sup> Of course, if I am grossly deceived about the shape of the earth, then my evaluations of truth and falsity are mistaken. But this does not alter the basic point about *what makes* the statements true and false.

Each is true.<sup>3</sup> In fact, their truth is non-negotiable—or, at least, the truth of (2), (3) and (4) is more certain than that of typical metaphysical theories. From **T1**, then, it follows that there exist objects corresponding to their subject terms. But what objects? Focus, for the moment, on (3). Since (3) is true, **T1** tells us that there is an object corresponding to ‘Frodo’, and that ‘is a hobbit’ applies to it. But what is this object, and where is it? Tolkien enthusiasts may insist that Middle Earth exists in some serious sense, and that Frodo is a (fairly heroic) hobbit living there. But this will strike most of us as a non-starter—it is too bizarre to be believed. So maybe Frodo is some more ordinary object. Presumably, though, no ordinary concrete object is a hobbit. One might hold that Frodo is an idea. But no idea is a hobbit: Hobbits have hairy feet, and ideas have neither feet nor hair! Perhaps Frodo is marks on paper or, now, on film. But, again, *hobbits* are animals—where marks on paper and film are not. So Frodo cannot simply *be* marks. Etc. There remain the options of rejecting **T1**, and of denying that (3) is true. But we want to hold with common sense, and at least to resist denying (3). And rejecting **T1** leaves us with the question of what in the world makes (3) true—since, if it is true, *something* in the world makes it true. Answering this question, and others like it, forces us into real dilemmas about what there is—and so into the heart of contemporary metaphysics.

Let’s recap. We routinely talk as if there are numbers, fictional characters, musical works and the like. We even say things like, ‘*There is a prime number greater than five and less than nine*’ which, on the surface at least, seems a direct assertion of existence. That such statements are true is not seriously to be doubted, and therefore counts as a sort of data or starting point for metaphysical discussion. Metaphysicians offer more-or-less plausible theories—theories about what there is—to account for this truth. Of course, a theory that denies that there is some magical realm of numbers or fictional objects is itself a metaphysical theory. But any such theory will be judged relative to the data.

Perhaps some will respond that this is a fascinating sort of question. But others may think that I am pulling the wool over their eyes, and that the supposed difficulties are entirely superficial—so that the questions are uninteresting. To undercut this response, and to lay a foundation for further general comments on the nature of metaphysics, let’s turn to an extended example.

## *II. The Case of Mathematical Truth*

Most of us accept a great many things about arithmetic and the natural numbers (the natural numbers are 0, 1, 2, 3...). We accept that there are truths of arithmetic, and that we know some of them. It is clear enough that  $3 + 2 = 5$ , but less clear whether every even number greater than two is the sum of two primes. Still, either every even number

---

<sup>3</sup> Just in case: Frodo is a character in J.R.R. Tolkien’s *The Lord of the Rings* (my copy is New York: Ballantine Books, 1965; *The Lord of the Rings* was complete in 1956). *Fanfare for the Common Man* is a musical work by Aaron Copland.

greater than two is the sum of two primes or not.<sup>4</sup> We accept that there is no end to the natural numbers—for every natural number, there is one greater than it. And the truths of arithmetic are, in some sense, “eternal” and “necessary.” If  $3 + 2 = 5$ , then  $3 + 2$  has always been 5 and, in some sense, has to be 5. We might imagine a situation where the words ‘three plus two equals five’ mean, “all dogs can fly”—and so say something false—but this isn’t the same as imagining a situation where *three* plus *two* isn’t *five*. Also, truths about numbers have application to the world. If  $3 + 2 = 5$ , then the result of taking three things and two things is five things.

In this section, I sketch four attempts to account for this data.<sup>5</sup> Though each has its modern-day supporters, each is problematic in one way or another. Naturally, these are not the only approaches to mathematical truth. And I do no more than sketch the four views along with some worries. So the discussion is hardly complete. It is, in fact, barely an introduction to the topic. But our aim, for now, is merely to see something of how the discussion might go, and something of its difficulty.

(A) *Platonism* or *realism* is a theory according to which numbers exist as such. On this view, numbers are as real as rocks and trees and, like rocks and trees, are what they are no matter what anyone thinks or says about them. However, on traditional versions of this view, numbers do not exist at any particular place or time. Where a *concrete* object is one that exists at some place and time, an *abstract* object is one that exists but not at any place or time. One might wonder whether there are any abstract objects. On this view, however, there are, and numbers are among them. Again, on a Platonic view, numbers do not therefore sacrifice independent being or existence. Plato seems to think of them as existing in a sort of “heaven.” Along with The Good, Justice, Beauty, and the rest of the “forms,” Numbers are part of a “pure” reality which our world “reflects” in an imperfect way.<sup>6</sup>

The realist can accept a perfectly straightforward account of mathematical truth. “There is a prime number greater than five and less than nine” is true precisely because *there is* a prime number greater than five and less than nine. She may accept something like **T1**, and hold that ‘seven’ refers to a number so that ‘seven is a prime’ is true because ‘is a prime’ applies to the object to which ‘seven’ refers. Insofar as arithmetical truth is a

---

<sup>4</sup> This is the famous Goldbach conjecture which, so far as I know, has resisted all attempts at proof. So, e.g.,  $2 + 2 = 4$ ,  $3 + 3 = 6$ ,  $3 + 5 = 8$ ,  $5 + 5 = 10$ ,  $5 + 7 = 12$ , etc. Nobody has ever found an even number greater than two that is not the sum of two primes. But this does not show that there isn’t one!

<sup>5</sup> The discussion of the first three draws especially on chapter 2 of M. Jubien, *Contemporary Metaphysics* (Malden: Blackwell Publishers, 1997).

<sup>6</sup> For Plato, see, e.g., *Phaedo* 73-76, 100-101, and *Republic* 514-526, in, e.g., their respective translations by Tredennick and Shorey, in *The Collected Dialogues of Plato*, ed. Hamilton and Cairns (Princeton: Princeton University Press, 1961). Gottlob Frege’s 1884 *The Foundations of Arithmetic*, trans. Austin (Evanston: Northwestern University Press, 1959), and Penelope Maddy, *Realism in Mathematics* (Oxford: Clarendon Press, 1990), are more recent versions of the view.

matter of correspondence, either every even number is the sum of two primes or not—no matter what anyone thinks or knows about it. Since all the numbers exist, there are infinitely many of them. And, insofar as the “heaven” is eternal and unchangeable, there is no problem about the eternity and necessity of arithmetical truths.

But there are problems as well. I’ll mention two. First, there is a problem about knowledge. We know that  $3 + 2 = 5$ . But on what ground do we claim that there are abstract objects at all, and how do we find out about particular features of 3, 2 and 5 so as to determine that  $3 + 2 = 5$ ? One might put the worry like this:

- (a) If Platonism is true, then numbers are outside of space and time.
- (b) If some things are outside of space and time, then there is no way for us to interact with them.
- \_\_\_\_\_
- (c) If Platonism is true, then there is no way for us to interact with numbers.
- (d) If there is no way for us to interact with some things, then we cannot have knowledge about them.
- \_\_\_\_\_
- (e) If Platonism is true, then we cannot have knowledge about numbers.
- (f) We can have knowledge about numbers.
- \_\_\_\_\_
- (g) Platonism isn’t true.

(c) follows from (a) and (b); (e) follows from (c) and (d); (g) follows from (e) and (f). (f) is just our datum that mathematical knowledge is possible; (a) states the Platonic view. So the argument hinges on (b) and (d). Insofar as (b) and (d) are plausible, Platonism has trouble with mathematical knowledge.<sup>7</sup> At one time, at least, Plato held a doctrine according to which we *do* interact with numbers: in this life we recollect a sort of “communion” with them from before we were born. This doctrine has problems of its own, and other philosophers suggest other solutions.

But, second, the problem about knowledge isn’t *merely* that numbers are difficult to investigate. Many people find reasons to think that there *aren’t* any. Consider what happens when a child concludes that there is no Santa Claus. Perhaps she discovers presents marked “from Santa” in her parent’s bedroom closet, perhaps she sees parents filling her stocking, etc. That is, there may be problems about evidence for Santa—evidence which once seemed to show that Santa exists falls away. So far, this is like the problem discussed above. But, further, the child may worry about how reindeer fly, how Santa visits so many houses, etc. That is, there may be problems about the *nature* of Santa Claus—the child has beliefs about the nature of the world, and Santa doesn’t fit in. The child forms a successful theory of the world on which reindeer don’t fly, there isn’t

---

<sup>7</sup> Much modern discussion of this problem derives from Paul Benacerraf, “Mathematical Truth,” *Journal of Philosophy* 70 (1973): 661-79. Reprint in many places including: *The Philosophy of Mathematics*, ed. W. D. Hart (Oxford: Oxford University Press, 1996).

time in one night to visit all the houses, etc. And from this theory it follows that there is no Santa. So the child reasonably concludes that Santa does not exist. But similarly for abstract objects generally, and numbers in particular. Contemporary science may seem to be a powerful and successful theory admitting only things located at places and times. But, if this is right, there are no abstract objects. So it seems rational to conclude that Platonism is false, and natural to seek an alternative. As we shall see, however, Platonism isn't all that easy to avoid.

(B) Like Platonism, *conceptualism* is a view according to which numbers exist. However, on this view, numbers are not abstract. Rather, they are to be identified with certain concepts or ideas in minds. Given this, mathematical truth arises in the usual way: The conceptualist accepts **T1**, and holds that 'seven is a prime' is true because 'is a prime' applies to the idea to which 'seven' refers.<sup>8</sup> In this case, it is important to distinguish an idea *of* a thing, from the thing of which it is an idea. So, e.g., my idea of the Parthenon isn't the Parthenon. But a person may have an idea of a mental entity, and the conceptualist might hold that seven, e.g., is some mental entity of which he has an idea. To make the view concrete, let's simply suppose that numbers can be identified with brain states. Then we avoid at least the objection, raised against Platonism, that numbers don't fit into an ordinary picture of what the world is like. And, given the proximity to minds, there should be no problem about mathematical knowledge.

But there are problems. Again, I'll mention two. First, the transient and fluctuating character of mental states seems inconsistent with the eternal and necessary character of mathematical truth. So, e.g., was  $3 + 2 = 5$  before there were any people? Will  $3 + 2 = 5$  after people are gone? Surely  $3 + 2$  has always been 5! However, if '3', '2', and '5' refer only to *mental states* then, by something like **T1**, ' $3 + 2 = 5$ ' *isn't* true without the mental states—when there is nothing to which '3', '2' and '5' refer. Even worse, ideas are capable of variation, and this suggests that mathematical truths might themselves change. As Frege puts it, if numbers are psychological entities,

Astronomers would hesitate to draw any conclusions about the distant past, for fear of being charged with anachronism—with reckoning twice two as four regardless of the fact that our idea of number is a product of evolution and has a history behind it.... How could they profess to know that the proposition  $2 \times 2 = 4$  already held good in that remote epoch? Might not the creatures then extant have held the proposition  $2 \times 2 = 5$ , from which the proposition  $2 \times 2 = 4$  was only evolved later

---

<sup>8</sup> In II.xvi-xvii of his 1690 *An essay Concerning Human Understanding* in, e.g., the Fraser edition, (New York: Dover Publications, 1959), Locke develops a perspective of this sort. One might see modern day "intuitionism," e.g., M. Dummett, "The Philosophical Basis of Intuitionist Logic," reprint in *Philosophy of Mathematics: Selected Readings* 2nd ed., ed. Benacerraf and Putnam (Cambridge: Cambridge University Press, 1983), and *Elements of Intuitionism* (Oxford: Oxford University Press, 1977), as a descendent of this view.

through a process of natural selection in the struggle for existence? Why, it might even be that  $2 \times 2 = 4$  itself is destined in the same way to develop into  $2 \times 2 = 3!$ <sup>9</sup>

Sarcastic, but the point is clear. These difficulties are related to the transient and variable character of mental states. So one might respond by holding that there is some one idea—an idea that is not itself any particular mental state—of say, seven, that everyone *has* or *grasps*. But, if the idea isn't a mental state, what is it? If the idea is eternal and unchanging, it might start to seem like one of Plato's numbers! It's possible to appeal to ideas in the mind of God. But this also seems a (relatively) Platonic proposal. We gain eternity and necessity, at the cost of problems about knowledge and the nature of numbers. That is, these proposals seem to give away what conceptualism was supposed to gain.

Further, there is a problem about infinity. Part of our data is that there are infinitely many numbers. But there aren't infinitely many ideas. A 4-bit binary memory location can take 16 different states,

0 0 0 0	0 1 0 0	1 0 0 0	1 1 0 0
0 0 0 1	0 1 0 1	1 0 0 1	1 1 0 1
0 0 1 0	0 1 1 0	1 0 1 0	1 1 1 0
0 0 1 1	0 1 1 1	1 0 1 1	1 1 1 1

And, similarly, a computer with any finite amount of memory can take at most a finite number of states. Presumably, something similar is true of the brain. Without suggesting that brain states are discrete or that the brain is digital, it is plausible to suppose that a person can have only so many ideas in a lifetime. To pick a number, suppose persons are capable of just  $10^{100}$  ideas and, what seems unlikely, that each person has as many ideas as they can. Now suppose the total number of persons who will ever live is  $10^{25}$ . (Obviously, I'm picking numbers out of the hat.) Then the total number of ideas is  $10^{100} \times 10^{25} = 10^{125}$ . Supposing—what seems improbable (to say the least)—that each of these ideas is a different number less than or equal to  $10^{125}$ , the conceptualist can rest content about the existence of numbers less than or equal to  $10^{125}$ . But what about  $10^{125} + 1$ ? There is no idea for it, so it doesn't exist. Notice: There is no problem about having an idea of *there being*—according to which there are—infinitely many numbers. But this *isn't the same as having infinitely many ideas*. For the conceptualist, numbers are supposed to be ideas. So, for the conceptualist, there aren't infinitely many numbers.

One might protest that the infinity of the numbers isn't supposed to be an *actual* infinity, but is rather a merely *potential* one. It's not that there *are* infinitely many numbers, but only that for any number there *can* be one greater than it. For any number idea I have, it seems possible for me to have another. So, if the infinity of the numbers is only potential, ideas seem well-positioned (in this respect) to count as numbers. In response, I'll only comment that this proposal does not conform to mathematical practice.

---

<sup>9</sup> *The Foundations of Arithmetic*, vi-vii.

Mathematicians routinely deal with “transfinite” (that is, infinite) numbers and the like. On the face of it, this activity is subject to *mathematical* criticism, but isn’t the *sort of thing* that could be brought down by considerations about the number of ideas that happen to obtain. And, further, it’s not easy to relegate this mathematical activity to the realm of meaningless mathematical games. So, e.g., physics depends crucially on the calculus, and the calculus on limits for functions of real numbers. To ground the account of such limits, mathematicians have understood real numbers as themselves sets containing infinitely many natural numbers. This procedure requires the actual, rather than the potential, existence of infinitely many natural numbers. The details are beyond the scope of our discussion. But they aren’t necessary.<sup>10</sup> The point should be clear: A move to the potential, as opposed to the actual infinity of the numbers jettisons an important part of mathematics, and so an important part of the data. The suggestion that there are infinitely many (extra-mental) ideas which people may have or grasp might fix the problem. But, again, this suggestion returns us in the direction of Platonism. And similarly for the proposal that numbers are ideas in the mind of God.

(C) Taking three things and two things results in five things—no matter what people think or have thought. This motivates the first objection against conceptualism. But maybe that is all we need! *Nominalism* is a view according to which, literally speaking, there are no numbers. Rather, statements apparently about numbers are reinterpreted so that they are about something else. Maybe statements, seemingly about numbers, are literally about no more than the “taking” of things. Then for, say, ‘seven is prime’, we can deny **T1**—we can deny that ‘seven’ refers to any particular thing—and give an alternate account of truth.<sup>11</sup> Here is one way to proceed: Sometimes ‘seven’ seems to work like a name, but sometimes it doesn’t. Consider, e.g.,

(5) There are seven dwarfs in the forest.

In this case, ‘seven’ seems to indicate a property of the group of dwarfs. The group is seven-membered. The nominalist’s idea is to take this usage as primary, and the “name” usage as derived. On this view, the property of being seven-membered does not involve any object that *is* seven. And statements which apparently involve numbers, are to be understood as short for relatively complex claims about the nature of groups.

We need some detail about what these complex claims are. If talk about numbers is short for talk about properties of groups, we need to explain the group properties without appeal to numbers. So, e.g., for a two-membered group, we might try,

---

<sup>10</sup> Those with the requisite background might look at chapters 17, 40, 41 and 51 of M. Kline, *Mathematical Thought from Ancient to Modern Times* (New York: Oxford University Press, 1972).

<sup>11</sup> A nominalistic approach is advocated in J. S. Mill’s *A System of Logic* (New York: Longmans, Green and Co., 1936) and, more effectively, in N. Goodman, *The Structure of Appearance*, 2nd ed. (Indianapolis: Bobbs-Merrill, 1966), and H. Field, *Science Without Numbers* (Princeton: Princeton University Press, 1980).

- (6) A group is *two-membered* if and only if it has members  $a$  and  $b$  such that  $a \neq b$ , and it has no other members.

(6) is clumsy, but it says what it is for a group to be two-membered without appeal to numbers. According to (6), “This group is two-membered” is no more than a (welcome) abbreviation for “This group has members  $a$  and  $b$  such that  $a \neq b$ , and it has no other members.” Similarly, we might say,

- (7) A group is *three-membered* iff it has members  $a$ ,  $b$  and  $c$  such that  $a \neq b$ ,  $a \neq c$ ,  $b \neq c$  and it has no other members.

And, in the same way, one might deal with (5). Suppose this is done and, for any natural number  $n$ , suppose that claims about  $n$ -membered groups have an expansion into claims without appeal to numbers. Given this, we are in a position to deal with other mathematical statements. So, e.g., maybe,

- (8)  $3 + 2 = 5$  iff given an arbitrary three-membered group  $a$ , and two-membered group  $b$ , with no members in common, the group consisting of all the members of  $a$  and all the members of  $b$  is a five-membered group.

So ‘ $3 + 2 = 5$ ’ is short for the relatively complex condition mentioned in (7); and, of course, the condition in (7) is itself short for a condition with ‘three-membered group’, ‘two-membered group’ and ‘five-membered group’ replaced by expanded forms. So far, perhaps, so good. We have the truth of ‘ $3 + 2 = 5$ ’—and, presumably, it has always been the case that  $3 + 2 = 5$ . Further, insofar as we can observe groups, there might be no problem about knowledge. And it is clear why we ordinarily revert to the language of arithmetic (despite its misleading suggestions about existence)—for we want to avoid the complexity associated with the cold, literal, truth.

But, again, there are reasons to worry. An initial problem has to do with the nature of a *group*. We have understood (5) as, in effect,

- (5') The group of dwarfs in the forest is seven-membered.

where this has an appropriate expansion. In its expanded form, (5') is supposed to be literally true. But (5') is in the subject-predicate form; so it is natural to think that **T1** should apply. That is, seemingly, the truth of (5')—and of (5)—requires the existence of a group to which ‘the group of dwarfs in the forest’ refers and to which (the expanded version of) ‘is seven-membered’ applies. But *what is* a group? Is the group an eighth thing in addition to the seven dwarfs? Ordinarily, we recognize that there are concrete things like dwarfs, rocks and trees. But which of these is the group of dwarfs? If it should turn out that the group isn't a concrete thing, we would seem to have returned to something like Platonism.



But perhaps there is a way out. Nelson Goodman suggests that there are complex and disjoint concrete objects: for any  $n$  things there is a thing that has just those  $n$  things as its parts. So, e.g., there are Sleepy, and Dopey, but there is also Sleepy-Dopey. The former are dwarfs, and the latter is not—as an arm may be part of a person but isn’t itself a person, so Sleepy and Dopey are parts of a complex and disjoint thing that isn’t itself a dwarf. On this view, given ordinary things  $b$ ,  $c$  and  $d$ , there are  $b$ ,  $c$ ,  $d$ , but also,  $(b + c)$ ,  $(b + d)$ ,  $(c + d)$  and  $(b + c + d)$ . But there is just one thing  $a$  whose parts are  $b$ ,  $c$ , and  $d$ —that is,  $a = b + c + d = b + (c + d) = c + (b + d) = d + (b + c)$ ; in each case, the combination of parts results in the same complex thing. So for any  $n$  things, there is just one thing that has those  $n$  things as parts. On this view, not all objects are “ordinary,” but at least all are concrete. Given Goodman’s approach to groups, ‘ $x$  is a member of group  $y$ ’ translates into something like, ‘ $x$  is a part of complex thing  $y$ ’. So *the dwarfs* is a complex thing—not itself a dwarf—and (5) is true iff the dwarfs happens to have seven parts that are, individually, dwarfs. (Of course, ‘seven’ is to be eliminated as above.) So the objection about the existence of groups seems put to rest.

But maybe not. Parallel to conceptualism, there is a problem about infinity. On this nominalistic account, it is the *complex things* that make mathematical claims true or false. Thus, ‘for any number there is one greater than it’ translates into something like ‘given an arbitrary thing  $a$ , there is a thing  $b$  with more parts than  $a$ ’. But, if there are only finitely many things, this latter claim is false: the thing with all other things as parts has more parts than any other. So, apparently, the nominalist must deny an important datum about mathematical truth. Even if there are as a matter of fact infinitely many concrete things, a problem remains if there *could have been* only finitely many things. For then the nominalist must allow that ‘for any number there is one greater than it’ *could* have been false. And this conflicts with the necessity of mathematical truth. One might protest that mathematics isn’t about actual groups, but about *ways groups could be* and, since there *could* be infinitely large groups, there is no problem about the infinity of the numbers. But this leaves us with a difficulty about the nature of these “ways.” If ways aren’t concrete, we seem cast into a form of Platonism.

Here’s another response. Suppose we reject Goodman’s account of groups, and allow that a group is something different from just the sum of its parts. The same individuals may “come together” to form different groups. Using curly brackets, ‘{ }’ to represent group membership,  $\{b, c, d\} \neq \{b, \{c, d\}\}$ ; that is, the group whose members are  $b$ ,  $c$  and  $d$  is not the same as the group whose members are  $b$  and the group whose members are  $c$  and  $d$ —the one has three members and the other only two! Similarly, on this view,  $\text{Bill} \neq \{\text{Bill}\}$ ; the group with Bill as a member is one thing, and the sum of its members—that is, Bill—is another. Further, if groups are distinct from the sum of their members,  $\{\{\text{Bill}\}\}$ —the group whose only member is the group whose only member is Bill, is another thing still. And similarly,

Bill,  $\{\text{Bill}\}$ ,  $\{\{\text{Bill}\}\}$ ,  $\{\{\{\text{Bill}\}\}\}$ ,  $\{\{\{\{\text{Bill}\}\}\}\}$ , etc.

are all different things. If this goes on forever—if for any group, there is a group with it as a member—then there are infinitely many groups, and so infinitely many things. Perhaps, then, the problem about infinity is solved! Observe that this way out won't work for Goodman; on his account, {Bill} is just Bill; thus {{Bill}} is Bill, etc.; the only concrete thing in the series is Bill, so the series doesn't generate infinitely many things. But this suggests that, if we have solved the problem, it comes at a cost: Goodman found a way to make groups concrete; but we have given up his solution, and if the groups in the series aren't concrete, then we seem to have accepted a sort of Platonism.

Say we are willing to accept whatever Platonism is implicit in this approach to groups. Still, Bill, and indeed every concrete object, might not have existed. There might have been no concrete objects. Thus mathematicians have moved to what we might call “pure” group (or set) theory. On this view, there is a group with no members—the “empty” group. The empty group does not depend for its existence on any concrete thing. Given this, we have,

$$\{ \}, \{ \{ \} \}, \{ \{ \{ \} \} \}, \{ \{ \{ \{ \} \} \} \}, \text{etc.}$$

the empty group, the group whose only member is the empty group, etc. But now the groups aren't concrete at all—they seem every bit as problematic as Plato's numbers. Once again, we seem to have sacrificed the benefits nominalism was supposed to supply. Thus, perhaps a bit discouraged by now, let's consider another option.

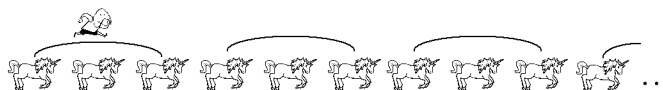
(D) *Deductivism* is another view on which there are, literally speaking, no numbers. On this view, ' $3 + 2 = 5$ ', say, is short for something like, ' $3 + 2 = 5$  follows from such-and-such premises in such-and-such logical system'. And, in general, mathematical claims are to be understood as saying no more than that *if* such-and-such premises are true, then such-and-such consequences follow.<sup>12</sup> There are two important points to be made: First, conditional claims of this sort may be *true*, even though there are no numbers. Second, reasonings about numbers may have a *form* or pattern with direct application to ordinary things.

By way of analogy, suppose some child is presented with a book which develops logical principles in application to unicorns: if every unicorn has a horn, and Morgan is a unicorn, then Morgan has a horn; if Morgan is a unicorn or a horse, and Morgan isn't a horse, then Morgan is a unicorn; etc. So far, principles in the book may be correct, though there are no unicorns. *If* every unicorn has a horn and Morgan is a unicorn, *then* Morgan has a horn—whether there are unicorns or not. If the book says only what follows from certain premises about unicorns, it need not be committed to the unicorns themselves. Even so, the book need not be *mere* fantasy. Its reasonings have a *form*

---

<sup>12</sup> This view is sometimes called “if-thenism.” Deductivism is closely related to (and sometimes treated as a version of) “formalism.” For discussion, and references to thinkers who have advocated these views, see chapters two and three of M. Resnik, *Frege and the Philosophy of Mathematics* (Ithaca: Cornell University Press, 1980).

which transfers to the concrete world: if all men are mortal, and Socrates is a man, then Socrates is mortal; if Socrates is a man or a woman, and Socrates isn't a woman, then Socrates is a man; etc. That is, we treat unicorns as “placeholders” for ordinary objects, and reason in the same way about the ordinary things. Similarly, suppose the book takes up some premises according to which there is an infinitely long unicorn parade, with amazing performers who leap from one unicorn to another. The task is to determine where the performers are after arbitrary series of uniform leaps.



To accommodate series and leaps of different lengths, the book develops an algorithm (with surprising analogies to ordinary multiplication) to determine where they are. But, if the algorithm is adequate to the unicorn case, it should be adequate to concrete cases as well: *if* things are lined up in the appropriate way, and we pass from one thing to another by uniform “leaps,” then the very same method should apply—and this whether there are unicorns (numbers) or not. Similarly, then, if mathematicians tell us merely about what follows from premises about numbers (or, for that matter, unicorns), they need not be committed to numbers themselves. And what the mathematicians tell us may exemplify forms with concrete applications. Mathematicians set up the forms, and those of us who apply mathematics—whether in simple counting or in complex physics—use what the mathematicians have done to reach substantive conclusions about the world.

Again, I'll mention two problems. First, it is part of our data that mathematical statements are either true or false. Either every even number greater than two is the sum of two primes or not. The deductivist can't admit this in a straightforward way. Rather, she offers, “Either it follows from the premises of arithmetic that every even number greater than two is the sum of two primes, or it follows from the premises of arithmetic that not every even number greater than two is the sum of two primes.” Say we are content with this. Still, it's not clear that we always *can* get this. In a series of astonishing results, it has been shown that there are important mathematical statements  $\mathcal{P}$  in basic mathematical systems such that neither  $\mathcal{P}$  nor not- $\mathcal{P}$  is a consequence of standard premises for those systems, and that for *any* appropriately specifiable premises adequate to the basic operations of arithmetic, if those premises are not contradictory, then there is sure to be at least *some*  $\mathcal{P}$  such that neither  $\mathcal{P}$  nor not- $\mathcal{P}$  is a consequence of the premises.<sup>13</sup> It is possible to argue about the significance of these results. They do show this much: *If*, as one might have thought, for any mathematical system, the deductivist proposes to find some *one* collection of premises such that for any  $\mathcal{P}$  in that system ‘ $\mathcal{P}$  is true’ translates into, ‘ $\mathcal{P}$  follows from the premises’, *then* the deductivist can't preserve the datum that for any  $\mathcal{P}$  in that system, either  $\mathcal{P}$  is true or not- $\mathcal{P}$  is true—because

<sup>13</sup> I'm thinking of the Gödel/Cohen proof that neither the continuum hypothesis nor its negation follows from the axioms of ZF set theory, and Gödel's incompleteness theorem. These results are a topic for advanced courses in logic, though Nagel and Newman, *Gödel's Proof* (New York: New York University Press, 1986) is a readable introduction to the latter.

it is sometimes the case that neither ‘ $\mathcal{P}$  follows from the premises’ nor ‘not- $\mathcal{P}$  follows from the premises’ is true. At least, then, we need to know more about how the deductivist will respond to this situation.

Second, there is a problem about application. The proposal is that those who apply mathematics use what the mathematicians do to reach substantive conclusions about the world. But under what conditions does one reach substantive conclusions about the world? Well, when there are the right analogies to premises for numbers. If the analogies hold, the methods of mathematics apply. But for what sorts of things do the analogies hold? We’re hoping for concrete physical structures. And this returns us to problems of nominalism: Standard premises for, say, real numbers or the calculus aren’t true except on complex infinite structures—and it’s at least not clear that ordinary things are this way. Even the algorithm for multiplication applies, in full generality, only on the assumption that there are infinitely many numbers. The child’s book assumes an *infinite* series of unicorns—as our ordinary algorithm for multiplication assumes infinitely many numbers. So long as it is unclear whether relevant premises are true, it is a mystery how or why methods of mathematics should apply. If the premises aren’t true, there is no reason why a statement of the sort ‘if the premises are true then the conclusion follows’ should be relevant or interesting. One might argue that there are abstract “extensions” of ordinary finite structures—so that the premises are true of the finite structures plus the abstract extensions—but this sounds like Platonism.

Of course, we haven’t yet considered all the approaches to mathematical truth or all the possible objections and replies. But it begins to look like a solution is no simple matter.

### *III. Truth as Correspondence*

In this section, I take up a first reaction to our project as illustrated above. Insofar as metaphysical questions arise when we ask what it is that makes ordinary theories and claims true, one motive for rejecting the project may be the thought that truth is somehow “up to us.” This response is sometimes encountered under the slogan, “true for me, but not for you.” Taken seriously and literally, I think this is a dark and mysterious saying. Here’s why: We have suggested that a statement is *true* if and only if reality is as the statement represents it to be—if and only if what it says corresponds to reality—and *false* if and only if it does not. Call this idea, “truth as correspondence.” But, given this approach to truth, there is an immediate problem for “true for me but not for you.” Suppose a statement correctly represents the world, and ask yourself: Is this person such that the statement is true? Is that person such that the statement is true? etc. If the questions make sense at all, you should answer ‘Yes’ to each—since, by hypothesis, the statement correctly represents the world. Or suppose some statement does not correctly represent the world, and ask: Is this person such that the statement is true? Is that person such that the statement is true? etc. This time, if the questions make sense, you’ll have to answer ‘No’—for, by hypothesis, the statement does not correctly represent the world. If truth is correspondence, then truth is “fixed” by the nature of the statement together with

the nature of what it represents; so long as a statement isn't about what people think, truth and falsity don't involve what people think; so truth and falsity don't vary from person to person. But if truth doesn't vary from person to person, there is no "true for me and false for you." The point might be worked out in different ways: One might say that if a statement is true (false) for anyone—if it correctly represents the way things are—then it is true (false) for everyone. Or, better, one might say that there is no "true for" at all; a statement isn't true *for* anyone, but is simply *true* or *false*.

Perhaps, though, saying that such-and-such is "true for me but not for you" does make sense. Perhaps when people say that such-and-such is "true for me but not for you" they don't mean to say that something is *true* for them but not for you; rather, they mean to say only something like, such-and-such is "what I believe but not what you believe," or "what I have good reasons for believing but you do not." These latter claims do make perfect sense. Even today, some people believe that the earth is round and others believe that it is flat. And, similarly, some people have access to reasons which others do not. Still, granting the coherence of these claims about beliefs and reasons, it's not clear that "true for me but not for you" is therefore vindicated. The saying remains, at best, a misleading and loose way of speaking. In fact, it seems a way of obscuring real and significant differences. Given this, I make the following simple proposal: In serious speech, at least, let's say what we mean. If "true for me but not for you" translates into some claim about belief or justification, leave "true for me but not for you" to the side, and make the corresponding explicit claim about belief and/or justification.

Perhaps you agree that it is appropriate to abandon "true for me but not for you" in serious speech. But perhaps you have questions to the effect that there is still some serious sense in which a thing may be true for one person and not for another. Because the issue is so important, I turn to a series of objections and replies.

*What if some people believe one way and other people believe another—doesn't this mean that the thing is true for some people and not for others?* As we have just seen, the answer is, "No." Suppose some people believe that the earth is round and others believe that it is flat. On a correspondence account, a belief is true if it correctly represents the way things are. In this case, the beliefs represent that the earth has a certain shape and so are true if the earth is that way. But, whatever shape the earth has, it isn't both round and flat. So not both beliefs are true. Of course, there is no problem admitting that the different people *have* different beliefs about the earth, but admitting this isn't the same as admitting that the different beliefs are both *true*.

*What if some people have good reasons for believing one way and other people have good reasons for believing another—doesn't this mean that the thing is true for some people and not for others?* No. Here's a silly, but dramatic, example: Perhaps some people have good reasons for thinking that you are an alien (they saw you emerge from a spaceship), while you have good reasons for thinking that you are not (you remember the event as a hoax). But having reasons is not the same as having truth: Maybe the event was a hoax—but maybe you are an alien and hoax memories were implanted when you

took human form! The reasons do not themselves *make* “You are an alien” true or false; rather, truth or falsity depends on the relation between what the statement says and the way things are—on whether you are, in fact, an alien. Reasons do not *make* truth, they are rather part of the process by which we *discover* it.

*What if there is no test that can decide for or against some statement—doesn’t this mean its truth or falsity depends on opinion?* No. Suppose the universe goes on forever, and consider the following claim,

There is, on some planet in the universe, a rock formation that is an exact duplicate of the Venus de Milo.

One can imagine exploring planets that are not too distant from our own. However, if the universe goes on forever (as we are assuming), it seems impossible to explore them all. In exploring the planets, we might discover that the statement is true, but its truth or falsity might remain forever unknown. In this case, should we conclude that the statement isn’t true or false, or that its truth or falsity is up to the individual? What people *think* about its truth or falsity may vary from person to person. But the truth or falsity doesn’t vary. The statement says something perfectly clear about the way things are. It is true if things are that way and false if they are not. As it turns out, our power to represent with language outruns our ability to discover with spaceships (or whatever). And along with the ability to represent beyond what we can discover comes the potential for truth and falsity beyond that which we can know.

*What if different people mean different things by the same words—if the same words represent the world being different ways, won’t they be true for one person and false for another?* Good point! This forces a certain clarification of our thesis: There is room to distinguish a *sentence* from *what one says* with a particular use of it. So, e.g., your use of “I am hungry” says that you are hungry, and another person’s use of it says that he is hungry. If you use the sentence just before lunch, and he uses it just after lunch, it is likely that what you say is true and what he says is false. Therefore it isn’t the *sentence*, taken apart from its uses, which is true or false, but what one says with particular uses, that is true or false.<sup>14</sup> If what you say (the *proposition* you express) with a particular use of “I am hungry” correctly represents the way things are, then the truth of what you say (the proposition you express) doesn’t vary from person to person; for its truth is a matter of correspondence between what you say and the world; if you say it, what you say is true just in case, as a matter of fact, you are hungry. And, similarly, if meanings shift across uses—if one person uses words one way, and another uses them a different way—it may be that different propositions are expressed with the same words. But the truth or falsity of the different propositions is a matter of correspondence, and so

---

<sup>14</sup> Perhaps what one says with any use (in a single language) of a sentence like “Dogs are mammals” is true, and if what one says with all a sentence’s uses is true (false) it is natural to say that the sentence is itself true (false); but, plausibly, this truth or falsity is parasitic on truth or falsity for what one says with the uses.

does not vary from person to person. Above, I have used the relatively vague term, “statement.” Properly understood, though, our thesis is this: The truth or falsity of a proposition—of what one says with a use of a sentence—does not vary from person to person. Given this, propositions aren’t true for one person and false for another, and the point about different meanings for words doesn’t count against our thesis.

*What about religious propositions*—aren’t they obviously true for some people and false for others? No—at least not obviously! Perhaps the difficulty is one about verification: It seems impossible to find a test that could decide, for everyone, whether certain religious propositions are true or false. Say there is no test. Then the situation for religion is like the situation for Venus-de-Milo rock formations. As before, though, a problem about *knowledge* isn’t the same as a problem about *truth*. Standard religious propositions like the one expressed by “God created the heavens and the earth” represent things as being a certain way. The propositions are true just in case things are in fact that way. So their truth and falsity is no different from truth and falsity for any others.

*What about taste*—doesn’t truth or falsity for propositions about taste obviously vary from person to person? No—at least not obviously! Consider, e.g., “Nose rings are cool.” If it is indeed a matter of taste whether nose rings are cool, then a use of “Nose rings are cool” should come with an understood “I think...” (or something of the sort) so that “Nose rings are cool” is like “I am hungry.”<sup>15</sup> One use says that one person or group likes nose rings, where another use may say that another person or group likes them. But the proposition expressed by “(I think) nose rings are cool” is true if and only if it correctly represents the way things are—and this is not something that varies from person to person. Similarly, for other propositions expressed by uses of “Nose rings are cool,” and for other propositions about taste.

*What about moral propositions*—aren’t at least these true for some people and false for others? No—at least not obviously! In this case, the worry might be like the worry about taste or like the worry about religion. On the one hand, if ethical claims like “Murder is morally wrong” are supposed to be a matter of taste, then their uses come with an assumed “I think...” (or something of the sort). But then, as before, truth or falsity for the corresponding propositions does not vary from person to person. Different uses may express different propositions, but the individual propositions aren’t true for some people and false for others. Perhaps, then, ethical claims aren’t supposed to be a matter of taste, and the worry is about knowledge—say, about discovering God’s commands. In this case, though, it’s not clear why the problem about knowledge is supposed to be a problem about truth. Consider the Venus-de-Milo case.

*But doesn’t the correspondence thesis promote intolerance and bigotry*—and isn’t it therefore best to reject it? Presumably, the idea is that “true for me but not for you” removes potential for error, so that each person and culture has as much “right” to their

---

<sup>15</sup> If it doesn’t come with an understood “I think...” (or something of the sort), then we aren’t expressing a matter of taste—as we have supposed.

beliefs as any other. This raises many issues. I'll make just a couple of comments about motivation: First, do not mistake the claim that truth and falsity don't vary from person to person, with the claim that any one group is justified in their beliefs, or that some one group knows what the truth is. So, e.g., scientific groups can disagree about some claim while respecting and understanding the process by which other groups come to their conclusions—all the while insisting that not all the groups can be right. Second, tolerance is a virtue which has its application precisely where there is disagreement and error. Insofar as “true for me but not for you” removes the potential for error, it removes the occasion for tolerance. But maybe it would be best to admit error and to promote the virtue.<sup>16</sup>

*But isn't there another account of truth*—one that would make sense of “true for me but not for you”? There are approaches to truth other than the correspondence theory. The correspondence theory fits most naturally with a common-sense (*realist*) view of the world according to which the world exists externally to us and independently of what we think. So, e.g., on a realist view, there may have been rocks and trees before there were people, and there might be rocks and trees after people are gone. But there are other (*anti-realist*) theories of the world which typically make the world itself somehow dependent on minds—and typically come with alternate approaches to truth. We'll get to this in part III. I think most people who say that such-and-such is “true for me but not for you” do not have in mind these sophisticated theories. More likely, either they mean to say something about belief or justification—and so do not say what they mean, or they mean to say something about truth which, as it turns out, does not mesh with their overall view of the world. To be clear and consistent, people in this position should give up saying “true for me but not for you.” Insofar as we are assuming that there is an “external” world which our theorizing has as its object, we should give it up as well—at least for now. And, if we do give it up, then the question of what in the world *makes* propositions true, reemerges as an important one.

#### *IV. Further Applications*

With our discussions of truth and mathematical truth as background, in this section, I make a series of remarks about the nature and significance of metaphysics. The remarks are a grab-bag, and come in no particular order.

(A) First, insofar as my application of the mathematical example is successful, the “the solution is easy...” response to sentences like (2), (3) and (4) should seem less plausible. The point isn't that there is no solution. Rather, it is that solutions do not lie on the surface, and that finding them requires work. Strangely, with respect to numbers,

---

<sup>16</sup> It isn't clear that “true for me but not for you” removes all potential for disagreement and error. When someone says that such-and-such is “true for me but not for you” is *this* supposed to be a correspondence truth, or is it supposed to be true for them and false for you? If it is supposed to be a correspondence truth, then there is room to disagree. If it is only supposed to be true for them, then you still disagree! (Think about it.) We will return to this in part III.



fictional objects, etc. people literally “do not know what they are talking about”! Insofar as one cares to know what one is talking about, this should undercut the “it’s uninteresting” reaction, and enhance the “how fascinating!” one.

(B) But maybe your reaction continues to be “the solution is easy...” Perhaps you have the answer! However, some words of caution are in order. First, details matter. The Newtonian theory of gravitation is pretty good. (!) As it happens, though, this theory fails to predict a certain motion of the orbit of Mercury—it misses by about 1/100 of a degree of arc per century. For the most part, Newton’s theory and Einstein’s theory of general relativity predict the same results. However, it matters that general relativity gets the motion of Mercury right. In physics, small anomalies can suggest sweeping theoretical changes, where these changes are enormously important insofar as we care to know what the world is like. And similarly for metaphysics. So, e.g., difficulties about infinity, or about making every mathematical proposition true or false, may suggest sweeping changes in an overall account of what exists. In the end, attention to details enables us to recognize issues and questions that might otherwise go unnoticed. Indeed, it’s only attention to details that motivates our questions in the first place.

Second, a related point. We have considered a series of proposals according to which such-and-such statements are true under such-and-such conditions—where our discussion centered on the adequacy of *those very conditions*. So, e.g., the conceptualist holds that numbers *are* ideas; we have held her to this, and so concluded that there couldn’t be more numbers than there are ideas. It is almost impossible to understand this discussion without taking the conceptualist’s position seriously and literally. It is important to set aside preconceptions according to which the answer is obvious or a condition “really” includes some unstated qualifications. It’s the specific proposal on the table that’s under consideration. Of course, it will be a legitimate criticism if, in the end, some important addition or qualification is left unconsidered. But that there is an alternative to some condition is not immediately relevant to discussion of *that very condition*. If you keep this in mind—if you consider proposed conditions literally, and on their own terms—then I think our discussions will take on the feel of a pleasing point-and-counterpoint dialogue, rather than that of an irritating series of arbitrary assertions. If your reaction is more like the latter than the former, or if you are continuing to have the “the solution is easy...” reaction, I suggest reconsidering the previous sections with the material of these last two paragraphs in mind.

(C) Perhaps, without any confusion of the “true for me but not for you” variety, someone just doesn’t care about truth—or doesn’t care about it so long as it lacks personal existential significance. This is the most difficult response to rebut. A person who wants arithmetic only to balance his checkbook is not going to care about numbers as such; and similarly for other metaphysical questions. I can think of two positive responses: The first, and most important, is “that the problems are there.” I suppose that knowledge and learning are goods for their own sake. It is part of being human to be curious. Questions about truth conditions expose fundamental gaps in understanding. So one wants to know.

But, second, pure research has its applications. It's not obvious that debates about mathematical truth are irrelevant to questions about the appropriate contexts for mathematical applications. Such issues seem relevant to our discussion of deductivism. And, as in discussion of nominalism, philosophical debates may even be relevant to which mathematical statements we think are true. This could be of substantial importance. Further, discussion of truth conditions plays directly into the way we think about inference patterns for logic and reasoning. Thus metaphysics influences other disciplines by way of the reasoning patterns they employ. So, e.g., one version of Anselm's ontological argument for the existence of God has premises which involve possibility and necessity. Similarly, a moral discussion may turn on whether some subject could or could not have done otherwise in some circumstance. But the way we reason with possibility and necessity depends on what we think makes things possible and necessary—on what we think about things like “possibilities.” And these are questions of metaphysics (to which we will turn in later chapters). So the religious and ethical conclusions seem to depend on prior metaphysical conclusions about possibility. Metaphysics thus has a sort of “fundamental” importance; though one is not interested in some metaphysical issue for its own sake, that very issue may matter crucially for matters with which one is directly concerned.

Of course, it is my conviction that you will be a better person—more enlightened, with better critical and reasoning skills, a better writer, with improved ability to take on tasks for their own sake—for having considered these issues. I would like to advertise powers to improve complexion, cure the common cold, etc. But that would be going too far.

(D) Finally, it should be clear from the mathematical example that, as I have suggested, the metaphysician offers theories about what there is in the world to account for the truth of ordinary theories and claims. Certainly, we have not attempted to *deduce* some one account of the numbers. And our theories *seem* to be about what there is (abstract objects, groups, etc.) in the external world. Our “window” into this discussion is consideration of what makes ordinary theories and claims true. And the data against which our theories are tested comes from the ordinary theories and claims as well. A couple of last comments on this:

First, it's hard to deny the sort of data with which we are working—but it can be done. In general, there are three options. (i) It is possible to deny some bit of data. Each of the mathematical proposals we have considered may seem to do at least some of this. (ii) It is possible to accept data and to account for truth by supplying appropriate objects. This is the primary Platonist and conceptualist strategy. And (iii) it is possible to accept data in some “reinterpreted” form, and to account for truth by supplying objects, if any, required by the modified version. This is the primary nominalist and deductivist strategy. There are different philosophical attitudes one can have toward the sort of data with which we are working. In the early part of the twentieth century, it wasn't all that uncommon for philosophers to use their metaphysical vision to divide among the ordinary claims and theories that were to count as acceptable, and those that were not. This

approach has seemed problematic precisely because it may end up denying claims which are obviously true. Statements like (2), (3) and (4) seem more secure than most metaphysical theories. I expect that *you* will go on believing that seven is prime even if you become convinced that there is no adequate account of the numbers! Insofar as data counts as secure, it provides an anchor or reference point so that metaphysical discussion is not mere speculation. I think a grasp of this dynamic makes sense of very much contemporary philosophy which may seem bizarre or obscure on first exposure.

Second, to say that our data raise metaphysical questions, and provide an anchor for their resolution, is not yet to say what metaphysics is. Different accounts are possible. In a sense, any question about what the world is like is a metaphysical one. But philosophical metaphysics is more narrow than this. We get closer to “metaphysics proper” when we ask for a catalog of the *kinds* of things there are. Is there a physical world? are there abstract objects? properties? possibilities? events? unobservable physical objects? minds? spirits? numbers? fictional objects? works of art? etc. Can some of these be understood as instances of others? All of these are metaphysical questions. Questions about unobservable physical objects, minds, spirits, numbers, fictional objects and works of art come up for discussion in the more metaphysical parts of specific branches of philosophy: philosophy of science, philosophy of mind, philosophy of religion, philosophy of mathematics, philosophy of art, etc. But questions about the physical world, abstract objects, properties, possibilities, events, and the like, remain, I think, in the sphere of metaphysics proper. These are so basic and so general that they do not fall into the specific sphere of the more particular branches. One might say that *metaphysics* (proper) is after a catalog of the fundamental constituents of the world—where other branches are more interested in how those constituents matter for the particular objects, which may or may not be fundamental, of concern to the branch in question. In this text, we will ask how to go about answering metaphysical questions in general, and focus on objects specific to metaphysics proper—on objects that seem to underlie work in other disciplines. I think a focus on this relatively narrow “core” will illuminate other topics that plausibly fall under the heading “metaphysics” and beyond.<sup>17</sup>

---

<sup>17</sup> It should go without saying that the content of this text is distinct from much of what falls under the heading “metaphysics” in a popular bookstore.

## Part II

### Quine's Method for Metaphysics

## Chapter Two

### Plato's Beard: A Problem About Method

In Part I, we developed an overall picture of the metaphysical project. In Part II, we think more carefully about the way metaphysical questions arise and are addressed. This discussion of method then forms a background for the discussion of particular metaphysical problems in Parts III and IV. It won't be possible to discuss metaphysical method without doing some metaphysics. And the particular problems have implications for method. So there is no hard-and-fast division between the sections. At any rate, in this section, we emphasize, or focus on, questions of method by working through W. V. O. Quine's classic article, "On What There Is." Quine may or may not be right—it is important to think carefully about what he says. Still, Quine's article raises issues which go to the heart of what contemporary metaphysicians do, and so provides an excellent focus for a discussion of method, and introduction to contemporary metaphysics. In this chapter, we take up just the very first part of Quine's article.

#### *I. Plato's Beard*

Quine begins by developing a problem about method. We will be able to see what is at issue, against the background of what we have already done. Thus, we'll turn to the problem itself, after saying a bit about what has gone before.

(A) Our idea is to "get behind" ordinary theories and claims and ask what does, or could, make them true. Roughly, when confronted with 'seven is prime', or 'Frodo is a hobbit', we have either taken the statement itself seriously and literally, or have looked for an alternative which, when taken literally, gives the truth condition for the original. Our literal understanding has itself been guided (at least in part) by **T1**. So, e.g., the Platonist applies **T1** directly to 'seven is prime'. But the nominalist looks for other expressions which, in the end, turn out to be about complex things and their parts; if it applies at all, **T1** applies to these. Thus our approach involves a sort of "regimentation" of language: Though truth conditions for ordinary expressions are not clear, it seems possible to regiment the interpretation of at least some uses of some expressions, so that truth conditions for those expressions are clear. We then use the expressions whose truth conditions are clear to explicate those whose truth conditions are not. But we need to think carefully about this process of regimentation.<sup>1</sup>

---

<sup>1</sup> One might think that there is a problem about using one bit of language to explicate truth conditions for another. Granted, one wouldn't want to say that one bit of language acquires its significance from another bit of language, which acquires its significance from another, which.... But that's not what we're doing. The question about the origin of significance is a question for philosophy of language. But we're interested in something much more modest: Insofar as language *has* significance, it should be possible to use it to say something about how language works. And that's all we need. Think about **T1**.

For our purposes, it is natural to suppose that *any* mode of regimentation will do. The point is *that* some expressions have definite truth conditions, not *how* they get them—that we can say literally what makes an expression true or false, not the particular manner by which we do so. Thus, e.g., one might give truth conditions by using some well-understood expressions of English, but one might also use Chinese, or some specially designed code. So far, we have applied **T1** to expressions of English, and this has had the desired result: it is clear what is required for the truth of subject-predicate statements to which **T1** is applied. Of course, **T1**'s *applicability* may be too narrow—so that it requires supplementation. For example, 'If dogs fly then I'll eat my hat' and ' $3 + 2 = 5$ ' aren't in the subject-predicate form, and we require some guidance as to the serious and literal condition for their truth. Perhaps there is some natural extension of **T1** that would apply to such expressions (in discussion of ' $3 + 2 = 5$ ' and the like, we have presumed that this is so); however, as it stands, **T1** isn't generally applicable. Thus it's natural to set out to extend it.

But this isn't Quine's strategy. Rather, Quine argues that **T1** is fundamentally flawed—that it shouldn't play even a role in regimentation. In "On What There Is," he argues for this conclusion by means of a debate with a couple of fictional (!) opponents, McX and Wyman.<sup>2</sup> Quine suggests that McX is led by **T1** into a sort of conceptualism, and Wyman into a sort of Platonism. Quine rejects their conceptualism and Platonism, and so rejects **T1**. Thus, in simplified form, he argues,

If we accept **T1**, we are led to accept conceptualism or Platonism.  
We shouldn't accept conceptualism or Platonism.

---

We shouldn't accept **T1**.

We'll think carefully about this reasoning. But notice: Even if Quine's specific arguments against conceptualism and Platonism are not decisive, it may be that **T1** remains problematic. For it would be inappropriate to address metaphysical questions by means of a mode of regimentation which is itself somehow biased in favor of Platonism or conceptualism. After a brief description of the pressure generated by **T1**, we'll move to the debate with McX, and the debate with Wyman.

(B) Suppose McX and Wyman differ with Quine about what there is, having the positive view that there is some thing which Quine thinks does not exist. Say the thing is Pegasus. They hold that Pegasus exists, and Quine that he does not. Quine's position seems perfectly intelligible—perhaps most of us agree. But **T1** puts Quine in a difficult

---

<sup>2</sup> Wyman seems to be a caricature of Alexius Meinong, an Austrian philosopher and psychologist who did much of his work around 1900. See, e.g., his "On the Theory of Objects," in *Realism and the Background of Phenomenology*, ed., R. Chisholm (Glencoe: The Free Press, 1960). McX may be a parody G. E. Moore and Bertrand Russell (at an early stage). See, e.g., Russell, *Principles of Mathematics* 2nd ed. (New York: Norton, 1964), p. xviii. We'll be thinking about McX and Wyman rather than Meinong, Moore or Russell.

position. Indeed, his position *seems* incoherent. Quine says, “Pegasus does not exist.” But, given **T1**,

- T1** A subject-predicate statement is *true* if and only if the subject term refers to some object, and the predicate term applies to the object to which the subject term refers.

Quine’s claim is true just in case there is some object to which ‘Pegasus’ refers and ‘does not exist’ applies. But the only things there are, to which ‘Pegasus’ might refer, are ones that exist. So ‘does not exist’ doesn’t apply to any of them. So, given **T1**, Quine’s claim *cannot* be true. This result is perfectly general. Given **T1**, *any* claim of the sort, “such-and-such does not exist” is true if and only if there *exists* some thing which *does not exist*. McX and Wyman prefer to avoid this incoherence, and so conclude that Pegasus exists. In this, they are not alone. This problem about denying the existence of particular things is an ancient one, which Quine says might be nicknamed, “Plato’s beard.”<sup>3</sup>

## II. McX’s “Confusion”

Although McX concludes that Pegasus exists, being a sensible fellow, he doesn’t say that Pegasus is a flesh-and-blood flying horse. Rather, McX holds that Pegasus is an *idea*. Thus McX advocates a sort of conceptualism. Quine hints that the idea proposal isn’t altogether clear. So, e.g., it’s not clear whether an idea is supposed to be a belief, a collection of beliefs, a mental image, or something else. And, whatever ideas may be, it’s hardly clear that there is only one Pegasus-idea: my Pegasus-idea may include associations which yours does not. Still, appeal to ideas is common in philosophy and, for the sake of argument, Quine allows that there is a mental Pegasus-idea. Unfortunately, however, this is not the end of the matter. We’ll get to Quine’s main objection after discussion of a preliminary worry.

(A) According to McX, Pegasus is an idea. Given **T1**, ‘Pegasus is an idea’ is true just in case ‘Pegasus’ refers to some thing to which ‘is an idea’ applies. On McX’s view, this is so, and ‘Pegasus is an idea’ is therefore true. But it’s also supposed to be true that,

- (1) Pegasus is a flying horse.

Given **T1**, (1) is true just in case ‘Pegasus’ refers to some thing to which ‘is a flying horse’ applies. But, of course, no idea is a flying horse—no idea has hair, hooves or wings. So ‘is a flying horse’ doesn’t apply to the thing to which ‘Pegasus’ refers and, on McX’s view, given **T1**, (1) isn’t true.

---

<sup>3</sup> But the ancient figure whose reasoning is most like McX and Wyman’s is Plato’s contemporary Parmenides. Parmenides seems to have used such reasoning to conclude that being is eternal, unchanging, undivided and homogeneous. So, e.g., it makes no sense to talk about what isn’t; but if being isn’t eternal, it comes from what isn’t, and if it changes, it becomes what it isn’t. Understandably, many philosophers have wanted to resist these conclusions. But it’s not always been entirely clear how to do so.

Clearly, McX needs some reply. One option is to focus on the *representative* character of Pegasus (the idea). Perhaps (1) is elliptical for,

- (2) Pegasus represents a flying horse.

Given this, **T1** applies in the usual way: (2) is true just in case ‘Pegasus’ refers to some thing to which ‘represents a flying horse’ applies. Plausibly, Pegasus (the idea) represents a flying horse, and (2) is therefore true. The move from (1) to (2) simply converts ‘is’ to ‘represents’. But this doesn’t always work: we wouldn’t want to convert ‘Pegasus is winged’ simply to ‘Pegasus represents winged’. Still, one might treat ‘Pegasus is winged’ as elliptical for ‘Pegasus is a winged creature’ and this, in turn, as elliptical for ‘Pegasus represents a winged creature.’ Then **T1** applies in the usual way. Good!

On this account, it is important that there be some sense in which Pegasus *represents* a flying horse without there being a flying horse it represents. On pain of introducing a Pegasus-object in addition to the Pegasus-idea, somehow, ‘represents a flying horse’ should apply apart from any object represented. Let’s assume there is some way to make this work. Also, in accepting this proposal, McX postulates systematic ambiguities for ordinary predicates. Sometimes “is a such-and-such” says “(literally) is a such-and-such”—as when we say that Pegasus is an idea. Sometimes “is a such and such” says “represents a such-and-such”—as when we say that Pegasus is a flying horse. Sometimes “is such-and-such” means “is a such-and-such creature,” which in turn means “represents a such-and-such creature”—as when we say that Pegasus is winged. And that is not all. In saying that Pegasus exists, one might be saying that Pegasus (the idea) exists, but one might also be saying that Pegasus (the idea) is an idea of some concrete object. On pain of incoherence, however, when one says that Pegasus does not exist, one is presumably saying just that Pegasus (the idea) is not the idea of any concrete object. The suggestion that there are such systematic ambiguities may or may not be plausible. And the ambiguities may or may not pose a difficulty for regimentation. Perhaps, though, McX has a workable response to the difficulty about (1).

(B) Quine objects not against McX’s treatment of the predicate, but rather against his treatment of the subject.<sup>4</sup> He accuses McX of a simple confusion: In ordinary cases, it is natural to distinguish the idea of a thing from the thing of which it is an idea. So, e.g., the Parthenon is distinct from the Parthenon-idea.<sup>5</sup> The Parthenon is physical, but the Parthenon-idea is not; the Parthenon is visible, but the Parthenon-idea is not; etc. As Quine observes, “we cannot easily imagine two things more unlike, and less liable to confusion, than the Parthenon and the Parthenon-idea” (e3). According to Quine, McX

---

<sup>4</sup> There are different ways to understand Quine’s objection. It may be that Quine is unhappy with predicate ambiguities. But Wyman requires related ambiguities, and Quine does not object to them (see below). So, plausibly, Quine’s point is just the one I give him.

<sup>5</sup> Just in case: the Parthenon is a Greek temple from the fifth century BC, with remains in Athens to this day.



distinguishes the Parthenon from the Parthenon-idea. Somehow, though, McX thinks Pegasus and the Pegasus-idea are one. So the objection is simple: When we talk about the Parthenon, we're not talking about the Parthenon-idea. Similarly, it may seem obvious that when we talk about Pegasus we're not talking about the Pegasus-idea. If this is right, McX's view is false. To see the force of Quine's point, it's worth considering how McX might handle reference for terms like 'Pegasus' and 'Parthenon'.

Perhaps McX is located in a certain "classical" tradition. Many, including Descartes, Hobbes, and Locke, have thought that ordinary names always refer, in the first instance, to ideas. Reference to "external" objects is only through or by means of the ideas.<sup>6</sup> This "extended" or "derivative" reference is required because we sometimes want to talk about the *object* of an idea, rather than about the idea itself. As J.S. Mill observes against this tradition in his 1867 *A System of Logic*,

When I use a name for the purpose of expressing a belief, it is a belief concerning the thing itself, not concerning my idea of it. When I say, "the sun is the cause of day," I do not mean that my idea of the sun causes or excites in me the idea of day: or in other words, that thinking of the sun makes me think of day. I mean, that a certain physical fact, which is called the sun's presence... causes another physical fact, which is called day.<sup>7</sup>

In application of **T1**, we need at least a *sort* of reference to things (as opposed to ideas) to make the truth of a statement about the sun depend on whether its predicate applies to *the sun*. So far, we have thought of an idea's representing as something entirely "internal" to the idea. When one says an idea represents that such-and-such, one is talking about the idea alone. But this isn't sufficient. An idea of a thing may be one way, and the thing it represents another—as, e.g., I may be mistaken about the number of pillars in the Parthenon. In this case, 'has so many pillars' might apply—in the form, 'represents a so many pillared object'—to my Parthenon-idea but, if the idea is mistaken, not to the Parthenon. It is one thing to say that an idea represents in such-and-such way, and another to say that the thing it represents is that way. Thus McX needs something like a *subject* ambiguity: so far as application of **T1** is concerned, sometimes we should think of reference as to ideas, but sometimes we should think of it as to objects represented.<sup>8</sup>

---

<sup>6</sup> See, e.g., Locke, *An Essay concerning Human Understanding*, III.

<sup>7</sup> Mill, *A System of Logic*, chapter 2, in e.g. the 8th ed. (London: Longmans, 1961).

<sup>8</sup> This conclusion isn't restricted to the "classical" tradition in which McX has been located. However reference may work, McX requires some reference to things, and some reference to ideas. And McX may not want to be located in the classical tradition. In recent times, S. Kripke, *Naming and Necessity* (Cambridge: Harvard University Press, 1980) and others have subjected the proposal that reference to other objects is through or by means of idea *content*, to a withering attack (though their arguments are not directed specifically at this classical proposal—but rather at a descendent of it). Thus, e.g., in many cases, the content of an idea of a thing isn't sufficient to pick out the thing it is an idea of. So many of us have, at best, only the vaguest idea of Empedocles. Perhaps you have never heard of him. Perhaps you think of him only vaguely as "an ancient Greek" or "an ancient Greek philosopher." Of course,

But it's not entirely clear how this is supposed to work. Consider, first, a case where someone asks whether Pegasus is such that her idea of him is accurate. Surely it is *possible* to have the wrong idea about Pegasus—as someone who thinks that Pegasus is a flying pig. If her question makes any sense at all, given the beard argument, it looks like we need an object for Pegasus distinct from the idea of him. Or consider cases where existence is controversial. Consider, e.g., 'Arthur never had a round table.' Is this a false statement about an idea, or a potentially true statement about a historical figure? Say it is intended as the latter. But, if it turns out that there never was a historical Arthur, will it therefore turn out that 'Arthur never had a round table' was about the idea all along? Surely this is absurd. It is natural to think that we are in better "control" of our language than this. We can talk about things, and we can talk about ideas. But doing the one isn't the same as doing the other. In ordinary cases, we know what we are talking about, and 'Arthur never had a round table' simply *isn't* about the Arthur-idea—if it's about anything at all, it is about *Arthur*. Say McX accepts this—'Arthur' doesn't refer to the Arthur-idea. Given this, if it should turn out that there is no historical Arthur, McX is in the awkward position of accepting the "beard" argument according to which 'Arthur' refers something, but allowing that 'Arthur' doesn't refer to an idea; so he is in the awkward position of having to find some object other than an idea for 'Arthur'. If he does supply some such object, it is hard to see why he shouldn't depend entirely on objects of that sort, and so why he shouldn't abandon the appeal to ideas altogether.

And similarly for Pegasus and the Parthenon. In general, it is possible to distinguish reference to the idea of a thing from reference to the thing of which it is an idea. Indeed, this seems *required*, on McX's view, if we are to know the form in which a given predicate is supposed to apply. In ordinary cases, reference isn't to ideas. Given this, the "beard" argument presses in the direction of *non-idea* objects for 'Pegasus', 'Parthenon' and the like. That is, McX seems to reach the *wrong* conclusion from his own argument: its conclusion is that there are things to which 'Pegasus', 'Parthenon' and the like refer; thus, if these terms don't refer to ideas, the conclusion isn't relevant to ideas—but rather to the things to which the terms refer. Perhaps it is likely that there is a concrete non-idea object for 'Parthenon'. But what if there isn't? Maybe we have all been deceived about the Parthenon by Greek tourism officials! And it is unlikely that there is a concrete non-idea object for 'Pegasus.' But McX admits a need of objects for both. If he supplies some non-idea objects for 'Parthenon', 'Pegasus' and the like, his proposal may begin to merge with Wyman's. At any rate, with its postulated ambiguities, it may seem like McX's view lacks the sort of cohesion that could make it into a plausible theory. At least there is a problem about its use as a vehicle for regimentation. So it is reasonable to consider Wyman's alternative.

---

you may be an authority on Empedocles and his philosophy. Plausibly, however, even those of us who don't know much about Empedocles can sensibly *refer* to him and ask, say, how many elements he thought there were. But if we aren't sure how many elements he thought there were, it's unlikely that our idea of him is sufficient to distinguish him from various other Greek philosophers including, say, Anaximenes. So our idea of him isn't sufficient for reference to him.

### III. Wyman's "Slum"

Wyman accepts the "beard" argument according to which Pegasus exists. But he doesn't think that Pegasus is an idea. Rather, Wyman holds that Pegasus is an unactualized possible being. He thinks there *is* an unactualized possible being such that it is Pegasus. We should think of unactualized possible beings as abstract objects—and so of Wyman's view as a sort of Platonism. On this view, some beings are actualized, and others are not. You and I are actualized. But Pegasus, Superman, etc. lack actualization. This doesn't make them any less real than you and I—they only lack a property (actualization) that we happen to have. As it turns out, Wyman introduces a sort of vocabulary problem: According to Quine, the wily Wyman *agrees* that Pegasus does not exist; but maintains instead that Pegasus *is*. In granting that Pegasus does not exist, Wyman grants no more than that Pegasus is not concrete; for him, this does not entail that there is no Pegasus; rather it is to say only that Pegasus isn't actual. Quine agrees to give Wyman the word 'exist', and proposes to stick with 'is'. Let us do so as well.

Given Wyman's use of 'exist', Quine thinks not only that Pegasus does not exist, but also that Pegasus *is not*—there is no object, abstract or otherwise, that is Pegasus. Quine brings at least three objections against Wyman. Though he may not make his own best case, the first two, especially, are significant. After discussion of a preliminary worry, we'll take up Quine's objections one by one.

(A) Here's the preliminary worry: Like McX, Wyman postulates a sort of predicate ambiguity. According to Wyman, Pegasus is an unactualized possible and, given **T1**, 'Pegasus is an unactualized possible' is true just in case 'Pegasus' refers to some thing to which 'is an unactualized possible' applies. On Wyman's view, this is so, and 'Pegasus is an unactualized possible' is therefore true. But, as before, it is also supposed to be true that Pegasus is a flying horse. Given **T1**, 'Pegasus is a flying horse' is true just in case 'Pegasus' refers to some thing to which 'is a flying horse' applies. But, of course, no abstract object is a flying horse—no abstract object has hair, hooves or wings. So 'is a flying horse' doesn't apply to the thing to which 'Pegasus' refers and, on Wyman's view, given **T1**, 'Pegasus is a flying horse' isn't true.

Like McX, Wyman needs some reply. And, like McX, he might postulate a predicate ambiguity. Perhaps unactualized possibles somehow "represent" or "build in" certain features they *would* have *if* they were actualized.<sup>9</sup> If so, 'Pegasus is a flying horse' might be elliptical for,

(3) Pegasus "represents" a flying horse.

---

<sup>9</sup> T. Parsons, *Nonexistent Objects* (New Haven: Yale University Press, 1980), develops a sophisticated proposal along these lines. See also E. Zalta, *Abstract Objects* (Dordrecht: D. Reidel, 1983) and, for a relatively accessible introduction to his views, <http://mally.stanford.edu/theory.html>.

where “represents” is understood in the appropriate way. Given this, **T1** applies as usual: (3) is true just in case ‘Pegasus’ refers to something to which “represents” a flying horse’ applies. Plausibly, Pegasus (the unactualized possible) does build in the appropriate features, and (3) is therefore true. Similarly, ‘Pegasus is winged’ might be elliptical for ‘Pegasus “represents” a winged creature’. And, on Wyman’s terminology, ‘Pegasus does not exist’ says just that ‘Pegasus is unactualized’. Given the “beard” argument, ‘Pegasus is not’ is incoherent. But Wyman has no desire to say that Pegasus isn’t. So far, maybe, so good.

Though predicate ambiguity is required, Wyman’s view requires *no* reference ambiguity. Where McX allows that names may refer to objects or to ideas, Wyman insists that they consistently refer to objects. ‘The Parthenon’ consistently refers to the Parthenon and ‘Pegasus’ to Pegasus. One happens to be actualized and the other not. But this does not make reference in the one case any more or less to the object than in the other. And similarly for Arthur. Thus the possibility that Arthur does not exist (isn’t actualized) puts no pressure on Wyman to introduce *extra* objects. ‘Arthur’ refers to Arthur—whether he exists or not. (Of course, historical investigations may be required to determine whether Arthur has the property of existing—of being actualized.) Thus Wyman seems to have grasped the force of the “beard” argument, and to have accepted its full consequences. So Quine’s objections against Wyman aren’t the same as his objections against McX.

(B) First, Quine suggests a sort of personal revulsion against Wyman’s position. Wyman’s is a “bloated universe,” a “slum,” with a “rank luxuriance”; it “offends the aesthetic sense of us who have a taste for desert landscapes” (pp. e3-e4). On p. e2, he observes that Plato’s beard historically “has proved tough, frequently dulling the edge of Occam’s razor.” And Quine’s other negative remarks, especially the last about “desert landscapes,” seem also an allusion to Occam’s razor. But Occam’s razor might be wielded to explain Quine’s prejudice in a positive way. Occam’s razor is a principle according to which, “entities should not be multiplied beyond necessity”—we should only accept that there *is* some entity when it is “necessary” that we do so. If this is right, a “desert landscape” position according to which there are relatively few objects should be our natural starting position, and one that we would resist abandoning except on pain of “necessity.”<sup>10</sup>

Unfortunately, Occam’s razor is a notoriously slippery principle. First, one application of the razor may compete with another. In a case from recent philosophy, David Lewis has claimed that there are non-actual concrete worlds or universes distinct from ours, where these are in different space-times. Lewis thinks his worlds can “take the place” of certain abstract objects, and that an advantage of his position is that all of the objects he admits there are are concrete. According to Occam’s razor we should be parsimonious about admitting that there are various objects. But, Lewis says,

---

<sup>10</sup> The principle was propounded by William of Occam, a philosopher and theologian (c. 1300) who used it against Platonists of his own day.

Distinguish two kinds of parsimony... qualitative and quantitative. A doctrine is qualitatively parsimonious if it keeps down the number of fundamentally different *kinds* of entity: if it posits sets alone rather than sets and unreduced numbers, or particles alone rather than particles and fields, or bodies alone or spirits alone rather than both bodies and spirits. A doctrine is quantitatively parsimonious if it keeps down the number of instances of the kinds it posits; if it posits  $10^{29}$  electrons rather than  $10^{37}$ , or spirits only for people rather than spirits for all animals. I subscribe to the general view that qualitative parsimony is good in a philosophical or empirical hypothesis; but I recognize no presumption whatever in favor of quantitative parsimony. My realism about possible worlds is merely quantitatively, not qualitatively, unparsimonious. You believe in our actual world already. I ask you to believe in more things of that kind, not in things of some new kind.<sup>11</sup>

Still, other philosophers have thought it is a disadvantage to suppose that there are so many concrete universes—better to reduce the number of concrete objects and let in the abstract! It's not clear how Occam's razor works in this situation. Further, Occam's razor may come into conflict with other important considerations. So, e.g., we can imagine a scientist faced with a choice between a complex theory according to which there are relatively few entities, and a simple one according to which there are relatively many. Is theoretical pressure from simplicity the sort of "necessity" that legitimates admitting that there is some entity? As it turns out, it is far from clear what this "necessity" is supposed to be, and it is very difficult to formulate a generally applicable version of the principle.

Despite these difficulties, there does seem to be something right about Occam's razor. Maybe it is this: Something like Occam's razor is a consequence of very natural principles about *justification*. It is natural to think that one isn't justified in accepting *any* claim without some positive evidence for it, and so isn't justified in accepting that there are some entities without positive reasons for thinking that they are. Given this, the "desert-landscape" position is at least an appropriate starting point (or prejudice). We should be moved to include plants and trees—or whatever—only by the force of strong evidence. If we begin with an "empty canvas" on which our justified beliefs about plants and trees are to be painted, the desert-landscape position becomes the effective "starting point" for justification—we begin with no beliefs about what there is, and add beliefs only with sufficient justification. One might object that these considerations do not at all justify a claim that the world *is* a desert landscape—for all we have said about justification, the world might *be* a jungle populated by leprechauns, numbers and the like. But, on the current account of Occam's razor, we're not saying that we should believe that the entities *don't* exist; rather we're saying just that we shouldn't believe that they *do*. Thus, e.g., I have no reason to believe that you aren't wearing Power-Ranger underwear; surely, however, I also have no justification for thinking that you are.

---

<sup>11</sup> Lewis, *Counterfactuals* (Cambridge: Harvard University Press, 1973), 87.

But there may be a way to get more than this from Occam's razor. Suppose I offer as a *complete* theory of the world a view on which there are no entities of such-and-such sort. The theory is justified by its ability to explain various phenomena just like any other. The theory's *completeness* is justified by its ability to explain the *full* range of phenomena. Then, to the extent that my theory counts as *best*, reasons to accept it might seem to be reasons to accept that there are no entities of the specified sort. Thus, e.g., suppose my car doesn't run, and consider the theory that it doesn't run because it is out of gas. To the extent that this theory is *best*, it seems to trump the view that there are leprechauns under the hood gumming up the works. In this case, we get a real negative result, insofar as the negative hypothesis makes a corresponding contribution to explanatory power. If this is what Occam's razor comes to, it cuts closer than before. All the same, we don't have immediate or unconditional reason to favor, say, Lewis's theory over those of his opponents. Rather, the point is just that we should *deny* the existence of whichever entities are not allowed by the theory that is the best overall explanation. Perhaps some have wanted more than this from Occam's razor; and there may be concerns about the different rationales.<sup>12</sup> I don't know how much should be read into Quine. But there is at least this much to be said in favor of his preference.

(C) Second, Quine develops an objection related to criteria of identity. He thinks that Wyman cannot provide criteria of identity for unactualized possibles and that it is therefore implausible to think that there are any unactualized possibles. Any thing is identical to itself and distinct from other things; so, Quine reasons, if there are no criteria of identity and distinctness for unactualized possibles, there are no unactualized possibles.

Take, for instance, the possible fat man in that doorway; and, again, the possible bald man in that doorway. Are they the same possible man, or two possible men? How do we decide? How many possible men are there in that doorway? Are there more possible thin ones than fat ones? How many of them are alike? Or would their being alike make them one? Are no *two* possible things alike? Is this the same as saying that it is impossible for two things to be alike? Or, finally, is the concept of identity simply inapplicable to unactualized possibles? But what sense can be found in talking of entities which cannot meaningfully be said to be identical with themselves and distinct from one another? These elements are well-nigh incorrigible (e4).

Quine thinks any plausible view according to which there are some things must have answers to such questions—or, at least, that an inability to answer them is a theoretical weakness in Wyman's position. I agree that there is a theoretical difficulty for Wyman.

---

<sup>12</sup> Suppose my theory isn't that there are leprechauns under the hood gumming up the works, but rather that they merely laugh and dance because of my trials. In this case, it isn't obvious how the gas theory is relevant, and we seem forced back to a larger "complete" theory, without room for the dancing leprechauns. The explanatory rationale seems to work, if anyplace, where theories compete, but not where they don't.

But I don't think it is about identity and distinctness. Rather, it is about reference. We'll see how this goes.

Let's begin by taking up some obscurities in Quine's presentation. First, why should Wyman agree that there is a possible fat man in the doorway? Unactualized possibles are supposed to be abstract. So they are outside of space-time, and therefore not candidates for being in the doorway. Of course, Wyman will agree that

- (4) Possibly there is a fat man in the doorway.

But this is not enough for Quine's point. In the very next paragraph, Quine allows (at least for the sake of argument) that it is legitimate to prefix 'possibly' to a statement like 'there is a fat man in the doorway'. His problem has not to do with (4), but rather with the supposed possible entities. Maybe Quine means to ask about "the unactualized possible being which could be a fat man in the doorway." But this does not work either. On principles he himself endorses later in the article, a phrase of the sort, "the  $\mathcal{P}$ " picks out an object if and only if there is one and only one object that is  $\mathcal{P}$ . We'll think about this in the next chapter. Whatever we think about it, though, it is plausible that a phrase of the sort, "the  $\mathcal{P}$ " does not pick out any object when there is more than one thing that is  $\mathcal{P}$ —consider 'the desk at CSUSB' in 'the desk at CSUSB has graffiti on it.' Surely Wyman insists that there is more than one unactualized possible being that could be a fat man in the doorway, so that "the unactualized possible being that could be a fat man in the doorway" doesn't refer. Insofar as Quine's description fails to pick out a single unactualized possible being, questions of identity and distinctness do not arise. Though there may be problems about this (see below), we do better with questions about Pegasus or Zeus. Given the "beard" argument, Wyman, at least, must allow that such terms refer.

Second, it's not entirely clear how Quine's argument is supposed to work. What he says is susceptible to different interpretations. He does evoke a feeling of mystery about Wyman's position. Perhaps, though, we can say that this feeling is evoked by reasoning as follows:

- 1) Any thing is such that there is a criterion of identity for it.
  - 2) There is no criterion of identity for unactualized possibles.
- 
- 3) There are no unactualized possibles.

If this is right, Quine's rhetorical questions mainly support (2). But different approaches to "criteria of identity" result in different accounts of this reasoning. First, one might think that it's a trivial or primitive fact that any thing is identical to itself and distinct from other things, so that no account is required. But Quine seems to require some account of identity facts in terms of other properties—thus, e.g., perhaps it's the way their

parts are bonded that distinguishes one brick from another.<sup>13</sup> In addition, a criterion of identity may involve facts that make a thing at one time the same as a thing at another time—thus, e.g. perhaps some physical continuity makes the brick here now the same as the one there then. Say we allow that things must satisfy identity criteria of this sort. Still, there are different ways to understand what Quine requires: One might think that a thing has a criterion of identity if there is some *fact* which makes it identical to itself and distinct from other things—where such a fact might obtain though nobody is in a position to say what it is. But one might think a thing has a criterion of identity only if there is some way to *say* what makes it identical to itself and distinct from other things—where we might be able to say what makes a thing identical to itself and distinct from other things without being in a position to tell which things are which in particular cases. And a thing might have a criterion of identity only if there is some way to *tell* whether it is identical to itself and distinct from other things in particular cases. Thus there are (at least) these three interpretations for the argument:

1a) Any thing is such that there are facts that make it identical to itself and distinct from others.	1b) Any thing is such that there is some way to say what makes it identical to itself and distinct from others.	1c) Any thing is such that there is a way to tell if it is identical to itself and distinct from others.
2a) There are no facts that make one unactualized possible identical to itself and distinct from others.	2b) There is no way to say what makes one unactualized possible identical to itself and distinct from others.	2c) There is no way to tell if one unactualized possible is identical to itself and distinct from others.
3) There are no unactualized possibles	3) There are no unactualized possibles	3) There are no unactualized possibles

It's not obvious which, if any, Quine would accept. Unfortunately, however, what he says isn't adequate to demonstrate that any of the arguments is sound. Let me explain.

The problem with the third is that (1c) is false. Consider two qualitatively identical marbles—two marbles which cannot be distinguished by color, weight, or any other generally applicable features—which are placed into some opaque and impregnable container where one is annihilated and the other returned. Say there is no way to tell which marble is returned. Still, the marble that is returned is identical to one of the marbles placed into the container and distinct from the other—maybe it is identical to the one I placed in the container with my left hand, rather than the one I placed in the container with my right. That there is no way to *tell* which it is, does not show that it isn't one of them—that there are no facts about material continuity or whatever that make it identical to itself and distinct from the other. And that there is no way to tell

---

<sup>13</sup> But this is far too simple for a general account. See, e.g., Peter Van Inwagen, *Material Beings* (Ithaca: Cornell Univ. Press, 1990): 56-60 and discussion in a later chapters.



which it is, certainly gives no reason for thinking that it does not exist! Thus (1c) seems implausible.

(1b) may itself be understood in different ways. Plausibly, any thing is such that God, at least, could say what makes it identical to itself and distinct from others. But Quine gives us no reason for thinking that God isn't in a position to do this! At best, his rhetorical questions show that we, or maybe Wyman, do not know what to say.<sup>14</sup> But if (1b) says that *we* can say, for any thing, what makes it identical to itself and distinct from others, then it seems false. A little thought about electrons, angels, minds, and even persons shows that giving criteria of identity is no simple matter. To take a much-discussed example, suppose you are "beamed" to another planet—your body is dissolved and relevant information is beamed to the other planet but, due to mechanical failure, multiple qualitatively identical copies of you are produced. Which, if any, *is* you? Different answers are possible. But it doesn't matter. Maybe Wyman attempts to say what makes unactualized possibles identical to themselves and distinct from others. But surely he is within his rights to insist that his inability to describe a thing is *irrelevant* to whether the thing is—for the realm of unactualized possibles is supposed to be without respect to what anyone thinks about it.

Thus, the argument, if it is to work at all, must be the first. And (1a) is plausibly a necessary truth. But I don't think Quine's questions successfully motivate (2a). Surely Wyman insists that Zeus is distinct from Pegasus. Why? Well, they have different properties. Zeus "represents" a burley fellow with thunderbolts in his hand. Pegasus "represents" a flying horse. But Quine's questions become more difficult for Wyman when we consider Zeus and his (unjustly ignored) twin brother. To make matters concrete, suppose two burley-looking fellows with thunderbolts suddenly appear in the doorway before you. Plausibly, there *is* no answer to the question, "Is the one on the left Zeus?" Even so, it's not obvious that this is a problem about identity. That is, the argument,

There is no answer to the question, "Is the one on the left Zeus?"—and similarly for other unactualized possibles.

There are no facts that make one unactualized possible identical to itself and distinct from others.

is invalid. It is possible for the premise to be true and the conclusion false—or so I claim. Here's the basic idea: We may admit that there is no answer to the question "Is the one on the left Zeus?" if there are no facts that make one unactualized possible identical to itself and distinct from others. Perhaps there is no answer to the question whether *this* cloud is the same as *that* one, precisely because of problems about conditions for cloud identity. But there may be *other* reasons why there is no answer

---

<sup>14</sup> As it happens, some philosophers propose answers to Quine's questions. See, e.g., note 9.

to such a question. In particular, there may be no answer, not because of problems about identity and distinctness, but because of problems about *reference*.

To see this, let's begin with an example. Suppose conditions for Kierkegaard's personal identity are unproblematic. Insofar as there are things to which 'Kierkegaard' and 'Climacus' refer, 'Kierkegaard is identical to Climacus' makes perfect sense. Similarly, since both refer, 'Is Kierkegaard identical to Climacus?' has an answer—as it turns out, he is.<sup>15</sup> But consider, 'Kierkegaard is identical to the Dane'. There is no one thing to which 'the Dane' refers. Thus it would seem that this expression doesn't assert the identity of Kierkegaard with any particular thing. Similarly, if we ask whether Kierkegaard is identical to the Dane, there seems to be no answer, just because no determinate question is asked. So there may be no answer to the question whether  $\alpha$  is identical to  $\beta$ , precisely because there is no particular thing to which ' $\alpha$ ' or ' $\beta$ ' refers.

Now imagine, for the moment, that unactualized possibles are like souls, on the model of reincarnation. Suppose, further, that some souls have always been in heaven, never entering the cycle of life. Presumably, the advocate of such a view will insist that there are facts about identity and distinctness for souls and, in particular, that there is some fact about whether T. Roy is identical to Julius Caesar. I would like to think that he is. Granted, there is no way to tell. As with the marbles, though, there will be some facts (of soul-continuity or whatever) that make it the case either that he is identical to Julius Caesar or that he is not. In this case, insofar as Roy and Caesar have entered the cycle of life and we refer to them in the usual way, 'Roy' and 'Caesar' refer unproblematically, so that we ask a sensible question when we ask whether Roy is identical to Caesar. But now consider those burley-looking fellows with thunderbolts in the doorway before you. Is there an answer to the question about whether the one on the left is Zeus? I think not—because I think 'Zeus' does not refer.

This is not the place to develop a theory of reference. But perhaps we can say enough to see the problem. It is natural to think that it is possible to refer to a thing by describing it. Unfortunately, general descriptions, by themselves, won't do the job here. The case is set up so that ordinary descriptions of Zeus apply equally to the fellow on the left, and the one on the right. But there may be ways to refer where descriptions fail; perhaps some *particular* spatio-temporal or causal relations make reference possible in such cases. So, e.g., maybe this is a world of two-way eternal recurrence—so that all events simply repeat in successive ages and every qualitative description of me applies equally to individuals in all the infinitely many ages. Even so, when you say 'Roy' you refer to *me*—that is, to the person who stands in various particular spatio-temporal and causal relations to *you*. So far, perhaps, so good. But there don't seem to be *these* sorts of relations with unactualized possibles—precisely because they are unactualized. Thus, in contrast to the Caesar-Roy case, it is not

---

<sup>15</sup> 'Johannes Climacus' is one of the many pseudonyms of the Danish philosopher and early existentialist, Søren Kierkegaard.

obvious that we have the resources to distinguish, and so to refer to, the unactualized possible that is Zeus: Insofar as Zeus has never entered the cycle of life, it's not clear that we have ever fixed on a particular "soul" for talk about him. And if we have not fixed on any particular soul, there is no point asking about whether that particular fellow is Zeus or not.

As Quine suggests, Wyman might reply that there aren't any qualitatively identical unactualized possible beings.<sup>16</sup> Then descriptions alone might be sufficient to do the job. In itself, this is surprising: why *can't* we talk about Zeus and his (qualitatively identical) twin? And why wouldn't the "beard" argument yield the result that both exist? But grant the point, and suppose the fellows in the doorway *aren't* qualitatively identical—the one has a mole on his toe that the other does not. Still, it's hardly clear that there is an answer to the question, "Is the one on the left Zeus?" For *our* descriptions of Zeus are satisfied equally by the one on the left and the one on the right. That is, we may reason,

Terms like 'Zeus' refer to unactualized possibles only if associated descriptions are sufficient to pick out unique unactualized possibles.

Descriptions associated with terms like 'Zeus' aren't sufficient to pick out unique unactualized possibles.

---

Terms like 'Zeus' do not refer to unactualized possibles.

Given this, as above, there may be no answer to the question whether the one on the left is Zeus. Notice that this reasoning doesn't depend on difficulties about identity or distinctness for unactualized possibles. In our "reincarnation" model, we have simply assumed that there *are* facts about identity and distinctness—and have still been able to explain why there might be no answer to the question, "Is the one on the left Zeus?" To the extent that the feeling of "strangeness" Quine evokes with his questions is adequately diagnosed as a problem about reference, Quine's ground for saying there *are* no unactualized possibles is removed.

But this result is, at best, cold comfort to Wyman. If terms like 'Zeus' do not refer to unactualized possibles, Wyman's use of the "beard" argument is itself in trouble. For Wyman supplies unactualized possibles as objects of *reference* for terms like 'Pegasus' and 'Zeus'. But when one uses the beard argument to conclude that there are some things to which 'Pegasus' and 'Zeus' refer, the things one supplies must be things to which the terms *can* refer! Thus, like McX, Wyman seems to reach the wrong conclusion from his beard argument. Perhaps, though, the upshot is a tie. Wyman's positive argument for the conclusion that Pegasus exists as an unactualized

---

<sup>16</sup> He suggests that Wyman's view might be rehabilitated by a "Fregean therapy of individual concepts." An important component of this therapy is that things are regarded as equivalent just in case uniquely identifying descriptions are necessarily equivalent.

is on the rocks. But we have found no support from problems about identity and distinctness for the conclusion that there aren't any unactualized possibles—it may be only that there are problems about referring to particular ones. Still, Quine might count the round as a win: if we agree that a “desert landscape” is the appropriate starting point, then Wyman *needs* a positive argument for the conclusion that there are unactualized possibles. But his argument is slipping away.<sup>17</sup>

(D) Quine's final objection centers on “the round square cupola on Berkeley College.”<sup>18</sup> Let's set worries about reference to the side. Presumably, there *is* no round square cupola on Berkeley College. Further, there *couldn't* be one. It is impossible for anything to be both round and square (in the same sense at the same time). So, though the round square cupola is indeed unactualized, it isn't an unactualized *possible*. *Prima facie* this is a problem for Wyman. For we would expect Wyman to argue that ‘the round square cupola on Berkeley College is not’ contradicts itself, and so to conclude that the round square cupola on Berkeley College is. Will Wyman therefore claim that it is an unactualized *impossible*? This seems even more absurd than before. As Quine suggests, “a good many embarrassing questions could be asked about [impossibles]. We might hope even to trap Wyman in contradictions, by getting him to admit that certain of these entities are at once round and square” (e4). If Wyman's view leads to the conclusion that there are round squares, then it would seem to be wrong.

But it's hardly obvious that Wyman's view commits him to the conclusion that there are round squares. Clearly, unactualized impossibles don't exist. So no such impossible is actually on Berkeley College. And, since it isn't possible, we can't say that there is an unactualized impossible which *could* be on Berkeley College. But there might be an unactualized impossible which “represents” itself as a round square cupola on Berkeley College. Call this entity ‘ $\alpha$ ’. Wyman has been committed to a predicate ambiguity all along. Thus,

- (5)  $\alpha$  is a round square cupola on Berkeley College.

needn't say that  $\alpha$  is itself round and square. Rather, it may be elliptical for,

- (6)  $\alpha$  “represents” a round square cupola on Berkeley College.

And there may be a way to understand this that doesn't require  $\alpha$ 's being both round and square. Indeed, insofar as  $\alpha$  is abstract, it shouldn't have a shape at all! Now, this doesn't obviously remove all difficulties. Our previous (p. 31) suggestion was that

---

<sup>17</sup> Wyman might beat the worry if there *is* some way to interact with unactualized possibles. So, e.g., if we follow Plato and allow that we interact with Zeus and/or his twin before our birth, then we might hold that some past relations with unactualized possibles ground reference to them in the usual way.

<sup>18</sup> Just in case: a cupola is an architectural feature—a small dome on top of a structure.

abstract individuals “represent” or “build in” certain features they *would* have *if* they were actual. And it is not clear what this means for individuals that *cannot* exist. Still, it seems possible to represent something as round and square, as,

- (7) The cupola on Berkeley College is both round and square.

represents that the cupola on Berkeley College is round and square. Though (7) represents something which cannot be, the existence of the *statement* is uncontroversial. And, similarly, Wyman might hold that  $\alpha$  unproblematically is, though  $\alpha$  could never exist.

On Quine’s account, however, this isn’t Wyman’s response. Rather Wyman rejects the proposal that there are unactualized impossibles. He does this on the ground that ‘round square cupola’ is meaningless because contradictory. If ‘the round square cupola does not exist’ is meaningless, then pressure from **T1** to accept that there is a round square cupola is removed. Nothing follows from “smella bella dupe,” precisely because it doesn’t mean anything.

But Quine thinks this strategy fails. He rejects the doctrine that whatever is contradictory is meaningless. In support of this conclusion, he raises two considerations. I am not impressed by the second: Quine observes that there is, in mathematical logic, no generally applicable test of contradictoriness. True enough. Quine seems to think that, if, with Wyman, we agree that whatever is contradictory is meaningless, there is therefore some problem. What problem? Perhaps we are supposed to accept that there *is* a generally applicable test of meaningfulness, so that contradictoriness and meaningfulness come apart. And if they come apart, we can’t move directly from contradiction to lack of meaning. Perhaps, then, the argument is this:

- 1) If Wyman is right, a well-formed statement in mathematical logic is contradictory if and only if it is meaningless.
- 2) There is no generally applicable test for contradictory statements of mathematical logic.

---

- 3) If Wyman is right, there is no generally applicable test for the meaningfulness of statements of mathematical logic.
- 4) There is a generally applicable test for the meaningfulness of statements of mathematical logic

---

- 5) Wyman isn’t right.

(1) isn’t a mere restatement of Wyman’s view. Wyman, we are assuming, asserts that *if* a statement is contradictory, then it is meaningless. But it would be absurd for him to admit that everything meaningless is contradictory. So, e.g., “smella bella dupe,” isn’t contradictory precisely because it doesn’t mean anything. But one might think

that well-formed (grammatical) statements of mathematical logic couldn't fall prey to this difficulty—and therefore that the only way a well-formed statement of symbolic logic could turn out to be meaningless is if it turned out to be contradictory. Given this, on the view being attributed to Wyman, (1) may seem plausible. (2) is just the result from mathematical logic. (3) follows from (1) and (2). (4) is a premise which may seem obvious. And (5) follows from (3) and (4). But Wyman might deny (4). Why think that there is or should be a generally applicable test for meaning? Perhaps the idea is that we can ordinarily “spot” meaning when it is there, and that this ability shows that there is some test sufficient to identify it. But maybe contradiction and meaningfulness are on a par in that we are often, but not *always* able to spot them. There is no result in mathematical logic according to which we can *never* identify contradiction! So Wyman might simply deny (4).

However, Quine's first point is more impressive. Consider a meaningless string of symbols, say the one mentioned above. It is implausible to suggest that such a string is the sort of thing that could be true or false or *follow* from anything. But if contradictions are meaningless, this jeopardizes a reasoning pattern known as *reductio ad absurdum*. The idea of any such argument is to show that some supposition leads to contradiction; but if a supposition leads to contradiction, then that supposition must be false; so the negation of the supposition is true. We have used this reasoning pattern many times. Here is an example (using the language of existence):

Suppose Pegasus does not exist; then, by **T1**, there exists an object corresponding to ‘Pegasus’ to which ‘does not exist’ applies; so there exists an object which does not exist; but this is impossible: the supposition that Pegasus does not exist is false, and Pegasus therefore does exist.

The *reductio* pattern requires that a contradiction follow from some supposition. But Wyman thinks contradictions are meaningless; so how can he make sense of a contradiction's *following* from something? If he can't, then he jeopardizes *reductio ad absurdum*, and his own positive argument along with it. Of course, again, Wyman might avoid these objections by avoiding the response Quine gives him, according to which contradictions are meaningless. So Quine's points about the round square cupola don't, by themselves, show that the postulation of unactualized possibles is absurd.

Let's sum up. Quine doesn't attempt to show that there are no ideas. And he doesn't generate any fatal blow against unactualized possibles. The discussion of identity and distinctness demonstrates what may be merely a problem about reference, rather than a problem about existence. And similarly, we have not been forced to admit that unactualized impossibles are themselves impossible—they may only represent what isn't possible. But Quine does land some blows against *motivation* for the views of McX and Wyman. For any supposed entities, it's natural to demand some justification before admitting that they exist. And the “beard” argument is in shambles. **T1** suggests that terms like ‘Pegasus’ and ‘Parthenon’ must refer to objects

in order for statements containing them to be significant. But, in discussion of McX, it turns out that the objects aren't always to be found in the mental and concrete realms. And, from discussion of Wyman, it looks like, as objects of reference, they aren't in the abstract realm either.

Thus, maybe we should give up looking for such objects; maybe we should look for some way to make sense of statements containing 'Pegasus' and 'Parthenon' which doesn't require that there are objects to which they refer. Of course, this would require rejecting **T1**. Further, as suggested at the beginning of the chapter, given its "bias" toward inflated ontologies, **T1** seems inappropriate as a vehicle for regimentation. At least, one would prefer some way of understanding statements containing 'Pegasus' and 'Parthenon' which doesn't itself *require* that there be objects to which they refer—maybe there are such objects and maybe not; the point is only that we should be able to argue about them in a neutral way. Developing such a mode of regimentation is the project of the next chapter.

## Chapter Three

### Russell's Way Out: A Plausible Canonical Notation

Given some sentences whose truth conditions are controversial, say some sentences of fiction or arithmetic, our idea is to use regimented expressions to say, literally, what makes them true or false. There may be controversy about whether some proposed truth condition is correct. But using the regimented sentences to state the condition should remove confusion and controversy about what conditions are proposed. Ideally, the expressive power of the regimented sentences should be sufficient to express arbitrary truth conditions. And we don't want regimented expressions to entangle us in Plato's beard. In this chapter, we'll develop a regimented system of expressions, or a "canonical notation," and argue that it escapes Plato's beard. In the next couple of chapters, we'll take up questions *about* the notation – including the question of whether it is sufficient to express arbitrary truth conditions. As it turns out, thinking about our notation will itself require consideration of interesting and important metaphysical questions. In this chapter, though, the focus is on introducing the notation and on Plato's beard. After some words of orientation, we'll introduce the notation, and see how the notation gets us out of Plato's beard.

#### *I. Orientation*

In this short section, I'll make some points that should aid in consideration of sections that follow. First, it's worth saying a bit more about what regimented expressions are supposed to do. We know they are to help say what makes ordinary claims true or false. But what is it to do this? (i) One might think that regimented expressions somehow expose the "inner workings" of ordinary language – so that once one has the regimented expressions, one has an account of how the ordinary ones work, and therefore an account of what makes them true. But (ii) one might think that regimented expressions merely "translate" or "paraphrase" ordinary ones. Quine seems to accept something along these lines. Again, though, there are different ways to understand the proposal. Insofar as we are interested in conditions under which an expression is true, we will consider regimented expressions adequate when their truth conditions are the same as ordinary ones. We therefore leave questions about "meaning" and the like to philosophy language. Observe that, so long as truth conditions are the same, and an argument is *valid* just in case there is no condition under which premises are true and the conclusions not (if this is not familiar, see the appendix), consequences of the expressions should be the same – for validity hinges on nothing more than truth conditions. So a test of adequacy for regimentation is that consequences of expressions are retained.

Second, a point related to the one about meaning: It is helpful to correlate expressions of the canonical notation with expressions of English. However, we do not thereby fix the significance of canonical expressions. So, e.g., a standard course in a foreign language is likely to present lists of the following sort,



<i>chou:</i>	cabbage
<i>petit:</i>	small
:	

where words of one language are said to correspond to words of another. But it is not a legitimate criticism of a Frenchman who refers to his sweetheart as “*mon petit chou*” to observe that she is no cabbage. French is not *defined* in a textbook relative to English. Rather, French has some conventions, such that sometimes use of ‘*chou*’ corresponds to use of ‘cabbage’ and sometimes it does not; the text *uses* this correspondence to teach the language. Similarly, it is helpful to introduce terms of the canonical notation as corresponding to some terms of English. But it will not be a legitimate criticism of the canonical notation to observe that this correspondence sometimes breaks down. Terms of the canonical notation have (stipulated) conventions of their own, and may or may not uniformly correspond to any term or terms of English.

Finally, our canonical notation, or something very much like it, is extensively studied in courses of symbolic logic. In such courses, the emphasis is on what follows from what. It turns out that there are powerful methods for addressing such questions.<sup>1</sup> But our objectives are more modest. We want to understand what makes the regimented expressions true, and to use them to state truth conditions. Our task overlaps enough with that undertaken in a course of symbolic logic that familiarity with symbolic logic may facilitate moving through this chapter, and familiarity with this chapter may ease the way into symbolic logic. But no exposure to logic is assumed here, and symbolic logic certainly requires no familiarity with this chapter. The introduction of some symbols may make the material of this chapter seem intimidating. But this chapter is, I think, the easiest in the text. Some definitions are introduced and manipulated. That is all. It will take work to become familiar with the various definitions and to use them, but there is no controversy or deep question about the nature of the world that is being addressed. So we get a break from “mind-bending” questions, and from working through detailed objections and replies.<sup>2</sup>

## II. The Canonical Notation

On the one hand, we can see our notation as simplifying or abbreviating expressions of ordinary language. We shall be able to get an initial fix on the notation in this way. But on the other hand, both the grammar and truth conditions for expressions of the notation are capable of very precise and independent expression. We

---

<sup>1</sup> See, e.g., Roy *Symbolic Logic: An Accessible Introduction to Serious Mathematical Logic* (unpublished but available at <http://philosophy.csusb.edu/~troy/int-ml.htm> ) or Bergmann, Moor and Nelson, *The Logic Book* 3rd ed. (New York: McGraw-Hill, 1998).

<sup>2</sup> The substance of the next section should not be new, if your background includes formal predicate logic with identity. Even so, it will be helpful to work through the material, to grasp the simplified and idiosyncratic way it is developed here.

shall be able to deepen our understanding by this means – and such understanding will be important as we move forward to applications.

(A) First, consider some simple subject-predicate sentences as from chapter 1 (p. 5). Thus we have both ‘Bill is happy’ and ‘Hillary is happy’. Such sentences may combine to form ones that are more complex. Thus, ‘Bill is happy *and* Hillary is happy’. We shall find it convenient to abbreviate this, ‘Bill is happy  $\wedge$  Hillary is happy’ with operator ‘ $\wedge$ ’. There are five operators of this sort,

$\sim$ Bill is happy	<i>It is not the case that</i> Bill is happy
Bill is happy $\wedge$ Hillary is happy	Bill is happy <i>and</i> Hillary is happy
Bill is happy $\vee$ Hillary is happy	Bill is happy <i>or</i> Hillary is happy
Bill is happy $\rightarrow$ Hillary is happy	<i>If</i> Bill is happy <i>then</i> Hillary is happy
Bill is happy $\leftrightarrow$ Hillary is happy	Bill is happy <i>iff</i> Hillary is happy

These may be combined in obvious ways, so that ‘Bill is happy  $\vee \sim$ Hillary is happy’ says Bill is happy or Hillary is not. Sometimes we will find parentheses convenient, so that ‘(Bill is happy  $\wedge$  Hillary is happy)  $\vee$  Hillary is president’ says Bill and Hillary are both happy *or* Hillary is president, where ‘Bill is happy  $\wedge$  (Hillary is happy  $\vee$  Hillary is president)’ says Bill is happy *and* Hillary is happy or president. You should be clear about the difference.

In addition, we will typically take *predicate* terms of the sort encountered in chapter 1, and represent them by capital letters; thus we may use  $Hx$  for ( $x$  is happy) and  $Px$  for ( $x$  is president). And we shall read ‘ $\forall xHx$ ’ to say for any thing  $x$  *it* is happy – so as to say that *everything* is happy. (The upside-down ‘A’ for ‘all’ is the *universal quantifier*). And we shall read ‘ $\exists xHx$ ’ to say there exists a thing  $x$  such that *it* is happy – so as to say that *something* is happy. (The backwards ‘E’ for ‘exists’ is the *existential quantifier*). As indicated by the official readings, the variable ‘ $x$ ’ here works very much like a pronoun in ordinary language. And it is possible to use different variables to the same effect. Thus ‘ $\forall yHy$ ’ says that any thing  $y$  is such that it is happy, and ‘ $\exists yHy$ ’ that there exists a thing  $y$  such that it is happy. The important point is that the variable or pronoun like the use of the quantifier to that of the predicate.

And, of course, our notions may be combined. Thus, ‘ $\exists xHx \wedge \exists xPx$ ’ says something is happy and something is president. ‘ $\exists x(Hx \wedge Px)$ ’ says something is both happy and president. These are not the same! The first might be the case so long as Hillary is happy and Bill is president. For the second, we would need some one thing, say Bill, that is both happy and president. You should try to see this given our official readings. But if this is not yet clear, do not despair. We will be able to be much more clear about such cases – and ones much more complex – once we have been more precise about how the notation works.

(B) A person can learn to read aloud in, say, Spanish or Hebrew without any understanding of what is being read. And one can imagine a child able to identify errors of grammar or punctuation before being able to read (maybe she catches missed capitals at the beginning of sentences). Similarly, in this section, we learn to identify and generate grammatical sentences of our canonical notation  $\mathfrak{U}$ , without direct thought about their meaning. It is possible to give so simple and concise a statement of the grammar of  $\mathfrak{U}$ , that a person may become a perfect grammarian independently of understanding of what is written.

We begin by introducing the *vocabulary* or symbols of  $\mathfrak{U}$ , and then turn to the way they are put together. The *vocabulary* for our canonical notation  $\mathfrak{U}$  consists of: (1) Predicate letters: for any  $n \geq 1$ , uppercase Roman letters with superscript  $n$ . Thus  $A^1$ ,  $B^1 \dots Z^1$ , are *one-place* predicate letters,  $A^2$ ,  $B^2 \dots$  are *two-place* predicate letters, etc. In addition, to be sure that we never run out of predicate letters of a given type, we allow also integer *subscripts*, in effect to generate new letters. Thus we would treat ' $A_1$ ' as a letter different from ' $A$ ', and ' $A_1^1$ ' as a one-place predicate letter different from  $A^1$ . (In practice, we will almost never need recourse to so many letters.) (2) Variables: lowercase Roman 'i' through 'z'. Again, to be sure that we never run out of variables, we allow integer subscripts so that ' $x$ ', ' $x_1$ ', ' $x_2$ ' etc. would all be different variables – though, again, we shall hardly ever need recourse to the device of subscripts. (3) Truth functional operators, tilde, caret, wedge, arrow, double arrow:  $\sim$ ,  $\wedge$ ,  $\vee$ ,  $\rightarrow$ ,  $\leftrightarrow$ . We have seen these before, and now give them names. (4) Quantifier symbols, universal and existential:  $\forall$ ,  $\exists$ . And, finally, the vocabulary includes, (5) punctuation symbols: opening and closing parentheses: (, ). And that is all. Insofar as these are all the symbols of the language, any expression which does not consist of some combination of these symbols could not be a grammatical expression of  $\mathfrak{U}$ .

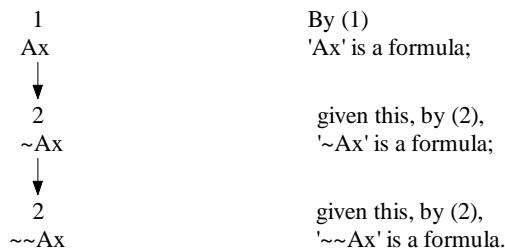
To proceed, we need some brief conventions for talking *about* expressions of  $\mathfrak{U}$ . Observe that the variables and predicate letters of  $\mathfrak{U}$  are lower- and upper-case Roman letters. We will let “curly” ' $\mathcal{P}$ ', ' $\mathcal{Q}$ '... ' $\mathcal{Z}$ ' represent arbitrary expressions of  $\mathfrak{U}$ , and lower-case italics, ' $w$ ', ' $x$ ', ' $y$ ', ' $z$ ' represent arbitrary variables. Truth functional operators, quantifier symbols, and punctuation symbols will stand for themselves. Insofar as these are symbols for symbols, they are “meta-symbols.” Concatenated or joined meta-symbols stand for the concatenation of that which they represent. Thus, e.g., any expression of the form  $\forall x$  is a universal  $x$ -quantifier, and any expression of the form  $\exists x$  is an existential  $x$ -quantifier. Here, ' $\forall x$ ' and ' $\exists x$ ' are not expressions of  $\mathfrak{U}$ . (Why?) Rather, we've said of expressions in  $\mathfrak{U}$  that ' $\forall x$ ' is a universal  $x$ -quantifier, ' $\forall y$ ' is a universal  $y$ -quantifier, ' $\exists x$ ' is an existential  $x$ -quantifier, ' $\exists y$ ' is an existential  $y$ -quantifier, etc. In the metalinguistic expression, ' $\forall$ ' and ' $\exists$ ' stand for themselves, and ' $x$ ' for the arbitrary variable. Thus, when we say something about expressions of the form  $\forall x$  we say something about a *range* of expressions, all at once. In the following, we *use* this method to say which expressions are formulas and sentences.

We are now ready to introduce the core notion of a *formula*. I'll give a complete statement, then turn to discussion and examples.

- (1) If  $\mathcal{P}$  is an  $n$ -place predicate letter followed by  $n$  individual variables, then  $\mathcal{P}$  is a *formula*.
- (2) If  $\mathcal{P}$  is a formula, then  $\sim\mathcal{P}$  is a *formula*.
- (3) If  $\mathcal{P}$  and  $\mathcal{Q}$  are formulas, then  $(\mathcal{P} \wedge \mathcal{Q})$ ,  $(\mathcal{P} \vee \mathcal{Q})$ ,  $(\mathcal{P} \rightarrow \mathcal{Q})$  and  $(\mathcal{P} \leftrightarrow \mathcal{Q})$  are *formulas*.
- (4) If  $x$  is a variable and  $\mathcal{P}$  is a formula with no  $x$ -quantifier then  $\forall x\mathcal{P}$  and  $\exists x\mathcal{P}$  are *formulas*.
- (5) Nothing is a formula unless it can be formed by repeated applications of (1) - (4).

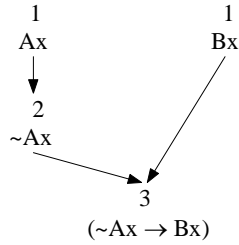
(1) tells us that ' $A^1x$ ', ' $B^2zw$ ' and ' $B^2zz$ ' are formulas. It does not allow that ' $A^1wz$ ' is a formula, because of the mismatch between the number of places and the number of variables; and it does not allow that ' $A^1$ ' is a formula, for the same reason. It is common to assume that predicate letters are followed by the correct number of variables, and to abbreviate by omission of superscripts. Let us do so as well. Given this, ' $Ax$ ' and ' $Axy$ ' involve *different* predicate letters, for ' $Ax$ ' abbreviates ' $A^1x$ ' and ' $Axy$ ' abbreviates ' $A^2xy$ '. We have seen this convention at work already in our initial intuitive presentation.

Clauses (2) through (4) tell us that one expression is a formula *given* that some other or others are formulas. So, e.g., by (1), ' $Ax$ ' is a formula; *given this*, by (2), ' $\sim Ax$ ' is a formula. (2) tells us that if *any* expression  $\mathcal{P}$  is a formula, then a tilde, followed by that expression, is a formula. And the process goes on! We have just seen that ' $\sim Ax$ ' is a formula; *given this*, by (2), ' $\sim\sim Ax$ ' is a formula. Etc. It is natural to represent this, and other cases, by a diagram which indicates how one rule application feeds into the next.



And the process could continue indefinitely.

Clause (3) works very much like (2), except that a given application of the rule requires *two* inputs. So, e.g., by (1), ' $Ax$ ' and ' $Bx$ ' are formulas. Given this, by (3), ' $(Ax \wedge Bx)$ ' is a formula. Again, the process goes on, and charts are helpful. For a case which combines rules (1) - (3),

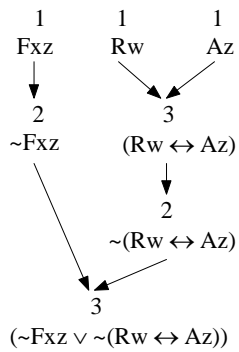


By (1),  
'Ax' and 'Bx' are formulas;

given this, by (2),  
'¬Ax' is a formula

given these, by (3),  
'(¬Ax → Bx)' is a formula.

Notice that different applications of (1) might produce the same formula; so, by a chart like the one above, but with 'Ax' uniformly replacing 'Bx', '(¬Ax → Ax)', is a formula. For a more complex case, let's use a chart to see that '(¬Fxz ∨ ¬(Rw ↔ Az))' is a formula:



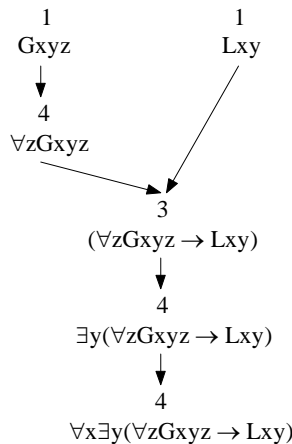
By (1),  
these are formulas;

given these, by (2) and (3),  
these are formulas;

given this, by (2),  
this is a formula;

given these, by (3),  
this is a formula.

Clause (4) is like (2) in that it requires only a single input. We thus say that  $\sim$  and the quantifiers are *unary* operators and  $\wedge$ ,  $\vee$ ,  $\rightarrow$ ,  $\leftrightarrow$  are *binary*. (4) differs from (2) insofar as it includes a constraint on the formulas that may legitimately count as inputs. By (1), 'Ax' is a formula; 'Ax' has an occurrence of 'x' and no x-quantifier so, by (4), ' $\exists xAx$ ' is a formula. But ' $\exists xAx$ ' has an x-quantifier; so we can't, by (4), move from this to the conclusion that ' $\forall x\exists xAx$ ' is a formula. Similarly, 'Ax' has no occurrence of 'y'; so we can't, by (4), move to the conclusion that ' $\forall yAx$ ' is a formula. To prefix a quantifier onto a formula, that formula must have an instance of the quantifier's variable, *and* not already have a quantifier of that variable-type. Again, charts are useful.



By (1),  
these are formulas;

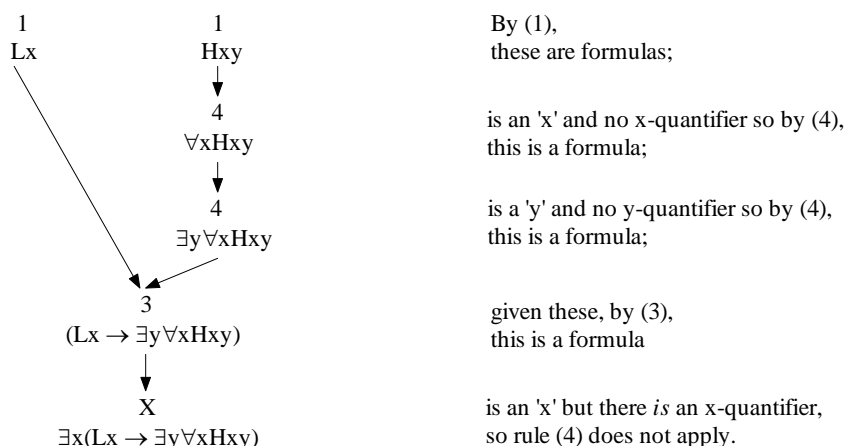
is a 'z' and no z-quantifier so by (4),  
this is a formula;

given these, by (3),  
this is a formula;

is a 'y' and no y-quantifier so by (4),  
this is a formula;

is an 'x' and no x-quantifier so by (4),  
this is a formula.

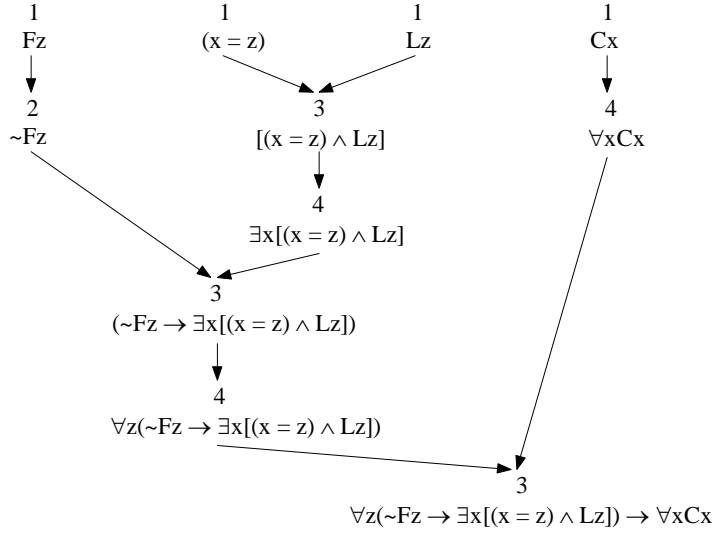
These charts *demonstrate* that a given expression is a formula. Clause (5) tells us that the only expressions that are formulas, are expressions that can be shown to be formulas this way. Thus, e.g., ‘ $(Ax)$ ’ is not a formula. Here’s one way to see it: clause (3) is the *only* clause that introduces parentheses; but (3) always introduces parentheses along with some binary operator; since ‘ $(Ax)$ ’ has parentheses but no binary operator, it could not be formed by (3), and so could not be formed by any rule; so, by (5), it is not a formula. Similarly, ‘ $(Ax \sim Bx)$ ’ is not a formula: ‘ $Ax$ ’ is a formula, and ‘ $\sim Bx$ ’ is a formula, but there is no way to put them together, by the rules, without a binary operator in between; so ‘ $(Ax \sim Bx)$ ’ is not a formula. If an expression is a formula, then there is some way to construct it by the rules on a chart; in general, a chart for an expression that is not a formula *must* break one of the rules. So, e.g., ‘ $\exists x(Lx \rightarrow \exists y \forall x Hxy)$ ’ may seem to be a formula. But, in its very last step, this chart breaks rule (4).



The only way to introduce the existential x-quantifier is by (4). Since the conditions for application of (4) are not met, ‘ $\exists x(Lx \rightarrow \exists y \forall x Hxy)$ ’ is not a formula. And similarly in other cases. Notice that we are in a position to deal with arbitrarily complex cases by our methods!

It will be convenient to have in hand a few more abbreviations. First, it is sometimes convenient to use a pair of square brackets ‘[ ]’ in place of a pair of parentheses ‘( )’. This is purely for visual convenience. E.g., ‘ $((()()))$ ’ may be easier to absorb as, ‘ $([()])$ ’. On a chart, it’s natural simply to introduce the square brackets, in an application of (3), in place of the parentheses. Second, if the *very last* step of some chart is justified by (3), then it is OK to take the last step with the *outermost* pair of parentheses (or brackets) dropped. Thus, e.g., ‘ $\forall x Fxy \rightarrow \sim Gz$ ’ abbreviates ‘ $(\forall x Fxy \rightarrow \sim Gz)$ ’; notice that it does not abbreviate ‘ $\forall x(Fxy \rightarrow \sim Gz)$ ’ – for we only drop parentheses associated with a *last* step which is justified by (3); the last step for ‘ $\forall x(Fxy \rightarrow \sim Gz)$ ’ is justified by (4), not (3), so we would not be allowed to drop these parentheses. Again, dropping parentheses is purely for visual convenience. Also, it will be convenient to have a fixed two-place predicate letter (say, ‘ $E_7^2$ ’) for the relation of equality. Supposing that we have fixed on a predicate letter  $\mathcal{P}$  for the relation of equality, we’ll capitulate to ordinary usage,

and abbreviate  $\mathcal{P}xy$  as  $(x = y)$ . In a chart, this abbreviation always appears in the top line, in application of rule (1). Putting all this together, the following (unannotated) chart shows that ' $\forall z(\sim Fz \rightarrow \exists x[(x = z) \wedge Lz]) \rightarrow \forall xCx$ ' is a(n abbreviation of a) formula.

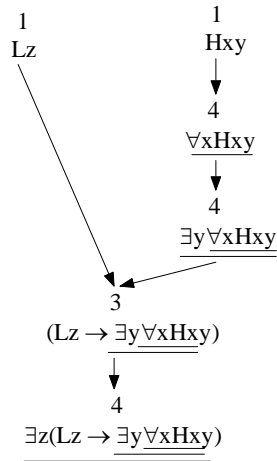


We drop superscripts on predicate letters, introduce the abbreviation for equality in the top row, use square brackets, and introduce the last arrow without parentheses. Where ' $E_7^2$ ' is the predicate letter for equality, ' $\forall z(\sim Fz \rightarrow \exists x[(x = z) \wedge Lz]) \rightarrow \forall xCx$ ' abbreviates the official expression, ' $(\forall z(\sim F^1z \rightarrow \exists x(E_7^2xz \wedge L^1z)) \rightarrow \forall xC^1x)$ '. In general, we won't distinguish between a formula and its abbreviations.

Our discussion of grammar wraps up with a few final definitions – including (finally) the definition of a *sentence*. The definitions all involve formulas, and are presented in relation to a formula's chart.

First, a formula's *main operator* is the *last* operator added in its chart. From the chart just above, ' $\rightarrow$ ' is the main operator of ' $\forall z(\sim Fz \rightarrow \exists x[(x = z) \wedge Lz]) \rightarrow \forall xCx$ '. Second, every formula in a chart for  $\mathcal{P}$  (including  $\mathcal{P}$  itself) is a *subformula* of  $\mathcal{P}$ . A subformula is *atomic* iff it appears in the top row, and so is justified by (1). From the above chart, the atomic subformulas of ' $\forall z(\sim Fz \rightarrow \exists x[(x = z) \wedge Lz]) \rightarrow \forall xCx$ ' are ' $Fz$ ', ' $(x = z)$ ', ' $Lz$ ' and ' $Cx$ '. An *immediate* subformula of  $\mathcal{Q}$  is a subformula to which  $\mathcal{Q}$  is directly connected by lines. From the chart above, ' $\forall z(\sim Fz \rightarrow \exists x[(x = z) \wedge Lz])$ ' and ' $\forall xCx$ ' are the immediate subformulas of ' $\forall z(\sim Fz \rightarrow \exists x[(x = z) \wedge Lz]) \rightarrow \forall xCx$ '. Of course, a formula with a unary main operator has just one immediate subformula. We sometimes speak of a formula by means of its main operator: a formula of the form  $\sim \mathcal{P}$  is a *negation*;  $\forall x\mathcal{P}$  is a *universal generalization*;  $\exists x\mathcal{P}$  is an *existential generalization*; a formula of the form  $(\mathcal{P} \vee \mathcal{Q})$  is a *disjunction* with  $\mathcal{P}$  and  $\mathcal{Q}$  as *disjuncts*; a formula of the form  $(\mathcal{P} \wedge \mathcal{Q})$  is a *conjunction* with  $\mathcal{P}$  and  $\mathcal{Q}$  as *conjuncts*; a formula of the form  $(\mathcal{P} \rightarrow \mathcal{Q})$  is a (*material*) *conditional* where  $\mathcal{P}$  is the *antecedent* of the conditional and  $\mathcal{Q}$  is the *consequent*; and a formula of the form  $(\mathcal{P} \leftrightarrow \mathcal{Q})$  is a (*material*) *biconditional*.

Moving now toward the definition of a *sentence*: If a formula includes a quantifier, that quantifier's *scope* includes just the subformula in which the quantifier *first* appears. Using underlines to indicate scope,

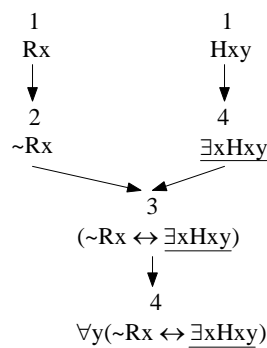


The scope of the x-quantifier is ' $\forall xHxy$ ';

the scope of the y-quantifier is ' $\exists y\forall xHxy$ ';

the scope for the z-quantifier is the entire formula.

A variable  $x$  is *bound* iff it appears in the scope of an  $x$ -quantifier, and a variable is *free* iff it is not bound. In the above chart, each variable is bound. The  $x$ -quantifier binds both instances of ' $x$ '; the  $y$ -quantifier binds both instances of ' $y$ '; and the  $z$ -quantifier binds both instances of ' $z$ '. In ' $\forall xFxy$ ', however, both instances of ' $x$ ' are bound, but the ' $y$ ' is free. Finally, an expression of  $\mathcal{U}$  is a *sentence* iff it is a formula and it has no free variables. To determine whether an expression is a sentence, use a chart to see if it is a formula. If it is a formula, use underlines to check whether any variable  $x$  falls outside the scope of an  $x$ -quantifier. If there is a chart, and no such variable, then the expression is a sentence. Thus, from the chart above, ' $\exists z(Lz \rightarrow \exists y\forall xLxy)$ ' is a sentence. From this chart,



The scope of the x-quantifier is ' $\exists xHxy$ ';

The scope of the y-quantifier is the entire formula.

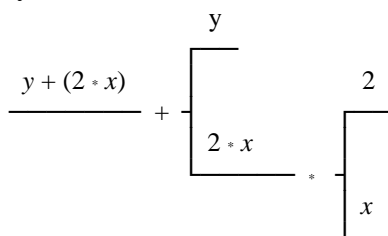
' $\forall y(\neg Rx \leftrightarrow \exists xHxy)$ ' is not. It has a chart so it is a formula. The  $x$ -quantifier binds the last two instances of ' $x$ ', and the  $y$ -quantifier binds both instances of ' $y$ '. But the first instance of ' $x$ ' is free. Since it has a free variable, ' $\forall y(\neg Rx \leftrightarrow \exists xHxy)$ ' is not a sentence. Insofar as it already has an  $x$ -quantifier, we cannot *prefix* another quantifier to bind the ' $x$ ' that is free. (But an  $x$ -quantifier could have been added to, say, ' $\neg Rx$ ', and ' $\forall y(\forall x\neg Rx \leftrightarrow \exists xHxy)$ ', e.g., is both a formula and a sentence.) That is all. If you've understood these rules and definitions, you are an expert in the grammar of  $\mathcal{U}$ !



(C) Having said which expressions are sentences of  $\mathcal{U}$ , we need to say something about their *semantics* – about the way they connect with the world and, in particular, the conditions under which they are true and false. Roughly, we are going to say that the “status” of a complex formula is determined by its main operator with the status of its immediate subformulas, and the status of an atomic subformula is determined directly. This makes it possible to calculate the status of a complex formula, moving subformula by subformula, from the atomics to the whole. Following Frege, from whom Russell and Quine adapt their notation, it is thus natural to see the fundamental semantic notion as that of a *function*. After a brief word about functions generally, we’ll take up atomic formulas, formulas with truth functional main operators, formulas with quantifier main operators, and work some examples.

(i) One might think of a function as a “black box” which takes some object(s) as input and, in response, extrudes an object as output. The output may or may not be the same as an input, but the output is always the same for any given inputs. The functions with which we are most familiar are those from mathematics. Set aside metaphysical questions about the nature of numbers and functions. Then we may think of, e.g.,  $2 * x$  as a function which takes an input object in the  $x$ -place and supplies another object as output. When the input is 1, the output is 2; when the input is 2, the output is 4; etc. Similarly,  $x + y$  is a function which takes one object in the  $x$ -place, and another in the  $y$ -place. With 1 in the  $x$ -place and 2 in the  $y$ -place ( $\llbracket 1/x, 2/y \rrbracket$ ), the output is 3; with  $\llbracket 2/x, 415/y \rrbracket$  the output is 417, etc.

It is important that functions may combine, so that the output of one is the input to another. Such combination may be illustrated by a “tree” diagram. So, e.g., we may understand  $y + (2 * x)$  by,



$y + (2 * x)$  results from operation of the addition function on  $y$  and  $2 * x$ , and  $2 * x$  results from operation of the multiplication function on 2 and  $x$ . We do the calculation from the “tips” to the whole. So, e.g., with  $\llbracket 2/x, 1/y \rrbracket$ ,  $2 * x$  is 4, and  $y + (2 * x)$  is 5. This way of seeing things may seem to make complicated what is intuitively obvious. But it may also make explicit what has been going on all along. It will be helpful to have this explicit understanding as we turn to the functions of our canonical notation.

It is possible to understand predicate letters, along with each of the truth functional connectives and quantifiers, as designating functions. In each case, the functions have, as output, *the true* or *the false* ( $T$  or  $F$ ). Again, let’s set aside metaphysical worries about functions, and about  $T$  and  $F$ . As it turns out, related entities

are commonplace in mathematics. So, for now at least, we'll proceed in faith that somehow there is an acceptable account of them, and exploit the function picture for our understanding of the canonical notation. Collectively, we may think of conditions that follow as a second account of truth, **T2**.

(ii) Like the predicate terms of chapter 1, predicate letters designate functions which *apply* to objects or not. If a predicate function applies to an object then the output is T, if it does not apply then the output is F. In a given context, we *specify* the relevant functions. Thus, as before, we might say  $Px$  designates the same function as ( $x$  is president). Then an atomic formula with some assignment of objects to its variables designates a truth value, T or F depending on whether the predicate applies to the thing. So  $Px$  with  $\llbracket \text{George}/x \rrbracket$  is T, and  $Px$  with  $\llbracket \text{Bill}/x \rrbracket$  is F. Similarly, if  $Lxy$  is ( $x$  loves  $y$ ), then  $Lxy$   $\llbracket \text{Søren}/x, \text{Regina}/y \rrbracket$ , is T and,  $Lxy$   $\llbracket \text{Regina}/x, \text{Søren}/y \rrbracket$  is F.<sup>3</sup> Notice that we designate functions *in the metalanguage*. Thus, e.g., if  $Px$  designates ( $x$  is president),  $Px$  is ( $x$  is president) and  $Py$  is ( $y$  is president), and if  $Lxy$  is ( $x$  loves  $y$ ),  $Lzw$  is ( $z$  loves  $w$ ) and  $Lzw$   $\llbracket \text{Regina}/w, \text{Søren}/z \rrbracket$  is T. In practice, it will be convenient sometimes to produce “word salads” mixing ordinary language with  $\mathcal{U}$ . So, e.g., ‘ $\exists x \forall y (x \text{ loves } y)$ ’ is like ‘ $\exists x \forall y Lxy$ ’ where  $Lxy$  is ( $x$  loves  $y$ ).

A few comments: First, it is an idealization to suppose that ordinary predicates represent functions. Functions are never vague. So, e.g., if  $Bx$  is ( $x$  is bald),  $Bx$  may seem definitely to return T for some objects, and F for others. But, where hair is merely thin, we may be tempted to say that  $Bx$  returns neither T nor F – or maybe a bit of both. We'll simply assume that the application of predicate functions is or can be specified so that there is always some definite output. Thus we might take  $Bx$  to be ( $x$  has  $\leq 64$  hairs on his head). Even so,  $Bx$  may seem *undefined* on certain objects: it is now clear enough what it is for  $Bx$  to apply to you or to me, but is it so clear what it is for it to apply to a computer? Again, we'll simply assume that predicate functions are or can be specified so that they have some definite output for any input object. So, e.g., we might take  $Bx$  to be the same as ( $x$  has a head with  $\leq 64$  hairs on it), or we might resort to specifying its application by list, say,  $Bx$ : {Telly, Yule...}.

Second, observe that predicate functions, once they are made precise, *do* correspond to lists. Thus, where  $Sx$  is ( $x$  is over 6 feet tall) there is a corresponding collection {Shaq, Wilt...}; and asking whether  $Sx$   $\llbracket \text{Shaq}/x \rrbracket$  is true is the same as asking whether Shaq is on the list (he is). Similarly, where  $Lxy$  is ( $x$  loves  $y$ ) there is a collection, { $\langle \text{Søren}, \text{Regina} \rangle$ ,  $\langle \text{Romeo}, \text{Juliet} \rangle$ ,  $\langle \text{Juliet}, \text{Romeo} \rangle$ ...} of *ordered pairs* such that the first member loves the second. In this case, asking whether  $Lxy$   $\llbracket \text{Søren}/x, \text{Regina}/y \rrbracket$  is true is the same as asking whether  $\langle \text{Søren}, \text{Regina} \rangle$  is in the list (it is); and asking whether  $Lxy$   $\llbracket \text{Regina}/x, \text{Søren}/y \rrbracket$  is true is the same as asking whether  $\langle \text{Regina},$

---

<sup>3</sup> Kierkegaard was much influenced by his unrequited love for Regina Olsen. That relationship is said to be behind his discussion of the Knights of Faith and of Infinite Resignation in his *Fear and Trembling*.

Søren) is on the list (it is not). Sometimes we will find it easier to think of the functions in these terms.

Finally, I have used names to say how predicate functions work.  $Px$  is T with *George* in the  $x$  place. But don't suppose that predicate functions therefore assume or require names for their operation. It's *George*, not 'George' that is in the  $x$  place. A predicate function applies or doesn't apply *to the thing* – apart from any names of it. When we get around to saying whether a predicate function applies or not (has output T or F), it is natural to use a name. But the function doesn't itself depend on names. This is important. For Quine's idea is that we may avoid Plato's beard by avoiding the use of proper names. Of course, we'll need to see how to do without proper names. But, insofar as the canonical notation does use predicates, it shouldn't thereby smuggle in the use of proper names as well. So it's a good thing that predicate functions work without names.<sup>4</sup>

(iii) The “truth-functional” operators  $\sim$ ,  $\wedge$ ,  $\vee$ ,  $\rightarrow$ , and  $\leftrightarrow$  designate functions which take truth values as inputs and have a truth value as output. Say  $\mathcal{P}$  and  $\mathcal{Q}$  are formulas, and  $\llbracket \alpha \rrbracket$  includes assignments to all the free variables in  $\mathcal{P}$  and  $\mathcal{Q}$ . Then  $\mathcal{P} \llbracket \alpha \rrbracket$  and  $\mathcal{Q} \llbracket \alpha \rrbracket$  have truth value outputs. Complex functions generated by the operators may be represented on a sort of “tree” diagram. Trees have a “forward” and a “backward” direction. In the forward direction, a formula is broken into its parts according to its main operator. In the backward direction, truth or falsity of the whole is calculated from truth or falsity of the parts. For formulas with truth functional main operators:

Forward:

Backward:

$$\frac{\sim \mathcal{P} \llbracket \alpha \rrbracket}{\sim} \quad \mathcal{P} \llbracket \alpha \rrbracket$$

The trunk is T iff the branch is F;  
the trunk is F iff the branch is T.

$$\frac{\mathcal{P} \wedge \mathcal{Q} \llbracket \alpha \rrbracket}{\wedge} \quad \begin{array}{c} \mathcal{P} \llbracket \alpha \rrbracket \\ \hline \mathcal{Q} \llbracket \alpha \rrbracket \end{array}$$

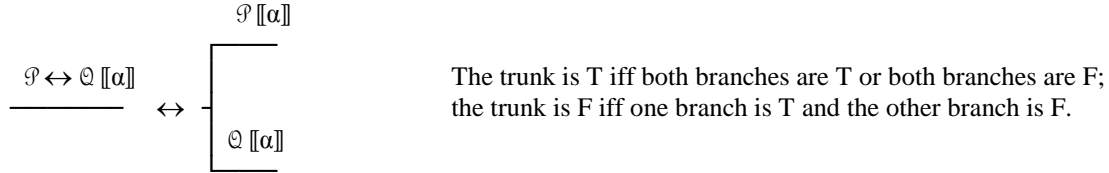
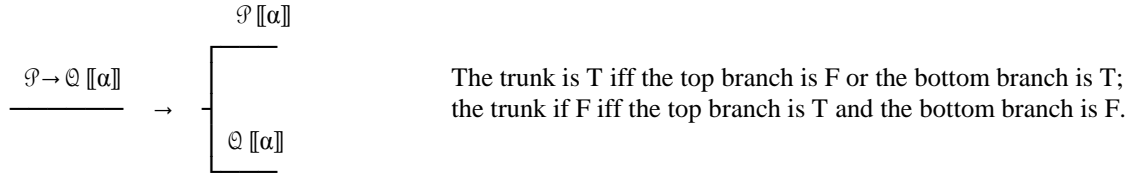
The trunk is T iff both branches are T;  
the trunk is F iff at least one branch is F.

$$\frac{\mathcal{P} \vee \mathcal{Q} \llbracket \alpha \rrbracket}{\vee} \quad \begin{array}{c} \mathcal{P} \llbracket \alpha \rrbracket \\ \hline \mathcal{Q} \llbracket \alpha \rrbracket \end{array}$$

The trunk is T iff at least one branch is T;  
the trunk is F iff both branches are F.

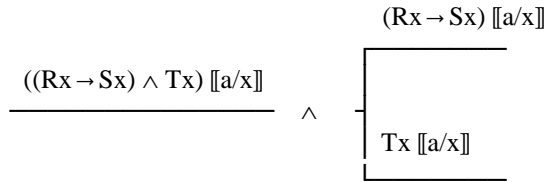
---

<sup>4</sup> G. McCulloch, *The Game of the Name* (Oxford: Clarendon Press, 1989) argues that there is some deep sense in which our notation requires the possibility of proper names. But he argues only that it is possible for there to be names of existing objects—and names for existing objects don't get us into Plato's beard.



An atomic subformula  $\mathcal{P}$  does not branch.<sup>5</sup>

To see how this works, let us consider an example. Suppose we are given that  $(Rx \rightarrow Sx) \llbracket a/x \rrbracket$  is T and  $Tx \llbracket a/x \rrbracket$  is F. The main operator of ‘ $((Rx \rightarrow Sx) \wedge Tx)$ ’ is  $\wedge$ , so  $((Rx \rightarrow Sx) \wedge Tx) \llbracket a/x \rrbracket$  branches as follows:



If we are given that the upper branch is T, and the lower F, at least one branch is F; so, by the “backwards” condition for  $\wedge$ ,  $((Rx \rightarrow Sx) \wedge Tx) \llbracket a/x \rrbracket$  is F.

<sup>5</sup> It is traditional to represent this information in table form as follows.

$\mathcal{P} [\alpha]$	$\mathcal{Q} [\alpha]$	$\sim \mathcal{P} [\alpha]$	$\mathcal{P} \wedge \mathcal{Q} [\alpha]$	$\mathcal{P} \vee \mathcal{Q} [\alpha]$	$\mathcal{P} \rightarrow \mathcal{Q} [\alpha]$	$\mathcal{P} \leftrightarrow \mathcal{Q} [\alpha]$
T	T	F	T	T	T	T
T	F	F	F	T	F	F
F	T	T	F	T	T	F
F	F	T	F	F	T	T

Thus when  $\mathcal{P} [\alpha]$  has output T,  $\sim \mathcal{P} [\alpha]$  is F; when  $\mathcal{P} [\alpha]$  is F,  $\sim \mathcal{P} [\alpha]$  is T; when  $\mathcal{P} [\alpha]$  is T and  $\mathcal{Q} [\alpha]$  is F,  $\mathcal{P} \rightarrow \mathcal{Q} [\alpha]$  is F; etc. Again,  $\sim \mathcal{P}$  is *any* formula with main operator  $\sim$  and immediate subformula  $\mathcal{P}$ ;  $\mathcal{P} \wedge \mathcal{Q}$ , is *any* formula with main operator ‘ $\wedge$ ’ and immediate subformulas  $\mathcal{P}$  and  $\mathcal{Q}$ ; etc. You should be able to see that this comes to the same thing as the trees.

$$\frac{((Rx \rightarrow Sx) \wedge Tx) \llbracket a/x \rrbracket^F}{\wedge} \begin{array}{l} (Rx \rightarrow Sx) \llbracket a/x \rrbracket^T \\ \hline Tx \llbracket a/x \rrbracket^F \end{array}$$

We'll get to an extended example in a moment. For now, observe that these conditions for trees mesh with the intuitive presentation with which we began, and at the same time make precise just what is involved. Thus the standard reading of  $\sim \mathcal{P}$  is, "it is not the case that  $\mathcal{P}$ ":  $\sim \mathcal{P}$  is T just when  $\mathcal{P}$  isn't T. The standard reading of  $\mathcal{P} \wedge \mathcal{Q}$  is, " $\mathcal{P}$  and  $\mathcal{Q}$ ":  $\mathcal{P} \wedge \mathcal{Q}$  is T just when both  $\mathcal{P}$  and  $\mathcal{Q}$  are T. The standard reading of  $\mathcal{P} \vee \mathcal{Q}$  is, " $\mathcal{P}$  or  $\mathcal{Q}$ ":  $\mathcal{P} \vee \mathcal{Q}$  is T just when  $\mathcal{P}$  is T or  $\mathcal{Q}$  is T (or both). The standard reading of  $\mathcal{P} \rightarrow \mathcal{Q}$  is, "if  $\mathcal{P}$  then  $\mathcal{Q}$ ":  $\mathcal{P} \rightarrow \mathcal{Q}$  is T just when one can move from the truth of  $\mathcal{P}$  to the truth of  $\mathcal{Q}$ . And the standard reading of  $\mathcal{P} \leftrightarrow \mathcal{Q}$  is, " $\mathcal{P}$  if and only if  $\mathcal{Q}$ ":  $\mathcal{P} \leftrightarrow \mathcal{Q}$  is T just when one can move from the truth of  $\mathcal{P}$  to the truth of  $\mathcal{Q}$ , and from the truth of  $\mathcal{Q}$  to the truth of  $\mathcal{P}$ . We'll say more about associations with ordinary language, after taking up quantifier functions.

(iv) An  $x$ -quantifier designates a function which takes as inputs truth values from the immediate subformula for *every* assignment to  $x$ . Again, it will be helpful to specify quantifier functions in terms of trees. For any variable assignment  $\llbracket \alpha \rrbracket$ , let  $\llbracket \alpha, \beta/x \rrbracket$  include all the assignments in  $\llbracket \alpha \rrbracket$  and  $\beta/x$  as well. Suppose all the things in the universe are  $a, b, \dots$ . Then quantifier conditions are:

Forward:

Backward:

$$\frac{\exists x \mathcal{P} \llbracket \alpha \rrbracket}{\exists x} \begin{array}{l} \mathcal{P} \llbracket \alpha, a/x \rrbracket \\ \hline \mathcal{P} \llbracket \alpha, b/x \rrbracket \\ \vdots \\ \text{one branch for} \\ \text{each thing} \end{array}$$

The trunk is T iff at least one branch is T;  
the trunk is F iff every branch is F.

$$\frac{\forall x \mathcal{P} \llbracket \alpha \rrbracket}{\forall x} \begin{array}{l} \mathcal{P} \llbracket \alpha, a/x \rrbracket \\ \hline \mathcal{P} \llbracket \alpha, b/x \rrbracket \\ \vdots \\ \text{one branch for} \\ \text{each thing} \end{array}$$

The trunk is T iff every branch is T;  
the trunk is F iff at least one branch is F.

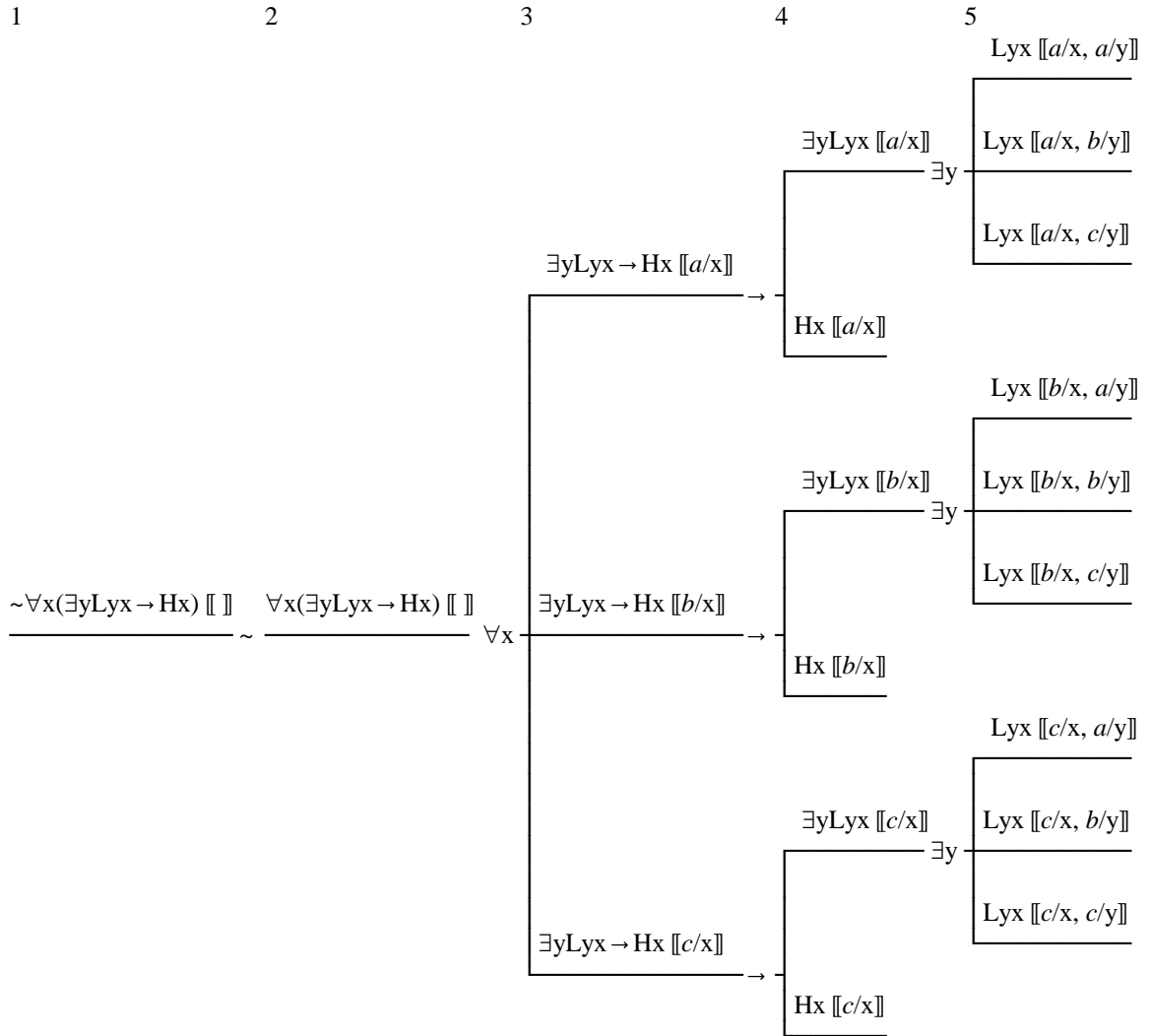
In the special case when there are no things, and so no branches, we understand the existential to be false (no branch is T), and the universal to be T (no branch is F). Completing a tree is impractical if there are more than three or four objects, and impossible if there are infinitely many. Still, we can *use* trees to see how the different functions work. As we have seen, the standard readings of  $\forall x\mathcal{P}$  are “for any  $x$ ,  $\mathcal{P}$ ” and “any  $x$  is such that  $\mathcal{P}$ ”. The standard readings of  $\exists x\mathcal{P}$  are “for some  $x$ ,  $\mathcal{P}$ ” and “there is an  $x$  such that  $\mathcal{P}$ ” – where “some” is understood to mean, “at least one.”

A full tree typically begins with a *sentence*  $\mathcal{P}$ , and an *empty* assignment  $\llbracket \rrbracket$ ; the sentence branches according to its immediate subformula(s), which branch according to their immediate subformula(s), etc. to atomics at the tips. Since a sentence has no free variables, for any  $\mathcal{Q} \llbracket \alpha \rrbracket$ , in the tree,  $\llbracket \alpha \rrbracket$  is sure to include assignments to all the free variables in  $\mathcal{Q}$ ; so, if  $\mathcal{Q} \llbracket \alpha \rrbracket$  is at a tip,  $\llbracket \alpha \rrbracket$  is sure to include assignments to all the variables in  $\mathcal{Q}$ ; so the backward calculation is sure to be possible. If  $\mathcal{P}$  returns T with the empty assignment,  $\mathcal{P}$  is simply *true*, and if  $\mathcal{P}$  returns F with the empty assignment,  $\mathcal{P}$  is simply *false*.

Again, an example should help. For now, let’s pretend there are only three things in the universe,  $a$ ,  $b$ , and  $c$ . Assume that  $a$  loves  $a$ ,  $a$  loves  $b$ ,  $a$  loves  $c$  and  $b$  loves  $a$ , where these are the only loving relations; also assume that  $b$  is happy and  $c$  is happy, where these are the only happy individuals.  $Lxy$  is ( $x$  loves  $y$ ), and  $Hx$  is ( $x$  is happy). In compact form, we consider an *interpretation*  $I$  of the language as follows,

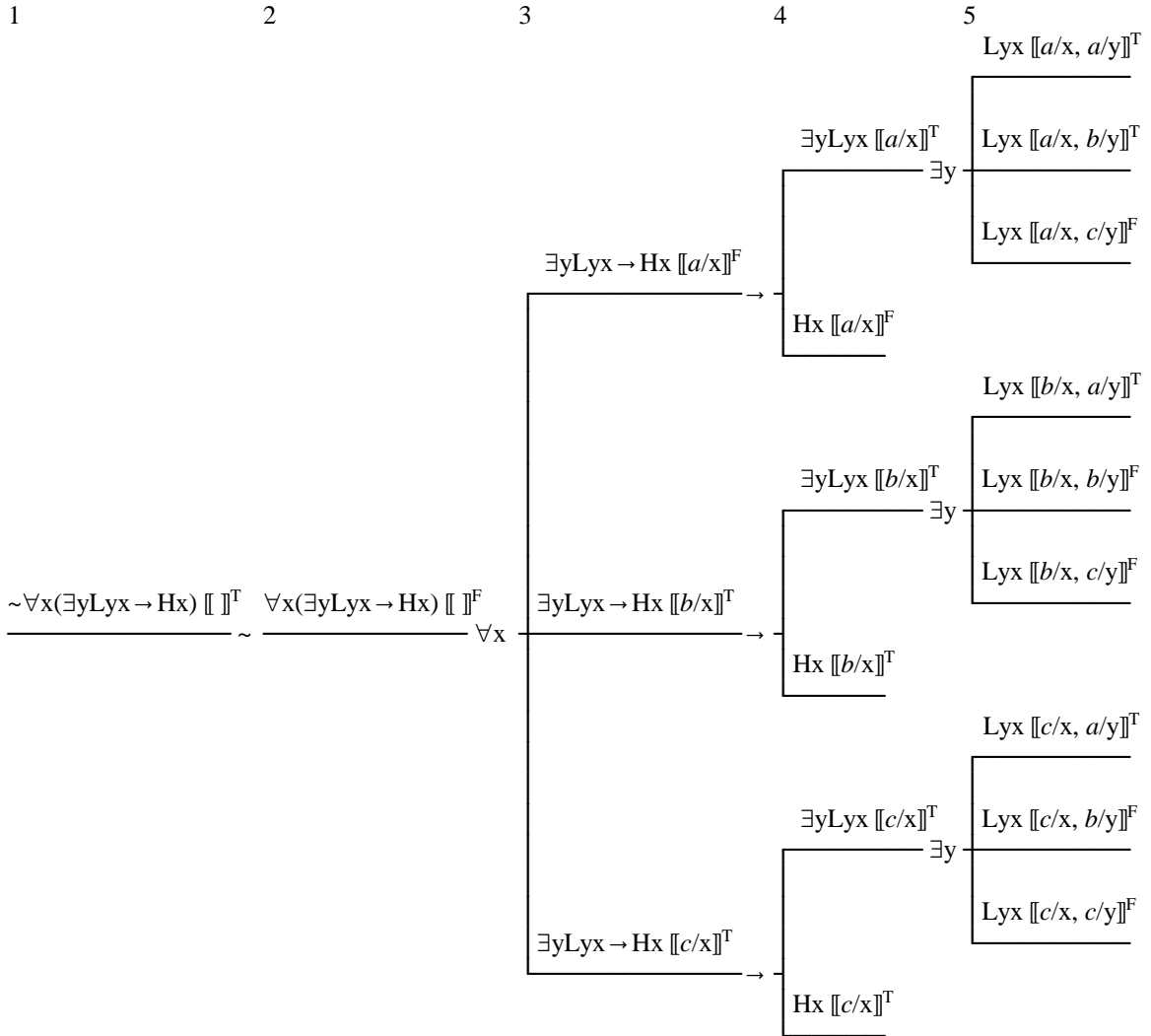
$$\begin{aligned} U &= \{a, b, c\} \\ I(H) &= \{b, c\} \\ I(L) &= \{\langle a, a \rangle, \langle a, b \rangle, \langle a, c \rangle, \langle b, a \rangle\} \end{aligned}$$

The *universe*  $U$  of the interpretation consists of  $\{a, b, c\}$ . We understand  $H$  as applying just to  $b, c$ . And  $L$  applies just to the ordered pairs,  $\langle a, a \rangle, \langle a, b \rangle, \langle a, c \rangle, \langle b, a \rangle$ . We’ll consider the function  $\sim \forall x(\exists y L y x \rightarrow H x)$ . This is complicated! But we can attack it, one operator at a time. First, in the forward direction, we begin with the sentence, and empty assignment, working in the forward direction:



At stage (1) we begin with the sentence  $\sim \forall x(\exists y \text{Lyx} \rightarrow \text{Hx})$  and the empty assignment  $\llbracket \rrbracket$ . The main operator is  $\sim$ , so there is only one branch at stage (2). Notice that, as is always the case for truth functional operators, the variable assignment simply carries forward. The main operator of the formula at stage (2) is  $\forall x$ . In this case, there are as many branches at the next stage (3) as there are things, though assignments on the branches have the different objects in the  $x$  place. Now  $\rightarrow$  is the main operator on each branch; so, for each stage (3) branch there are two new branches at stage (4). Again, variable assignments carry forward from the branch before. There is nothing to be done with the lower limbs of each pair at stage (4). But the upper limbs get as many new branches as there are things. Notice that existing variable assignments are augmented with appropriate assignments in the  $y$  place. This completes the forward tree.

Now we are ready to calculate truth values from the tips back toward the trunk. Recall that  $H$  applies just to  $b$ ,  $c$ , and  $L$  just to  $\langle a,a \rangle$ ,  $\langle a,b \rangle$ ,  $\langle a,c \rangle$ ,  $\langle b,a \rangle$ .

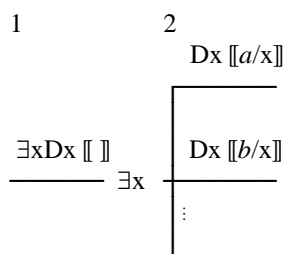


The easiest to see are the lower branches of the pairs at (4). It is given that  $b$  and  $c$  are happy but  $a$  is not. So  $\text{Hx } \llbracket b/x \rrbracket$  and  $\text{Hx } \llbracket c/x \rrbracket$  are T but  $\text{Hx } \llbracket a/x \rrbracket$  is not. For the top tip at (5), we have that  $a$  loves  $a$ , so  $\text{Lyx}$  has output T with  $\llbracket a/x, a/y \rrbracket$ ; for the next,  $b$  loves  $a$ , so  $\text{Lyx}$  has output T with  $\llbracket a/x, b/y \rrbracket$ ; but  $c$  does not love  $a$ , so  $\text{Lyx}$  is F with  $\llbracket a/x, c/y \rrbracket$ . You should be able to check other tips at (5) by similar reasoning. Now, moving back from the tips, an existential is T if and only if there is at least one branch that is T. For each existential quantifier, there is at least one branch that is T, so the upper branches of the pairs at (4) are each T. The only way for  $\rightarrow$  to be F is if the top branch is T and the lower branch is F. So the first branch at (3) is F, and the others are T. Since the universal at (2) has a branch that is F, the result is F. Finally, the branch for the negation is F; so the trunk at (1) is T. So, on our assumptions,  $\sim \forall x (\exists y \text{Lyx} \rightarrow \text{Hx})$  is T. You should convince yourself that it would be false if, other things being equal, nobody loved  $a$ .



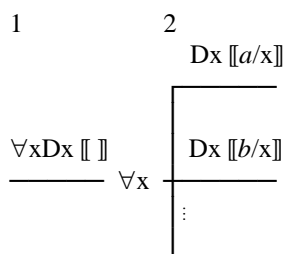
(D) So far, we’ve been able to calculate the truth or falsity of a sentence from an arrangement of truth values at the tips of its tree. Now, in order to see “what a sentence says” about the world, we need to think more generally about the range of conditions under which it is true. I’ll continue to work with trees, though you should *use* them to reach the point where you can “read” sentences on your own. Let  $Lx$  be ( $x$  is lucky);  $Dx$ , ( $x$  is a dog); and  $Bxy$ , ( $x$  barks at  $y$ ). Say all the things in the universe are  $a, b, \dots$

‘ $\exists xDx$ ’ is read, “there is an  $x$  such that  $Dx$ ”; it is true iff there is at least one dog.



$\exists xDx \llbracket \rrbracket$  is T at stage (1) just in case at least one branch at (2) is T. So  $\exists xDx \llbracket \rrbracket$  is T iff there is at least one object  $o$  such that  $Dx \llbracket o/x \rrbracket$  is T – that is, iff there is at least one dog. The only way for  $\exists xDx \llbracket \rrbracket$  to be F is if there are no dogs. This, of course, is just what we thought it said from our original intuitive presentation.

‘ $\forall xDx$ ’ is read, “for any  $x$ ,  $Dx$ ”; it is true iff everything is a dog.



$\forall xDx \llbracket \rrbracket$  is T at (1) just in case each of the branches at (2) is T. So  $\forall xDx \llbracket \rrbracket$  is T iff for any object  $o$ ,  $Dx \llbracket o/x \rrbracket$  is T; so it’s T iff everything is a dog.  $\forall xDx \llbracket \rrbracket$  is F, if there is even one thing that isn’t a dog.

‘ $\sim \exists xDx$ ’ is read, “it is not the case that there is an  $x$  such that  $Dx$ ”; it is true iff there are no dogs.

1	2	3
		$Dx \llbracket a/x \rrbracket$
$\sim \exists x Dx \llbracket \rrbracket$	$\sim$	$\exists x$ $Dx \llbracket b/x \rrbracket$
		$\vdots$

$\sim \exists x Dx \llbracket \rrbracket$  is T at (1), just in case  $\exists x Dx \llbracket \rrbracket$  is F at (2); but the only way for  $\exists x Dx \llbracket \rrbracket$  to be F is for each of the branches at (3) to be F. If even one branch at (3) is T, then  $\exists x Dx \llbracket \rrbracket$  is T at (2), and  $\sim \exists x Dx \llbracket \rrbracket$  is F.

‘ $\exists x \sim Dx$ ’ is read, “there is an  $x$  such that it is not the case that  $Dx$ ”; it is true iff at least one thing isn’t a dog.

1	2	3
	$\sim Dx \llbracket a/x \rrbracket$	$Dx \llbracket a/x \rrbracket$
$\exists x \sim Dx \llbracket \rrbracket$	$\sim$	$\sim$
	$\sim Dx \llbracket b/x \rrbracket$	$Dx \llbracket b/x \rrbracket$
	$\sim$	$\sim$
	$\vdots$	

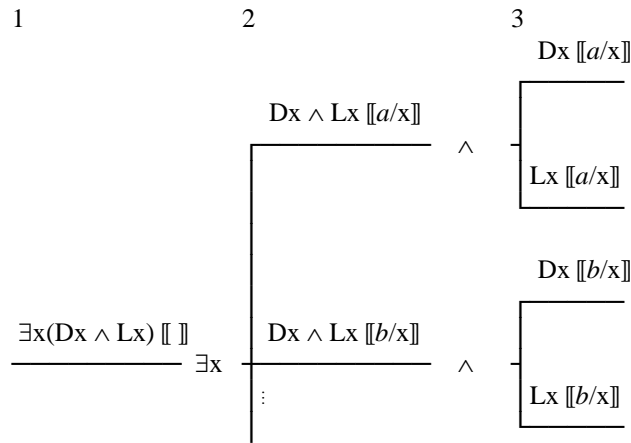
$\exists x \sim Dx \llbracket \rrbracket$  is T at (1) just in case at least one branch at (2) is T, and the only way for a branch at (2) to be T, is for the corresponding branch at (3) to be F; if a branch at (3) is F, the corresponding branch at (2) is T, and the existential at (1) is T. The only way for  $\exists x \sim Dx \llbracket \rrbracket$  to be F, is for everything to be a dog; if everything is a dog, then all the branches at (3) are T; so all the branches at (2) are F; so  $\exists x \sim Dx \llbracket \rrbracket$  at (1) is F.

‘ $\exists x Dx \wedge \exists x Lx$ ’ is read, “there is an  $x$  such that  $Dx$ , and there is an  $x$  such that  $Lx$ ”; it is true iff something is a dog, and something is lucky.

1	2	3
		$Dx \llbracket a/x \rrbracket$
	$\exists x Dx \llbracket \rrbracket$	$Dx \llbracket b/x \rrbracket$
	$\exists x$	$\vdots$
$\exists x Dx \wedge \exists x Lx \llbracket \rrbracket$	$\wedge$	$Lx \llbracket a/x \rrbracket$
	$\exists x Lx \llbracket \rrbracket$	$Lx \llbracket b/x \rrbracket$
	$\exists x$	$\vdots$

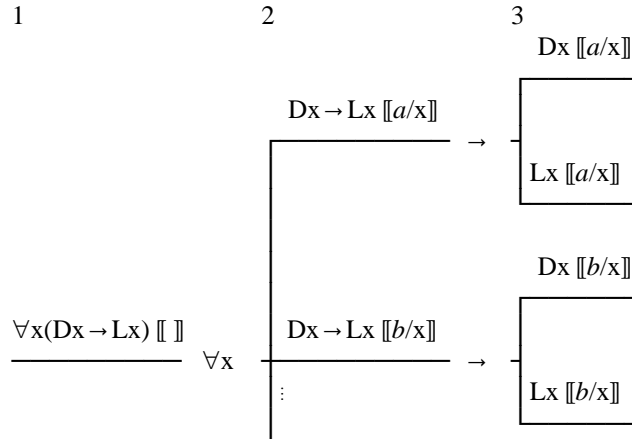
$\exists xDx \wedge \exists xLx$  is T at (1), just in case both branches are T at (2); for the top branch to be T, at least one thing has to be a dog; for the bottom branch to be T, at least one thing has to be lucky. It doesn't matter *which* thing is a dog, and which is lucky, so long as there is at least one of each (notice that nothing prevents the lucky thing from being the dog).  $\exists xDx \wedge \exists xLx$  is F at (1) if one or both of the branches at (2) is F; to make the top branch F, there would have to be no dogs; to make the bottom branch F, there would have to be no lucky things. If there are no dogs, or there are no lucky things,  $\exists xDx \wedge \exists xLx$  is F.

' $\exists x(Dx \wedge Lx)$ ' is read, "there is an  $x$  such that  $Dx$  and  $Lx$ "; it is true iff there is at least one lucky dog.



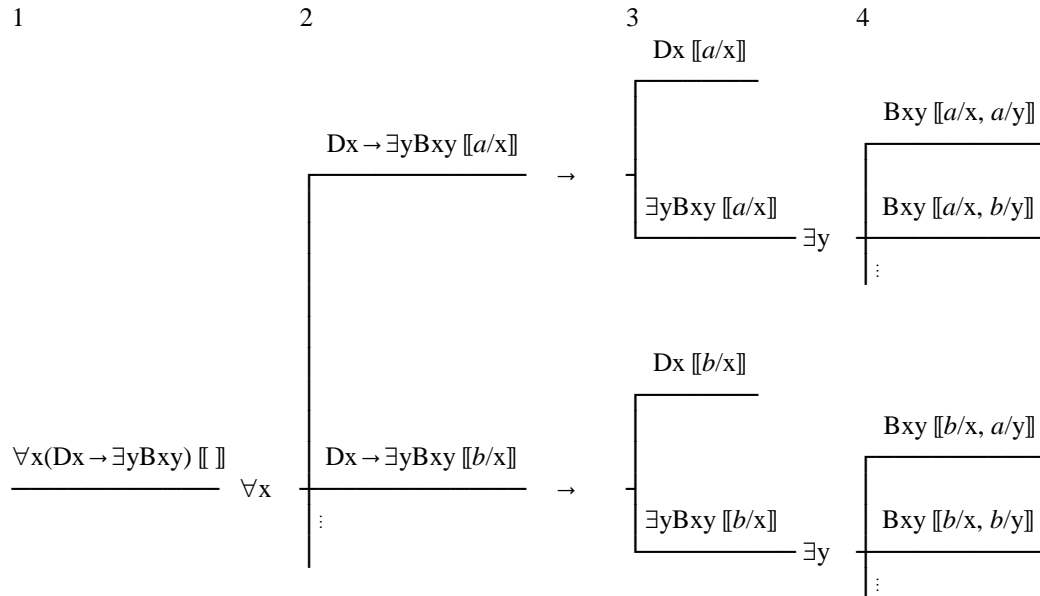
$\exists x(Dx \wedge Lx)$  is T at (1) just in case at least one of the branches at (2) is T; for one of those branches to be T, both branches of the corresponding pair at (3) must be T; so if  $\exists x(Dx \wedge Lx)$  is T, some one object  $o$  must be both a dog and lucky.  $\exists x(Dx \wedge Lx)$  is F iff all the branches at (2) are F; this is the case if no dogs are lucky. Notice: if  $\exists x(Dx \wedge Lx)$  is T,  $(\exists xDx \wedge \exists xLx)$  is T as well, but not the other way around; for  $\exists x(Dx \wedge Lx)$  is T only if some *one* thing is both lucky and a dog. This is just the contrast with which we left off in our original intuitive presentation.

‘ $\forall x(Dx \rightarrow Lx)$ ’ is read, “for any  $x$ , if  $Dx$  then  $Lx$ ”; it is true iff all dogs are lucky.



$\forall x(Dx \rightarrow Lx) \llbracket \rrbracket$  is T at (1) just in case all of the branches at (2) are T; for this to be the case, there can be no pair at (3) for which the top is T and the bottom is F; so  $\forall x(Dx \rightarrow Lx) \llbracket \rrbracket$  is T just in case anything that is a dog is also lucky. Notice: If  $o$  isn’t a dog, then the corresponding branch at (2) is T *whether  $o$  is lucky or not*. It’s only when  $o$  is a dog, that the corresponding branch at (2) has any chance of being F. But if some  $o$  is a dog and isn’t lucky, then the corresponding branch at (2) is F, and  $\forall x(Dx \rightarrow Lx) \llbracket \rrbracket$  is F as well.

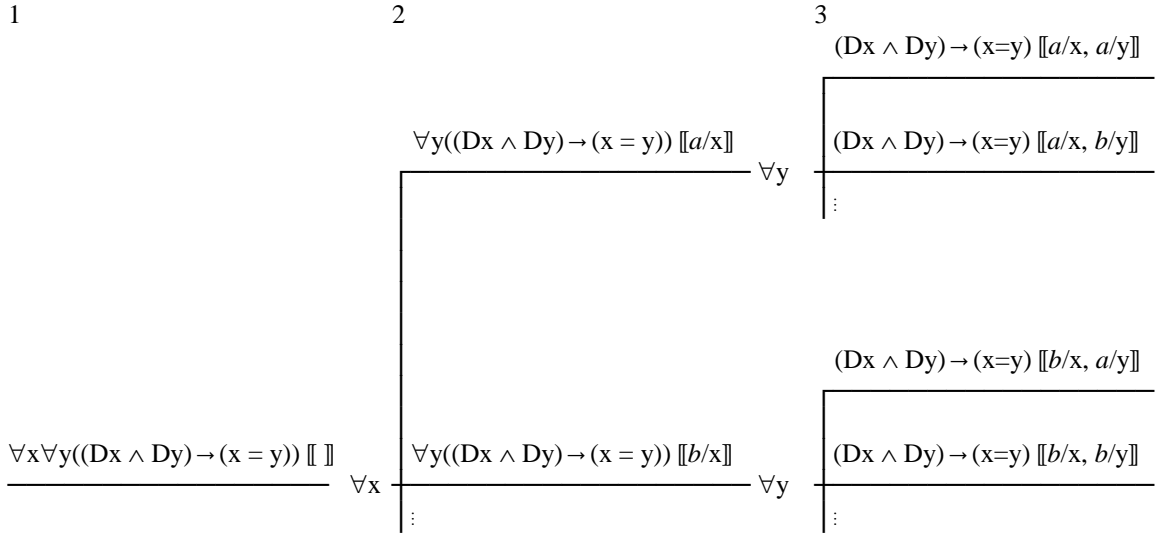
‘ $\forall x(Dx \rightarrow \exists yBxy)$ ’ is read, “for any  $x$ , if  $Dx$  then there is a  $y$  such that  $Bxy$ ”; it is true iff every dog barks at something (which may or may not be itself).



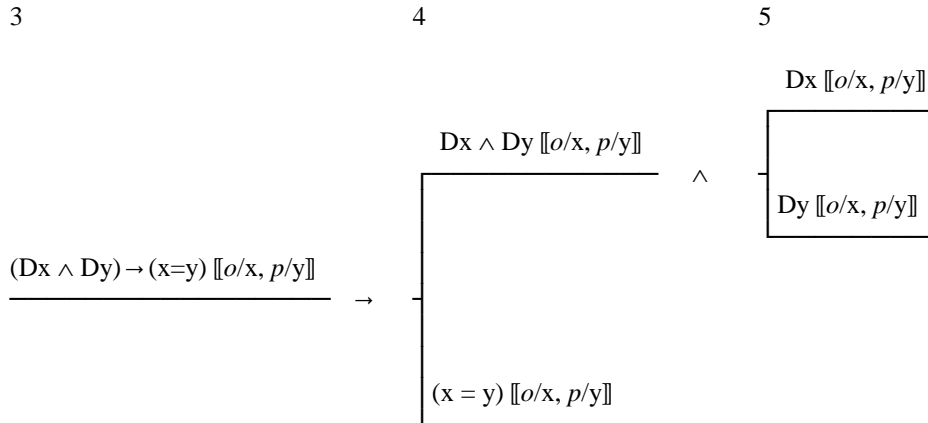
As in the previous example,  $\forall x(Dx \rightarrow \exists yBxy) \llbracket \rrbracket$  is T at (1) just in case all of the branches at (2) are T; for this; there can be no pair at (3) for which the top is T and the bottom is F.

As before, if  $o$  isn't a dog, the corresponding branch at (2) is automatically T. But the cases differ with respect to the condition dogs must satisfy. If  $o$  is a dog, the bottom branch at (3) is T just in case one of the corresponding branches at (4) is T—that is, just in case there is some  $p$  such that  $Bxy \llbracket o/x, p/y \rrbracket$  is T. If some  $o$  is a dog and there is no  $p$  at which it barks, then at (3), a top branch is T and bottom F; so the corresponding branch at (2) is F, and the universal at (1) is F.

' $\forall x \forall y ((Dx \wedge Dy) \rightarrow (x = y))$ ' is read, "for any  $x$  and any  $y$ , if  $Dx$  and  $Dy$ , then  $(x = y)$ "; it is T iff there is at most one dog. Here is the beginning of its tree.



$\forall x \forall y ((Dx \wedge Dy) \rightarrow (x = y)) \llbracket \rrbracket$  is T at (1) just in case all the branches at (2) are T; and these are T just in case all the branches at (3) are T. But now consider what it is for one of the branches at (3) to be T.



Each branch at (3) is T, if no pair at (4) has the top T and the bottom F. For a top at (4) to be T, both  $o$  and  $p$  have to be dogs; if one or both isn't a dog, then the top at (4) is F, and the branch at (3) is T. But suppose  $o$  and  $p$  are both dogs, then the bottom is false unless  $o$  is  $p$ . If  $o$  and  $p$  are distinct dogs, then (4) will have the top T and bottom F; so the branch at (3) will be F, and the universals at (2) and (1) will be F as well. So  $\forall x \forall y ((Dx \wedge Dy) \rightarrow (x = y))$  is T if there are no dogs (all the tops at (4) will be F), or there is one dog (distinct dogs are required to make the bottom at (4) F), and F if there is more than one. In effect, if  $o$  is a dog, we're saying that *anything that is a dog, is it*.

### III. Russell's Way Out

Some languages very much like  $\mathcal{U}$  include a category of *individual constants*, where these work something like ordinary proper names. But Quine's idea is to avoid Plato's beard by avoiding proper names. So we have done without individual constants here. Following Russell, the idea is that the work ordinarily associated with proper names may be shifted onto predicates, truth functions, and quantification—and that doing so voids Plato's beard.<sup>6</sup> Russell seems to think of his solution along the lines of (i) from the first section of this chapter—and so to think that there is something illusory about ordinary proper names. This is controversial. But it's enough for our purposes that he gives us a way of saying 'Pegasus does not exist' without contradiction. Following Russell and Quine, we'll begin with definite descriptions—with phrases of the sort, "the so-and-so...", and then we'll move to ordinary proper names.

(A) Russell was concerned with, 'The present king of France is bald.' Given **T1**, for this to be true, there must be a present king of France to which the predicate applies. But, of course, there is no present king of France. So the sentence isn't true. Perhaps this seems right. But 'The present king of France does not exist' is more problematic. Given **T1**, it is true only if the present King of France both exists and does not exist. So we have a straightforward version of Plato's beard. We'll see how Russell uses our canonical notation to get out of Plato's beard, after we see how he uses it to say "The present king of France is bald."

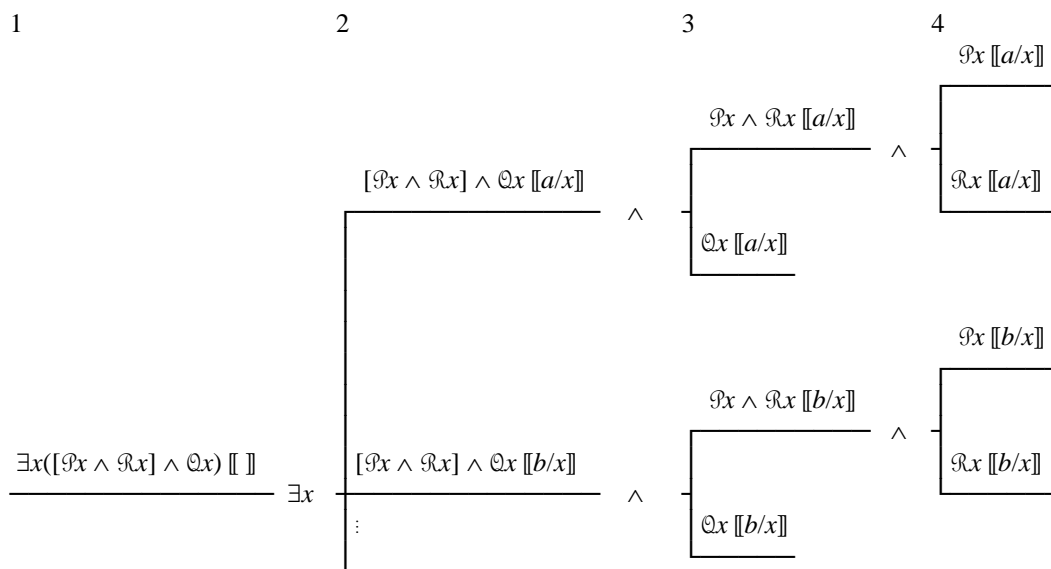
As observed in the last chapter, phrases of the sort, "the so-and-so..." seem to *fail* when there is more than one so-and-so. Similarly, phrases of the sort, "the so-and-so..." seem to fail when there aren't any so-and-sos. Thus, e.g., neither 'the desk at CSUSB has graffiti on it' nor 'the present king of France is bald' seem to be true. The first because the description fails to pick out just one object, and the second because the description doesn't pick out any object at all. Of course, if a description does pick out just one object, then the predicate must apply. So, e.g., 'The president of the USA is a woman' isn't true. There is exactly one object which is the president of the USA, but it isn't a

---

<sup>6</sup> See B. Russell, "On Denoting," originally in *Mind* 14 (1905). Reprint in many places, including *Meaning and Truth*, ed. Garfield and Kiteley (New York: Paragon House, 1991).

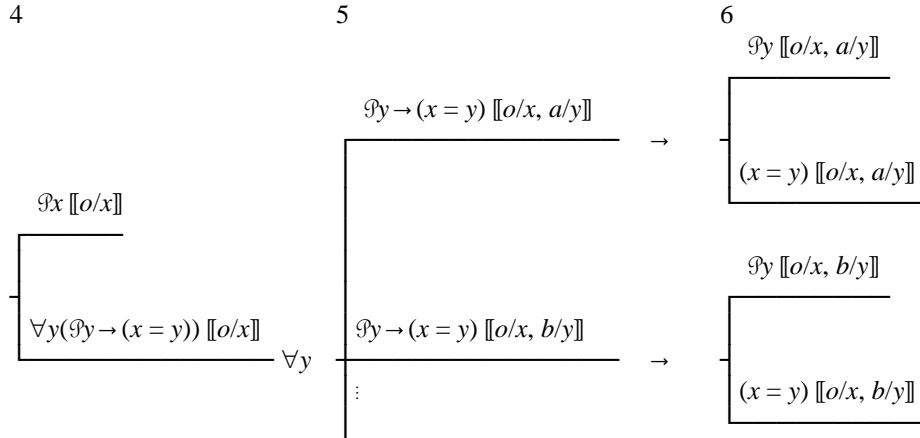
1. At least one thing is  $\mathcal{P}$ .
2. At most one thing is  $\mathcal{P}$ .
3. Whatever is  $\mathcal{P}$  is  $\mathcal{Q}$ .

Russell uses canonical notation for his account of truth conditions. Thus we'll express his conditions for "the  $\mathcal{P}$  is  $\mathcal{Q}$ " in  $\mathcal{U}$ . Speaking in the metalanguage, where  $\mathcal{P}x$  and  $\mathcal{Q}x$  are (maybe atomic) formulas with just  $x$  free,  $\exists x([\mathcal{P}x \wedge \forall y(\mathcal{P}y \rightarrow (x = y))] \wedge \mathcal{Q}x)$  does the job. Now the trees for this get messy. We will see how this goes. But I think we can see up front what is happening. First, for visual convenience, let  $\mathcal{R}x$  be the subformula,  $\forall y(\mathcal{P}y \rightarrow (x = y))$ , then  $\exists x([\mathcal{P}x \wedge \forall y(\mathcal{P}y \rightarrow (x = y))] \wedge \mathcal{Q}x)$  appears as  $\exists x([\mathcal{P}x \wedge \mathcal{R}x] \wedge \mathcal{Q}x)$ . Then what we need is an object that satisfies each of the three conditions,  $\mathcal{P}x$ ,  $\mathcal{Q}x$ , and  $\mathcal{R}x$ . Here's that much of the tree:



$\exists x([\mathcal{P}x \wedge \mathcal{R}x] \wedge \mathcal{Q}x) \llbracket \rrbracket$  is T at (1) just in case at least one of the branches at (2) is T. But a branch at (2) is T just in case both of the corresponding branches at (3) are T; and a top branch at (3) is T just in case both of the corresponding branches at (4) are T. Thus a branch at (2) is T just in case each of the three corresponding branch tips is T. If an object isn't  $\mathcal{P}$  or isn't  $\mathcal{Q}$ , then the corresponding branch at (2) is F. So far, so good.

Now suppose we are on a branch where some  $o$  is such that  $\mathcal{P}x \llbracket o/x \rrbracket$  and  $\mathcal{Q}x \llbracket o/x \rrbracket$  are T. What does the condition  $\mathcal{R}x, \forall y(\mathcal{P}y \rightarrow (x = y))$ , require? Given that  $\mathcal{P}x \llbracket o/x \rrbracket$  is T, it requires that *anything that is  $\mathcal{P}$  is identical to object  $o$* . Thus, as in the last example we considered above, it requires that at most one thing is  $\mathcal{P}$ . This is just right. Thus, the tree continues:



$\forall y(\mathcal{P}y \rightarrow (x = y)) \llbracket o/x \rrbracket$  at (4) is T just in case all the branches at (5) are T; for this to be the case, there can be no pair at (6) for which the top is T and the bottom is F. Suppose some object  $p$  is  $\mathcal{P}$ ; then the top at (6) is T, and the bottom is false unless  $o$  is  $p$ . If  $o$  and  $p$  are both  $\mathcal{P}$  but distinct, then the corresponding branch at (5) is F, the universal at (4) is F, and the conjunctions at (2) and (3) are F as well. If  $o = p$  then the top at (6) is T, but the bottom is T as well, so the branch at (5) is T. Having determined that  $o$  is  $\mathcal{P}$ ,  $\mathcal{R}$  requires that *anything that is  $\mathcal{P}$  is it*. So, again, it's like the condition from the last example above, and  $\mathcal{R}$  requires that there is at most one  $\mathcal{P}$ . So  $\exists x([\mathcal{P}x \wedge \forall y(\mathcal{P}y \rightarrow (x = y))] \wedge \mathcal{Q}x)$  is T just in case some  $o$  is  $\mathcal{P}$ , is the only  $\mathcal{P}$ , and is  $\mathcal{Q}$ . So each of Russell's conditions is imposed.

Let's return to 'The present King of France is bald'. Relaxing constraints on predicate vocabulary, with  $\text{PKFx}$  for ( $x$  is present king of France) and  $\text{Bx}$  for ( $x$  is bald), we get,

$$\exists x([\text{PKFx} \wedge \forall y(\text{PKFy} \rightarrow (x = y))] \wedge \text{Bx})$$

If there is no present king of France, this is false because it fails the top tips at (4). If there is more than one present king of France, some top tips at (4) may be T, but the corresponding bottom tips are (F). If there is just one present king of France who isn't bald, then a pair at (4) may be T, but the corresponding tip at (3) is F. If there is just one present king of France who is bald, then a pair of tips at (4) is T, as well as the corresponding tip at (3); so the conjunctions at (3) and (2) are T, and the existential at (1) is T as well. Notice: we do *not* treat 'the present king of France' as a name—as we would



under the influence of **T1**. Rather, whatever things there are, we consider each, and ask whether *it* meets certain conditions. If none of the things meets the conditions, then ‘The present king of France is bald’ is false; if some thing meets them, then ‘The present king of France is bald’ is true.

This may or may not be a correct account of the way ordinary definite descriptions work. But we do have at least a natural way to say that the present king of France doesn’t exist. Again, what’s key is that statements “about” the present king of France don’t require that there be a present king of France. We don’t talk about what there isn’t. Rather, a quantifier merely “checks” each object there is, to see whether *it* is a present king of France. If none of them meet the condition, then there isn’t a present king of France. Thus  $\sim\exists xPKFx$  is straightforwardly true. And  $\sim\exists x(PKFx \wedge \forall y[PKFy \rightarrow (x = y)])$ —which requires that there is no *unique* present king of France—is true as well.  $\sim\exists xPKFx$  is T iff  $\exists xPKFx$  is F;  $\exists xPKFx$  is F iff for no object in the *x* place is  $PKFx$  true. Since  $PKFx$  isn’t true for any object in the *x* place,  $\sim\exists xPKFx$  is true. And this is so without any reference to a mysterious non-existing present king of France. So we seem to have said just what we had in mind by ‘The present king of France doesn’t exist’ without mystery. And similarly for ‘The round square cupola on Berkeley college does not exist’.

(B) So much for definite descriptions. What about proper names? Russell’s idea is that we may understand ordinary proper names as “disguised” definite descriptions. So, e.g., by “Aristotle” maybe we mean “the student of Plato and teacher of Alexander”. Then, applying the analysis of definite descriptions, ‘Aristotle was a philosopher’ becomes ‘ $\exists x[(x \text{ is a student of Plato and teacher of Alexander}) \wedge \forall y((y \text{ is a student of Plato and teacher of Alexander}) \rightarrow (x = y))] \wedge (x \text{ is a philosopher})$ ’. And similarly for other cases. Again, Russell seems to think of this as an account of the way ordinary proper names work. But it is enough for us that our notation can be used to say things about particular individuals.<sup>7</sup> In particular, it’s easy to say that Pegasus does not exist. Again, the key is that statements “about” Pegasus don’t require that Pegasus is. Rather, the quantifier “checks” each object there is to see whether *it* is Pegasus and, if so, what features it has. So, e.g., perhaps “the winged horse that was captured by Bellerophon” can stand in for “Pegasus” and ‘Pegasus does not exist’ is true iff  $\sim\exists x(x \text{ is a winged horse captured by Bellerophon})$ —or, maybe better, iff  $\sim\exists x[(x \text{ is a winged horse captured by Bellerophon}) \wedge \forall y((y \text{ is a winged horse captured by Bellerophon}) \rightarrow (x = y))]$ . The latter is true if and only if no unique winged horse was captured by Bellerophon.

One might worry that there isn’t always a description to go with every ordinary proper name. And one might worry that proper names (‘France’, ‘Bellerophon’, etc.) *reappear* in descriptions. But Quine thinks this is no real problem. It is, he thinks, possible to “convert” the content of any ordinary proper name into the predicate position. Thus the content of the name ‘Pegasus’ transfers to the predicate ‘is-Pegasus’, the

---

<sup>7</sup> See note 8 in chapter 2.

content of the name ‘Quine’ transfers to ‘is-Quine’, etc.<sup>8</sup> So, e.g., ‘Arthur didn’t have a round table’ becomes, ‘ $\exists x[(x \text{ is-Arthur}) \wedge \forall y((y \text{ is-Arthur}) \rightarrow (x = y))] \wedge \sim(x \text{ had a round table})$ ’ and ‘Arthur doesn’t exist’ becomes ‘ $\sim \exists x[(x \text{ is-Arthur}) \wedge \forall y((y \text{ is-Arthur}) \rightarrow (x = y))]$ ’. If there never was any Arthur, the second is true and the first is false. If there was an Arthur, then the second is false, and the first depends on whether Arthur had a round table. Again, it seems possible to talk “about” Arthur whether or not there ever was one.

This name-to-predicate suggestion may seem to have problems of its own. But it’s not obviously absurd. Let’s agree that sentences containing ‘Pegasus’ *are* somehow meaningful. Then it’s not obvious why content associated with ‘Pegasus’ can’t be shifted to the predicate position. It is one thing to be me, and another to be you. Quine merely suggests that such properties are expressible by predicates. Even so, if ‘Pegasus’ is meaningless without an object to which it refers, ‘is-Pegasus’ may require an object as well. If a name’s content doesn’t depend on reference then, plausibly, the corresponding predicate doesn’t depend on reference. But if a name’s content does depend on reference then the corresponding predicate may depend on reference as well. Given this, our worry isn’t about the shifting as-such. It’s rather that derived predicates may themselves require objects of reference.

In one place, Quine suggests that this way of putting things may beg the question against him. Perhaps the predicates are basic, and ordinary names are derived from them! Indeed, from his position that Pegasus does not exist, and our assumption that sentences containing ‘Pegasus’ are meaningful, it *follows* that ‘Pegasus’ is meaningful apart from reference. Of course, McX and Wyman might object that this reasoning begs the question against them—for we begin with the assumption that there is no Pegasus. But Quine does raise some considerations which are supposed to tell positively in his favor. First, on Russell’s account, ‘The present king of France is bald’ and ‘The present king of France does not exist’ are sentences with meaning but without reference. Perhaps McX and Wyman would argue that Russell’s account is mistaken as an account of ordinary language, and that **T1** should apply to our ordinary expressions. Even so, ‘ $\exists x([PKFx \wedge \forall y(PKFy \rightarrow (x = y))] \wedge Bx)$ ’ and ‘ $\sim \exists x(PKFx \wedge \forall y(PKFy \rightarrow (x = y)))$ ’ are sentences with meaning, but without reference—at least they require no reference to a thing that is the present King of France. Their semantics is specified by us, so that **T1** does not apply to them. So at least some cases seem to work Quine’s way. And, if this is right, reference or naming isn’t *necessary* for meaning.

Quine suggests that McX and Wyman go wrong by supposing that it is. And he

---

<sup>8</sup> In addition to “On What There Is,” see Quine, “The Scope and Language of Science,” in Quine, *The Ways of Paradox* 2nd ed. (Cambridge: Harvard University Press, 1976), p. 238.

thinks they might have recognized their error by reflecting on an example of Frege's.<sup>9</sup> Frege observes that,

The Morning Star = The Morning Star.

is obvious but,

The Morning Star = The Evening Star.

is not. The first can't be doubted. But the second seems known only as the result of an important discovery. Since it's the same object, Venus, that is named in every case, there is more to the meaning of these terms than that they name such-and-such object. For the meanings differ while the object remains the same. So naming isn't sufficient for meaning. It's possible to make a similar point with proper names. 'Hesperus' and 'Phosphorus' name the Evening and Morning stars. But,

Hesperus = Hesperus

is obvious while,

Hesperus = Phosphorus

is not. Again, the meanings seem to differ while objects remain the same. So there is more to the meaning of these terms than that they name such-and-such object. Of course, this doesn't show that naming isn't *necessary* for meaning. McX and Wyman might very well *grant* Frege's point, and continue to hold that referents are necessary for significance—it's only that something more is required as well. But this response applies only to the Frege case, and not to examples from the previous paragraph—Quine might continue to insist that reference is not *necessary* in order for expressions involving 'The present King of France' to be meaningful. And, in his discussion, Quine says only that reflection on Frege's case might have given McX and Wyman an "inkling" of the difference between a statement's having significance, and its naming something.

Whatever we may conclude about naming and significance, let's notice what has been happening: It has seemed natural to *argue* about the way predicates work, and about whether Pegasus exists. But this is precisely what Quine wants the right to do, and what, given Plato's beard, once seemed impossible. That is, shifting content to the predicate position *relocates* the problem. It's no longer part of our semantic story—our account of the conditions under which sentences are true and false—that Pegasus and the like exist.

---

<sup>9</sup> G. Frege, "On Sense and Meaning" trans. Black in *Translations from the Philosophical Writings of Gottlob Frege* ed. Geach and Black (Totowa: Rowman & Littlefield, 1980). Other translations (sometimes under the title "On Sense and Reference" or "On Sense and Nominatum") and reprints in many places.

As in the last chapter, Quine hasn't shown that there *are* no objects corresponding to 'Arthur', 'Pegasus' and the like—he hasn't even shown that terms in the predicate position do not require them. But he does seem to have found a “neutral” context where we aren't *forced* from the start to admit that there are such objects. Russell's way out, at least leaves *room* for meaning without reference. And that's how it gets us out of Plato's beard.

## Chapter Four

### Quine's First Thesis: How the Notation Commits

We want to “get behind” ordinary claims and theories and ask what does or could make them true. Our idea is to say what makes them true in the canonical notation  $\mathcal{U}$  and so to be very clear about what is going on. Having given, in the last chapter, an account of  $\mathcal{U}$ 's semantics, we want now to focus more directly on the question of *what there is* that makes sentences of  $\mathcal{U}$  true. In this chapter, we'll take up Quine's answer. His answer is organized around the (e12) slogan, “to be is to be the value of a variable.” We'll take some time to understand this. As it turns out, this response generates controversy in two directions. First, it is possible to argue that Quine's criterion commits us to too little—that the truth of canonical statements depends on more than just the values of variables. And, second, it is possible to worry that Quine's criterion commits us to too much—that variables require values which do not exist. We'll get to the second objection in the next chapter. In this chapter, we introduce Quine's criterion, consider the objection that it commits us to too little, and take up some examples which illustrate and apply Quine's method for metaphysics.

#### *I. Quine's Criterion*

Obviously the slogan, “to be is to be the value of a variable” requires expansion and/or clarification. Perhaps it will help to begin with a point only just mentioned in the previous chapter. On a tree, a universal quantifier gets one branch for each thing there is, and the trunk is T iff every branch is T. If there aren't any things, there aren't any branches, and we understand the universal to be T. And, similarly, even if there are things,  $\forall x(\mathcal{P}x \rightarrow \mathcal{Q}x)$  is T so long as  $\mathcal{P}x$  isn't T for any object—if  $\mathcal{P}x$  is never T,  $\mathcal{P}x \rightarrow \mathcal{Q}x$  is always T and the universal is T as well. So accepting such universals does not obviously commit us to the existence of anything at all. In contrast,  $\exists x\mathcal{P}x$  has output T iff  $\mathcal{P}x$  is T for at least one object in its  $x$  place. Thus it is impossible to make an existential statement true, without there being some object which makes it true.

This suggests a simple approach to ontological commitment. Suppose some theory is expressed in the canonical notation. A statement is *universally* quantified iff its main operator is a universal quantifier; a statement is *existentially* quantified iff its main operator is an existential quantifier. Given this, it is natural to suggest that,

EQ The ontological commitments of a theory are shown in its affirmative, existentially quantified statements.<sup>1</sup>

---

<sup>1</sup> A proposal very much like this one is considered on p. 175 of K. Campbell, *Metaphysics: An Introduction* (Encino: Dickenson Pub. Co., 1976).

Condition (EQ) has both a positive and a negative part. Positively, affirmative, existentially quantified statements commit us to the existence of objects; negatively, other statements do not. Thus, e.g.,  $\exists x(x \text{ is a number})$  is true iff there is at least one thing that is a number; if we accept that the affirmative existentially quantified statement is true, then we are forced to accept that there is at least one number. On the other hand, the negation of an existentially quantified statement may be true without any objects at all. Thus, e.g.,  $\sim \exists x(x \text{ is a number})$  is true iff  $\exists x(x \text{ is a number})$  is F, and  $\exists x(x \text{ is a number})$  is sure to be F if there aren't any things. So the negation does not commit us to anything—or, rather, is such that its only commitments are negative. This is precisely Quine's point in the debate with McX and Wyman. Similarly, a nonnegated universally quantified sentence does not *force* us to agree that there is anything at all.  $\forall x(x \text{ is a number})$  requires that everything is a number. But, as above, this may be true in the case where there is nothing at all; so it is at least possible for a person to accept it without being *committed* to the existence of anything.

Unfortunately, however, (EQ) is too simple. To see this, notice that the truth condition for  $\sim \forall x \sim (x \text{ is a number})$  is the same as that for  $\exists x(x \text{ is a number})$ .

1	2	3	4
		$\sim Nx \llbracket a/x \rrbracket$	$Nx \llbracket a/x \rrbracket$
$\sim \forall x \sim Nx \llbracket \rrbracket$	$\forall x \sim Nx \llbracket \rrbracket$	$\sim Nx \llbracket b/x \rrbracket$	$Nx \llbracket b/x \rrbracket$
		$\vdots$	

$\sim \forall x \sim Nx$  is T at (1) just in case  $\forall x \sim Nx$  is F at (2);  $\forall x \sim Nx$  is F just in case at least one branch at (3) is F; a branch at (3) is F iff the corresponding branch at (4) is T. If a branch at (4) is T, the corresponding branch at (3) is F, so the universal at (2) is F, and the negation at (1) is T. That is,  $\sim \forall x \sim Nx$  is T just in case there is some  $o$  such that  $Nx \llbracket o/x \rrbracket$  is T. And this is precisely the condition under which  $\exists x Nx$  is T. Similarly, suppose we accept both,  $\exists x(x \text{ is a number})$  and  $\forall x[(x \text{ is a number}) \rightarrow \exists y((y \text{ is a number}) \wedge (y \text{ is greater than } x))]$ . The first commits us to the existence of just one number. Taken by itself, the second commits us to the existence of nothing. But—given natural assumptions about “greater than”—taken together, the two commit us to the existence of infinitely many numbers: Suppose some thing  $a$  makes the first true; then the second requires that there is a  $b$  greater than it; but then another application of the second requires that there is a  $c$  greater than  $b$ ; so another application of the second requires that.... The upshot of these examples is that, at least on its “negative” side, (EQ) is mistaken. The suggestion that we can simply scan a collection of canonical statements and directly “read off” ontological commitments by identification of existential quantification, is wrong.

This suggests a move to something more like Quine's "standard" formulation of his criterion of ontological commitment. Quine is notorious for having stated multiple, not entirely equivalent formulations of his criterion. However, for our purposes, it is clear enough what he has in mind. On e10 of "On What There Is" he says,

We are convicted of a particular ontological presupposition if, and only if, the alleged presuppositum has to be reckoned among the entities over which our variables range in order to render one of our affirmations true.

Or, at the top of the next page, and perhaps more perspicuous,

Q1 A theory is committed to those and only those entities to which the bound variables of the theory must be capable of referring in order that the affirmations made in the theory be true.

We know what bound variables are. The question is what things have to be assigned to variables, in order to make our sentences true. If one or more of our sentences would be false apart from the existence of some objects, then we are committed to the objects. If all our sentences would remain true without the objects, then we are not committed to them. The Q1 formulation is not tied to any particular sentence *form*. Given the examples of the previous paragraph, there is no problem admitting that  $\sim\forall x\sim(x \text{ is a number})$  commits us to the existence of a number, and that  $\exists x(x \text{ is a number})$  with  $\forall x(x \text{ is a number} \rightarrow \exists y(y \text{ is a number} \wedge y \text{ is greater than } x))$  commit us to infinitely many. And Q1 is compatible with Quine's response to Plato's beard:  $\sim\exists x(x \text{ is-Pegasus})$  is true though there is no Pegasus among the objects assigned to 'x'; so it doesn't commit us to Pegasus. So far, perhaps, so good.

In the next chapter, we'll take up some reasons why Quine himself might be uncomfortable with the Q1 formulation. For now, though, let's treat the formulation as clear enough, and explore consequences.

## *II. Plato's Stubble*

Q1 says that a theory is committed *only* to those entities to which the bound variables of the theory must be capable of referring in order that the affirmations made in the theory be true. That is, the only things to which we are committed when we accept a theory, are the ones that must be "fed" into the relevant functions if corresponding statements are to be true. But what about the functions themselves? Unless there *are* functions, it's hard to see how objects could be fed into them. If it is the case that the use of a predicate requires that there be a corresponding function, then a statement like ' $\exists x(x \text{ is happy})$ ' requires not only that there be at least one happy thing, but also that there be the (x is happy) and existential functions. This, of course, is inconsistent with Q1, for Q1 requires just that there be the happy thing.

As it turns out, McX thinks there are entities such as attributes, relations, classes, and functions, and thinks that we express or refer to them with predicates. Perhaps a *property* or *attribute* simply *is* a function with “output” T for any object that has it and F for any other. At any rate, if a *predicate* works by means of a property, then ‘x is a property’ works by means of something that returns T when applied to itself. But if this is so, Quine can’t be right if he says, “there are no properties (attributes, functions, or whatever).” For to say this is to say  $\sim\exists x(x \text{ is a property})$ ; this is the case only if (x is a property) returns F for any object; and this is the case only if (x is a property) returns F when applied to itself. Once again, then, Quine seems to be at a disadvantage when advocating a negative position according to which some entities do not exist. This time, the canonical notation is no help. If anything, the notation only exacerbates the problem; for, as we have seen, it depends heavily on predicates. Thus we seem to have stumbled upon a stubble remaining from the supposedly razed Plato’s beard.

But we need to see *why* McX thinks predicates refer to properties (attributes, functions, or whatever). Given our presentation in the previous chapter, it may seem obvious that Quine requires such entities. But there are other ways to develop notations like  $\mathcal{U}$ . If we had said—as some do—that *formulas* are T or F, depending on whether *predicates* apply to objects, etc. we might have seemed to depend just on ordinary things along with bits of language—so that there would be no obvious appeal to functions (as such) in the account of the language. In this section, we’ll explore ways McX and Quine defend their respective positions. After a preliminary version of the debate about properties, I turn to an extended response on behalf of McX, and Quine’s reply. We won’t resolve the debate between Quine and McX here. Still, strangely perhaps, the *way they argue* will enable us to reach positive conclusions about the status of Q1.

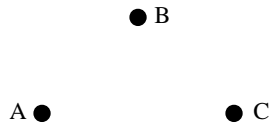
(A) Since Plato, various philosophers have thought that predicates pick out something *common* among things. McX is a philosopher of this sort. According to Quine, speaking of attributes, McX says,

There are red houses, red roses, red sunsets; this much is prephilosophical common sense in which we must all agree. These houses, roses, and sunsets, then, have something in common; and this which they have in common is all I mean by the attribute of redness (e8).

According to McX, then, attributes are *universals* insofar as they may be shared by multiple objects. Since they are *shared* by particular objects, they aren’t identical with any of the particular objects that share them. But, if they aren’t identical with ordinary concrete objects, there is pressure to think of them as abstract. Plato lives!

But let’s think about this reasoning more carefully. Since red is typographically difficult, let me switch to black. Consider these three dots:





It is natural to think that the color of A is the *same* as the color of B, the distance between A and B is the *same* as the distance between B and C, etc. Given this, McX argues,

- (1) If the color of A is the same as the color of B, then the color is a universal.
  - (2) The color of A is the same as the color of B.
- 
- (3) The color is a universal.

And similarly for distances, etc. This argument is valid in the sense that *if* the premises are true, then the conclusion is true as well. Are the premises true?

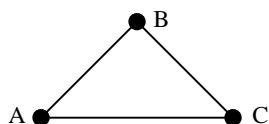
It's natural to reply that the color of A isn't literally the *same thing* as the color of B. Rather, it's only that each dot has a particular color, and the color of the one exactly *resembles* the color of the other. And similarly, the distance between A and B exactly resembles the distance between B and C, etc. This sets up a "two-pronged" reply to McX's argument. On the one hand, if "is the same as" means "is the same *thing* as," then we may grant (1), but deny (2): If the color of A is the same thing as the color of B, the color is indeed a universal; so we grant (1). But if the colors merely resemble, it is not the case that the color of A is the same thing as the color of B. So, on this reading, (2) is false and the argument is unsound. On the other hand, if "is the same as" means "exactly resembles" then we may grant (2), but deny (1): By assumption, the color of A exactly resembles the color of B; so we grant (2). But, insofar as the things merely resemble, there is no reason to think they are universal. So, on this reading, (1) is false and, again, the argument is unsound. So McX is charged with equivocation. There are two ways to understand "is the same as." Understood one way (2) is false, and understood the other way (1) is false. Either way, the argument fails. From now on, let's understand "is the same as" as "is the same *thing* as" so that, officially, the charge is that (2) is false.

(B) Quine gives only a glimpse of McX's reasoning. So far, however, McX's position is remarkably similar to one Russell advocates in his, "The World of Universals." It's worth considering a reply which Russell (now in the role of McX) offers against the above objection.<sup>2</sup> Russell finds a problem about the resemblance suggestion—he thinks that if we adopt the position that the color of A merely resembles the color of B, then we are led into an unacceptable regress. Suppose, in response to the worry about whether A, B and C have the *same* color, we reply that the color of A merely

---

<sup>2</sup> Of course, insofar as Russell writes before Quine, this is sure to be anachronistic. "The World of Universals" is a chapter in Russell, *The Problems of Philosophy* (Oxford: Oxford University Press, 1912).

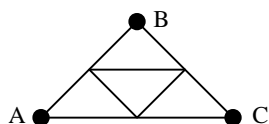
resembles the color of B, the color of B merely resembles the color of C, etc. That is, where the lines represent resemblances, we say,



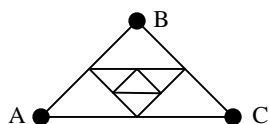
that there are color-resemblance relations between A, B and C as above. This is the basis on which we reject (2). Russell's strategy is to deflect this response—or, rather, to show that it introduces universals of its own. Working now with white patches, he says,

Since there are many white things, the resemblance must hold between many pairs of particular white things; and this is the characteristic of a universal. It will be useless to say that there is a different resemblance for each pair, for then we shall have to say that these resemblances resemble each other, and thus at last we shall be forced to admit resemblance as a universal. The relation of resemblance, therefore, must be a true universal. And having been forced to admit this universal, we find that it is no longer worth while to invent difficult and unpalatable theories to avoid the admission of such universals as whiteness and triangularity (e18).

How does this work? Well, given resemblances between A, B and C as above, Russell wants to know whether the resemblance relation between A and B is the *same relation* as the relation between B and C, etc. If it is, then the relation is a universal. So suppose we stick to our guns and insist that the relations merely resemble one another. That is, where the new lines represent resemblances, we say,



that there are resemblances among the color resemblances as above. What is the problem with this? Russell wants to know whether the new resemblance relations are the *same*. If they are, there is a universal. So suppose they merely resemble one another. Then we seem committed to the new resemblance relations:



Etc. Insofar as it is supposed to avoid commitment to universals, the “resemblance” strategy seems to commit us to an infinite series of resemblances.

Say this is right. Is it bad? Many have thought so. That is, Russell seems to argue,

- (4) Either there is a resemblance universal, or there is this infinite series of resemblance relations.
  - (5) This infinite series of resemblance relations is impossible.
- 
- (6) There is a resemblance universal.

And, on Russell's view, if there is one universal, there are likely to be more—so that there is at least room for color universals as above. Again, the argument is valid in the sense that *if* the premises are true, then the conclusion must be true as well. Are the premises true? We have just seen reasons to accept (4). What about (5)? Here, things are less clear. One might think that the very idea of an actual infinite series is absurd.<sup>3</sup> But this isn't obvious. Perhaps time is finite. But is it obviously *impossible* that every time have one before it? Aristotle, at least, did not think so. Or, again, perhaps space is somehow quantized. But is it obviously *impossible* that every space have one smaller than it? Many have thought that this is obviously the case! And, similarly, one might think that any resemblance relations are themselves related by further resemblance relations—that there can be an infinite series of resemblance relations.

But Russell's series isn't just any series. It's supposed to be an infinite *regress*, and to be rejected for that reason. Infinite regresses are not, I think, very well understood. But perhaps we can say enough to make progress here. First, drawing on a proposal from chapter 1, suppose that (i) groups are things and (ii) for any thing, there is a group with it as a member. Then, given the existence of Bill, there is a series,

Bill, {Bill}, {{Bill}}, {{{Bill}}}, {{{{Bill}}}}, ...

with infinitely many members. If Bill is a thing then, by (ii), there is a group {Bill} with Bill as a member; since {Bill} is a group, by (i), it is a thing; so by (ii), there is a group {{Bill}} with {Bill} as a member; etc. Supposing that the groups are all distinct, there are infinitely many. Despite this, there is nothing obviously inconsistent about premises (i) and (ii). In fact, (i) and (ii) are an integral part of standard set theory in mathematics. But this series isn't what one ordinarily thinks of as an infinite *regress*—it might be called an infinite “progress.” In this case, one member of the series leads to or guarantees the next. But in a regress, one member of the series *results from* or is *guaranteed by* the next.

---

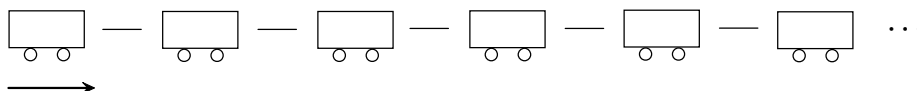
<sup>3</sup> Some arguments to this effect are explored by W. L. Craig in, *The Kalām Cosmological Argument* (London: Macmillian Press, 1979), and other places. I am inclined to think that these arguments fail. Medieval Arabic philosophers, in particular, develop interesting problems about actual infinities as such. But, *pace* Craig, these seem resolved with the rise of contemporary, Cantorian, mathematics.

For a regress, consider first an infinite series of cats stalking a mouse. Suppose any cat in the series bites the tail of the cat/mouse in front of it iff its tail is bitten by the cat/mouse behind.



(In case of worry about cruelty to animals, substitute some mechanical devices which “bite” iff bitten.) Now suppose the leftmost cat bites the mouse’s tail; then its tail must be bitten by the cat behind. So the second cat bites the tail in front of it, and its tail must be bitten from behind; so the third cat bites the tail in front of it; etc. We have a progress if the cats are faced the other way, and the mouse bites the tail of the leftmost cat; then the leftmost cat bites the one in front of it, who bites the one in front of it, etc. But as above, the leftmost cat’s biting results *from* the next cat’s biting, etc. Thus it’s a regress. “Inadequacy” in one member—its inability to bite on its own—leads to the conclusion that there is another member; inadequacy in that member, leads to the conclusion that there is another member, etc. Is such regress impossible? It’s mind-boggling! But, I claim, not impossible. Notice that, if some cats are arranged in a circle, either each bites, or none bites. And, plausibly, in an infinite series, either each bites or none bites—where the options are equally possible. God could create the series of cats “all at once” in either state. All that is required for biting is that each cat be in proximity to one that bites—where, seemingly, god could create them all that way. And similarly for not biting. If this is right, then the regress is possible. But contrast this case with the next.

Consider a series of traincars each of which moves iff it is moved by the one in front. Then, supposing that the leftmost car moves, there is an infinite series,



where each car moves. Since the leftmost car moves, it is moved by the car in front of it; since the second car moves, it is moved by the one in front of it; etc. Clearly, it’s a regress. “Inadequacy” in one member—its inability to move on its own—leads to the conclusion that there is another, etc. Is the regress impossible? This time it is. Notice that, if some cars are arranged in a circle, assuming friction and a level track, they don’t move. Similarly, even an infinite series of cars won’t move. It’s natural to think that if any series of traincars moves, then it has a member that moves on its own (is a locomotive).<sup>4</sup> The total force on the train is equal to the sum of the forces from the cars;

<sup>4</sup> Imagine that each car is only half the weight of the car it pulls. Then the weight of the entire train is finite, and we can sensibly imagine that a locomotive could pull it. For this, recall Zeno’s paradox of dichotomy: if one goes half some distance, half the remaining, etc., after any  $n$  steps,  $1/2^n$  the original distance remains. Thus, where our train has  $n$  cars, with the first  $1/2$  some unit of weight, its total weight  $W$

but if the force from each car is zero, the total force on the train is zero, and the train doesn't move: the inability of the individual cars to move on their own, transfers to the series as a whole, with the result that the cars do not move.

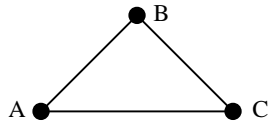
So far, we've distinguished a series that is a *progress* from one that is a *regress*. And it is traditional to distinguish regresses that are *vicious* from those that are *benign*. A vicious regress is impossible, a benign regress is not. Roughly, I suggest that we have infinite *regress* when there is some purported fact about a series (the leftmost cat bites, the leftmost car moves), and some purported facts about its members, which result in conflict for every finite case—so that any finite series is inadequate to support facts of both sorts. Thus, no member of a finite series of cats bites insofar as an “end cat” remains unbitten. And no member of a finite series of traincars moves because no member pulls the next (no member is a locomotive). We have *vicious* regress when claims about the series and the members conflict even in the infinite case. Thus, tension goes away with an infinite series of cats because, where the series infinite, there is no cat without another to bite its tail. But inconsistency remains with an infinite series of traincars, as there is still no locomotive. In the vicious case, we somehow “collect” contributions from the members of the series, where the collection is inadequate to support some purported feature of the series. In the benign case, we do not “collect” contributions in the same way; rather members of the series are characterized by a “self-sustaining” property, and the inadequacy which drives the regress is localized at the end. On this picture, then, the regress of cats is benign, and the regress of traincars is vicious.<sup>5</sup>

Perhaps the difference between the regress of biting cats and the regress of moving traincars remains difficult to grasp. Even so, Russell's series of resemblance relations may seem not a regress at all! Suppose it is established that our original three resemblances,

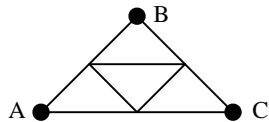
---

$= 1/2 + 1/4 + \dots + 1/2^n = 1 - 1/2^n$ ; and  $W$  approaches 1 as  $n$  approaches infinity.

<sup>5</sup> This is a rough version of a proposal I develop further in “What's So Bad About Infinite Regress?” (currently available at <http://philosophy.csusb.edu/~troy/regress-pap.pdf>). For another case which may help to see how the inadequacy which drives a series may or may not transfer to the whole, consider some blocks on a scale, where each block weighs half the one underneath it. If the scale reads 1 pound, and the first block weighs .5, then there are infinitely many blocks on the scale: the first weighs .5 and  $.5 < 1$ , so there is another; the first two weigh .75 and  $.75 < 1$ , so there is another; etc. But  $1/2 + 1/4 + 1/8 \dots = 1$ ! So the series of blocks weighs 1, and the regress is benign. Notice: in this case we “collect” contributions from members, but the collection is adequate just in the infinite case. Of course, if the scale reads 1.01 pounds, and the first block weighs .5, then there is a regress—where the series doesn't sustain the fact, so that the regress is vicious.



are in fact resemblances. Then it seems uncontroversial that they resemble one another. That they are resemblances, guarantees the next set of resemblances,



Etc. That is, this series seems to be a *progress* rather than a *regress*. It's the existence of the first relations, which leads to the next, etc. So far as his discussion in "The World of Universals" goes, one *might* think that this is all there is to his argument, and that he therefore gives us no reason to accept that the infinite series of resemblance relations is impossible (premise 5). And if there is room to reject this, then it would seem that the resemblance strategy enables us still to deny that the color of A is the same (thing) as the color of B (premise 2). So we aren't given an adequate reason to accept that there are universals.

However, it is possible to challenge this account of his argument. In another place (published around the same time as "World of Universals"), Russell suggests a different way to think about the series of resemblance relations.

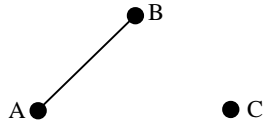
We may take a standard particular case of color-likeness, and say that anything else is to be called a color-likeness if it is exactly like our standard case. It is obvious, however, that such a process leads to an endless regress: we explain the likeness of two terms as consisting in the likeness which their likeness bears to the likeness of two other terms, and such a regress is plainly vicious. Likeness at least, therefore, must be admitted as a universal, and, having admitted one universal, we have no longer any reason to reject others.<sup>6</sup>

In both places, Russell is thinking of Berkeley and Hume and their attempts to account for generality by "standard" or "representative" particulars. On the resemblance view, *what makes* a color patch black (or what makes 'black' apply to a color patch) is its resemblance to a standard color patch. And, similarly, what makes one relation a color resemblance is its relation to some standard color resemblance. Russell claims that *this* leads to a regress that is "plainly vicious."

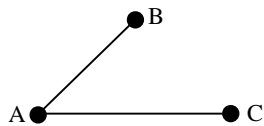
---

<sup>6</sup> Russell, "On the Relation of Universals and Particulars," *Proceedings of the Aristotelian Society* 12 (1911-1912): 9.

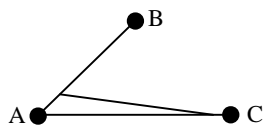
In this case, the point is not merely that the resemblances resemble one another, but rather that they must resemble the standard resemblance. Suppose dot A is the standard black patch, and the relation between A and B is the standard color resemblance.



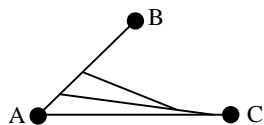
We want to know whether C is the same color as A. On the resemblance view, this depends on whether the color of C resembles the color of A. So consider the relation between the color of A and the color of C.



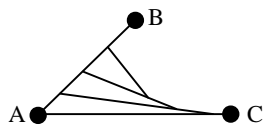
The colors count as the “same” if the relation between A and C is a color resemblance and not the same if it isn’t. But whether it is a resemblance, depends on the relation between *it* and the standard resemblance.



If this new relation is a resemblance, then the relation between A and C resembles the relation between A and B, so that the relation between A and C is a resemblance, and C is the same color as A. But whether this new relation is a resemblance itself depends on a relation between *it* and the standard resemblance.



And whether *this* is a resemblance depends on the relation between it and the standard resemblance.



Etc. Clearly, this is a regress. What makes one relation a resemblance, is the next; what makes the next a resemblance, is the one after that; etc. So it's a regress. It is less clear, perhaps, whether it is vicious. Russell merely asserts that it is. Uncertainty about this case may derive from obscurity in the resemblance proposal itself. However, we may suppose that, as with the train, the relations all sit "idle" apart from something that is a resemblance on its own (a resemblance locomotive) to make them go. Given this understanding of the infinite series, Russell seems to have provided a reason to accept that the infinite series of resemblance relations is impossible (his premise 5), and so to accept that there is a universal.

(C) Resemblances are offered as an alternative to universals. Russell rescues the argument for Platonism by arguing that the resemblance alternative doesn't work. Quine rejects universals—thus he denies that the color of A is the same thing as the color of B. But he also denies that resemblances are required to account for sameness of color. McX and the resemblance theorist each try to say *what makes* the different dots black. McX offers one answer, the resemblance theorist another. But Quine denies the need to respond. Thus Quine denies (2) in McX's argument and (4) in Russell's. In a famous passage, he not only denies that the dots have anything in common, but argues that he doesn't have to explain what makes them both black.

One may admit that there are red houses, roses, and sunsets, but deny, except as a popular and misleading manner of speaking, that they have anything in common. The words 'houses', 'roses', and 'sunsets' are true of sundry individual entities which are houses and roses and sunsets, and the word 'red' or 'red object' is true of each of sundry individual entities which are red houses, red roses, and red sunsets; but there is not, in addition, any entity whatever, individual or otherwise, which is named by the word 'redness', nor, for that matter, by the word 'househood', 'rosehood', 'sunsethood'. That the houses and roses and sunsets are all of them red may be taken as ultimate and irreducible, and it may be held that McX is no better off, in point of real explanatory power, for all the occult entities which he posits under such names as 'redness' (e8).

Quine doesn't here distinguish the "same thing" and "exact resemblance" strategies. However, I think we may take him as denying not only that the houses, roses, and the like have anything in common, but also that their being red depends, in any way, on their resembling one another. No doubt, he would agree that they *do* resemble one another—he only denies that their being red depends on resemblance.

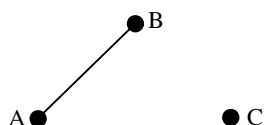
There are at least two things going on here. First, just as a person who asserts that there *are* such-and-such trends may mean no more than that people behave in certain complex ways, or a person who asserts that there *is* a prime number greater than 5 and less than 9 may assert only that some premises have certain consequences, so Quine offers an *alternative* to saying that there *is* something the roses, sunsets, etc. have in



common. Saying that the roses, sunsets, etc. have something in common just *is*, according to Quine, saying that they are all red. Or, rather, the truth condition for a claim according to which they have something in common is that they all be red. And, similarly, saying that there *are* color resemblances between the roses, sunsets, etc. just *is*, according to Quine, a way of saying that they are all red. The truth condition for the claim that they have something in common is not, what one might think, that  $\exists x(\text{roses, sunsets and the like have } x \text{ in common})$  but rather something more like,  $\exists x(x \text{ is a rose} \wedge x \text{ is red}) \wedge \exists x(x \text{ is a sunset} \wedge x \text{ is red})$ . Admitting the data, then, Quine thinks he is committed to the roses and sunsets, but not to any thing that they have in common.

Second, Quine denies that he has to explain what makes the different things red. Contrast a medieval who explains motion of the planets by their “motive power” with a Newtonian who explains their motion by “gravitation.” Perhaps, through his mathematical characterization of gravitation, the Newtonian explains and unifies phenomena which the medieval does not. But the idea of a gravitational force seems, at bottom, no less mysterious than that of a motive power. Each is, in its respective theory, primitive and unexplained. Perhaps gravitational forces are explained in general relativity. But this theory has primitives of its own. Thus, on the one hand, it is no knock on a theory that it has some primitive notions—every theory has them. On Quine’s view, that the houses, roses and sunsets are all red is primitive. That he has a primitive can’t be held against him. On the other hand, theories will count as better or worse insofar as they do or do not explain and unify phenomena which others do not. One may even have a view about what *sort* of thing is or can be primitive—and want to explain other phenomena in terms of those that *should* be primitive<sup>7</sup>. Quine thinks his view lacks no “real” explanatory power which McX’s has and, presumably, that his primitives are located in just the right place.

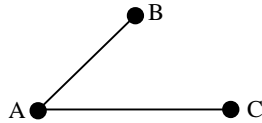
To understand Quine’s proposal, it may help to compare the way a (modified) Platonist could respond to a regress closely related to the one we’ve just seen. McX thinks there *are* properties and relations like *redness* and *being the same height as*, and that ordinary things like roses and people *have* or *instantiate* them. Suppose dot A is a color property, and the line between A and B is the *instantiation* relation.



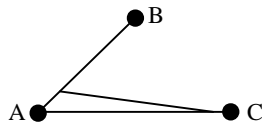

---

<sup>7</sup> On this, consider Epicurus’s atomic theory on which all the atoms were initially uniformly moving in a single direction. As a result of a “swerve” collisions were introduced into this system, resulting in the world with all its complexity. The atoms, initial motion, and swerve are all primitive. But one might think the atoms and initial uniform motion are legitimate as primitives, in a way that the swerve is not.

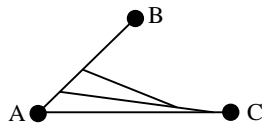
Then dot C has property A just in case C instantiates A—just in case the relation between A and C is an instance of the instantiation relation. So consider the relation between A and C.



The relation between A and C is an instance of instantiation, just in case *it* instantiates *instantiation*. So consider the relation between *it* and *instantiation*.



This new relation is an instance of instantiation just in case *it* instantiates instantiation, so we have,



etc. As in the resemblance case, this regress looks vicious against the Platonist if we think that a relation is an instance of instantiation just *because* it instantiates *instantiation*.<sup>8</sup>

But a (modified) Platonist might respond as follows: If you are six feet tall, and I am six feet tall, the way we are, individually, guarantees that we are the *same height*. The fundamental facts seem to be our individual heights, and the sameness arises because of the way we are individually. Similarly, a Platonist *might* say that a property is one way, and a thing is another; relations between them arises *because* of the way they are individually. In particular, on this view, the relation of *instantiation* arises between a thing and a property because of the way they are individually. Given this, our series appears as a progress, not a regress. The relation between A and C is an instance of instantiation *because* of the way A and C are—not because of some further relation to *instantiation*. There are further relations among the relations. But, again, the fundamental or primitive facts seem to involve A and C—and instantiation facts arise

---

<sup>8</sup> This, and the parallel relation regress are examples of what D. Armstrong calls the “fundamental relation” regress. See, e.g., Armstrong, *Universals: An Opinionated Introduction* (Boulder: Westview Press, 1989). It’s interesting that Russell seems not to have recognized that his own argument could be revamped to apply against Platonism.

because of the way they are. Insofar as one fact leads to the next, rather than arising from it, there is progress, not regress.<sup>9</sup> Now, we may see Quine as saying very much the same thing—only less. He agrees that patches may *be* red or black, and that other facts may arise because of the way they are. But he wants to say this without admitting any entities that *are* properties, relations or resemblances. Perhaps Quine allows that ‘black’ applies to things which resemble in a certain way. But that they resemble is to be explained by the way the things are—by their being black—and not by the existence of any “extra” things which are resemblances.

We find Quine using the same strategy in his discussion of meanings. McX suggests that words *have* meanings and that these may very well be attributes, properties, or the like—and so be universal. In response, Quine allows that words may be meaningful or significant, but denies that any things are meanings which words or utterances have. The truth condition of “This utterance has a meaning” isn’t that  $\exists x(x \text{ is a meaning} \wedge \text{this utterance has } x)$  but rather that  $\exists x(x \text{ is this utterance} \wedge x \text{ is meaningful})$ . Admitting the data, Quine is committed to the utterance, but not to the meaning. Similarly, what may seem more difficult, the truth condition for the claim that some utterances have the *same* meaning just amounts, for Quine, to the claim that they are synonymous. The truth condition for ‘this utterance has the same meaning as that’ isn’t that  $\exists x(x \text{ is a meaning} \wedge \text{this utterance has } x \wedge \text{and that utterance has } x)$  but that  $\exists x \exists y(x \text{ is this utterance} \wedge y \text{ is that utterance} \wedge x \text{ is synonymous with } y)$ . Quine takes some utterances’ being meaningful or synonymous as his primitive. And these primitives are used explain supposed reference to meanings. And, according to Quine, the “explanatory value of special and irreducible intermediary entities called meanings is surely illusory” (e12).

(D) We’re in the midst of an interesting and complex debate. Our debates about meanings and properties have turned into debates about what is supposed to explain what, and about the explanatory power of different theories. McX takes properties and meanings among his primitives and explains a thing’s being red or meaningful in terms of them. Quine takes a thing’s being red or meaningful as primitive, and explains talk about supposed properties and meanings on that basis. There’s a legitimate question about which theory has the right sort of explanatory power and which finds its primitives in the right sort of place. We’ll come back to properties in the last part of this book. For now, though, let’s cut this debate short, and notice something about it’s nature.

Notice: *the “stubble” argument has played no role in our discussion of properties.* Rather, McX and Quine seem to be arguing about what our *quantifiers* must range over in order to make sense of our use of predicates. If the truth condition for ‘this utterance has the same meaning as that’ is that  $\exists x(x \text{ is a meaning} \wedge \text{this utterance has } x \wedge \text{and that}$

---

<sup>9</sup> As we saw in consideration of the the first resemblance series from p. 81, a similar reply seems open to the resemblance theorist: The dots are both black when they resemble one another; but they resemble one another because of the way they are, not because of some further resemblance relation.

utterance has  $x$ ), we are committed to meanings. If the condition is merely that  $\exists x \exists y (x \text{ is this utterance} \wedge y \text{ is that utterance} \wedge x \text{ is synonymous with } y)$  then, so far, we don't seem forced to admit the meanings. That is, both McX and Quine seem to accept something like,

- Q1' A theory is committed at least to those entities to which the bound variables of the theory must be capable of referring in order that affirmations made in the theory be true.

McX attempts to use this principle to force us to recognize that there are entities corresponding to predicates, and Quine attempts to show that we are forced into no such position. So the debate takes place on grounds Quine can accept—if the negative side of Quine's criterion is mistaken, it would seem that we have to use the positive side to show it. It would seem, then, that on its negative side Quine's criterion of ontological commitment is correct in at least this sense: the original source of our commitments comes through quantifiers, not through predicates. Thus, whatever may be the outcome of this debate about properties, again, Quine seems to have found a “neutral” mode of expression which makes the debate possible.

### *III. Preliminary Applications*

Observe, then, that we've been *using* Quine's criterion to think about how it works. In this section, I take up preliminary discussion of some further cases intended to show how the method sets up metaphysical debate, and the shape such debate may take. We have already had something to say about how the method is supposed to work. But the examples should help to integrate what has come before.

(A) Quine himself begins with some simple examples involving dogs and zoölogical species.

'Some dogs are white' says that some things that are dogs are white; and, in order that this statement be true, the things over which the bound variable 'something' ranges must include some white dogs, but need not include doghood or whiteness. On the other hand, when we say that some zoölogical species are cross-fertile we are committing ourselves to recognizing as entities the several species themselves, abstract though they are. We remain so committed at least until we devise some way of so paraphrasing the statement as to show that the seeming reference to species on the part of our bound variable was an avoidable manner of speaking (e10).

The example about dogs rehearses what we have already seen.  $\exists x (x \text{ is a dog} \wedge x \text{ is white})$  is true if and only if there is at least one white dog. So anyone who accepts that  $\exists x (x \text{ is a dog} \wedge x \text{ is white})$  is committed to the existence of at least one white dog. But, if the

conclusions of the last section were correct, accepting that  $\exists x(x \text{ is a dog} \wedge x \text{ is white})$  doesn't itself commit us to abstract entities like doghood or whiteness. Presumably, though he denies that there are entities such as doghood and whiteness, Quine accepts that there are dogs. So this case is no problem for Quine.

In contrast, 'some zoölogical species are cross-fertile' seems a problem. If its truth condition is that  $\exists x \exists y (x \text{ is a zoölogical species} \wedge y \text{ is a zoölogical species} \wedge x \neq y \wedge x \text{ is cross-fertile with } y)$  then, if it is true, among the things that are fed into the functions are at least two zoölogical species. Perhaps we are inclined to agree that some things are dogs, cats, zebras and rhinoceroses, but to deny that there are any things in addition to these which are their *species*. Quine holds that 'some zoölogical species are cross-fertile' commits us to the existence of species at least until we find some "paraphrase" which does not require them. One might argue about whether "paraphrase" is required or not—recall our discussion from the beginning of chapter 3. And one might argue about whether the English 'some zoölogical species are cross-fertile' itself commits us to species. But it seems clear that anyone who accepts that some zoölogical species are cross-fertile, and that its truth condition is  $\exists x \exists y (x \text{ is a zoölogical species} \wedge y \text{ is a zoölogical species} \wedge x \neq y \wedge x \text{ is cross-fertile with } y)$  commits herself to the existence of species as such. A metaphysician who accepts that some zoölogical species are cross-fertile but denies that there are species, is faced at least with the challenge or task of producing an account of truth which does not commit to species. And if she admits that there are species, then she is faced with the task of saying what they are—perhaps, as Quine suggests, they are abstract!

But giving such an account isn't all that easy to do. If we wanted to say that dogs and cats are cross-fertile, we might go with ' $\exists x \exists y (x \text{ is a dog} \wedge y \text{ is a cat} \wedge x \text{ is cross-fertile with } y)$ '. This commits us only to the existence of one dog and one cat. So we aren't yet committed to species as such. To get that some zoölogical species are cross-fertile, we might try  $(\text{dogs and cats are cross fertile}) \vee (\text{dogs and mice are cross fertile}) \vee (\text{horses and donkeys are cross fertile}) \vee \dots$  for all the pairs of species there are. Of course, this might get a bit lengthy, particularly if there are, say, infinitely many species of angels or extraterrestrials. As an alternative, we might try something like, ' $\exists x \exists y \exists x_1 \exists y_1 (x \text{ is a species property} \wedge y \text{ is a species property} \wedge x \neq y \wedge x_1 \text{ has } x \wedge y_1 \text{ has } y \wedge x_1 \text{ is cross-fertile with } y_1)$ '. This depends on just whatever species properties there happen to be. But it commits to species properties. Here's another option: Drawing on a proposal of Goodman's from chapter one, we might say that the species *dog* just is the complex thing whose parts are all and only the dogs—and similarly for other species.<sup>10</sup> Then we may be unembarrassed about including zoölogical species among the things that are fed into our functions. Intuitively, a *species* doesn't change every time a member of it is born or dies. On this proposal, however, a species is a *different* thing every time a new member is born

---

<sup>10</sup> I thank students of PHI 380, at CSUSB in Fall 1997, for this suggestion. I think Jason Jones was a primary defender of this view, though others may have slipped my mind.

or dies. But maybe our notion of “thing” is still too rigid: Perhaps, like ordinary living things, zoölogical species are complex entities which remain the same even though parts are constantly being added and subtracted. So we admit zoölogical species as concrete entities without shame.<sup>11</sup>

So far as the method itself is concerned, what’s important is not the solution to the problem, but the overall dialectical situation: One might have thought that we were committed to abstract entities, or the like, when it seemed that we were committed to the existence of zoölogical species. Maybe we are so committed. But it’s not automatic. The method forces us to look for whatever things *really* make the original claims true. If abstract species are involved, so be it. But if abstract species aren’t involved, we are pressed to offer some alternative—and the notation thus presses us to be completely explicit about what things we think there are.

(B) Given all Quine has done to keep his “desert landscape” clean and pure, his treatment of mathematics may come as something of a surprise. In the end, he treats Platonistic mathematics as a powerful theory which is at least deeply integrated with ordinary science. Insofar as we accept ordinary science, there are therefore strong pressures to accept the mathematics. At the same time, as Quine suggests, classical Platonistic mathematics is “up to its neck in commitments to an ontology of abstract entities” (e11). As we have seen, one might end up with something like classical mathematics simply by observing that *mathematicians* require numbers—as when they say that  $\exists x(x \text{ is a prime} \wedge x \text{ is greater than a million})$ , etc. But, for the link with science, we might do better with something like, ‘the average American has 3.2 children’. Surely, we don’t want this to commit us to the existence of an average American with 3.2 children—as,  $\exists x(x \text{ is an average American} \wedge x \text{ has 3.2 children})$ . Rather, it should turn out merely that the total *number* of Americans divided by the total *number* of their children is 3.2. But this seems to commit us to *numbers*. We get the same kind of result from an even simpler case:

Suppose I want to say of some infant, Fred, that he is one foot high. That is, let’s suppose that,

(7) Fred is one foot high.

is true. As its truth condition, we might offer,

(8)  $\exists x(x \text{ is-fred} \wedge x \text{ is one foot high})$

This commits us to nothing but the existence of Fred. So far, so good. But the claim about Fred’s height does not exist in isolation. Presumably, we are offering a “complete”

---

<sup>11</sup> Or at least we put off problems for others related to concrete entities to be considered in part IV!

theory of the world—or as complete a one as we can—and, given that (7) is in our theory, we would expect something corresponding to, say,

$$(9) \quad \exists x(x \text{ is-fred} \wedge x \text{ is less than two feet high})$$

as well. But (8) and (9) are not independent. Part of the data is that the one follows from the other. And recall from the beginning of chapter 3 that we expect to retain entailments. So we might include in our theory that anything which is one foot high is less than two feet high.

$$(10) \quad \forall x(x \text{ is one foot high} \rightarrow x \text{ is less than two feet high})$$

Now it will be part of our theory that (8) and (9) are appropriately linked. But we presumably require many such principles corresponding to relations among all the different heights. And something about the resultant account would seem to miss the generality of the situation. When we learn arithmetic, we seem to reason generally about such contexts. To put it another way, our theory seems to leave out an important sort of unity and so to lack an important sort of explanatory power. As such, it is vulnerable to a theory without these weaknesses.

Here is how Quine removes the difficulty: he quantifies over numbers.<sup>12</sup> Thus the truth condition for (7) becomes something like,

$$(11) \quad \exists x \exists y (x \text{ is-fred} \wedge y \text{ is-one} \wedge y \text{ is the height-in-feet of } x)$$

where the thing that is-one is a number. Given this, we should be in a position to conclude that Fred is less than two feet tall—if we have a theory of numbers that tells us that  $1 < 2$ . In fact, we will be able to conclude that he is less than any number  $> 1$  feet tall if we have the corresponding fact about numbers.

So let's delve a bit into a corresponding theory for numbers. A thing's height-in-feet may be an arbitrary non-negative real number (there is a real number corresponding to each point on a number line). Let '*a*', '*b*' and '*c*' be variables which range over such numbers—that is, we take  $\forall a \mathcal{P}$  as short for  $\forall a(a \text{ is a real number} \rightarrow \mathcal{P})$ , and  $\exists a \mathcal{P}$  as short for  $\exists a(a \text{ is a real number} \wedge \mathcal{P})$ , etc. Suppose we accept the following theses:

$$(12) \quad \text{Any two numbers have a unique sum:} \\ \forall a \forall b \exists c [Sabc \wedge \forall y (Sabc \rightarrow (y = c))]$$

---

<sup>12</sup> See, e.g., p. 12 of Quine, "Things and Their Place in Theories," in Quine, *Theories and Things* (Cambridge: Harvard Univ. Press, 1981). In the end, though, he eliminates direct appeal to numbers, by thinking of them as certain abstract groups – or, rather, his view may be that the numbers just *are* the abstract sets. Better source??

Since any two numbers have a unique sum, we won't go wrong by setting up "names" for their sum – and things will be much simpler if we do. Actually, we'll introduce a *function*  $(x + y)$  such that with  $a$  and  $b$  assigned to the variables, the sum of  $a$  and  $b$  is taken as the output of the function and, as the output, is assigned to the symbol, ' $(x + y)$ '. Similarly,

- (13) Any two numbers have a unique product:  

$$\forall a \forall b \exists c [Pabc \wedge \forall y (Paby \rightarrow (y = c))]$$

Since any two numbers have a unique product, we won't go wrong by setting up "names" for their sum. Again, we'll introduce a function  $(x * y)$  such that with  $a$  and  $b$  assigned to the variables, the product of  $a$  and  $b$  is the output of the function and, as the output, is assigned to the function symbol, ' $(x * y)$ '. Now we introduce some principles to describe the operation of these functions. In particular, the sum and product functions have distinct *identity* elements. Thus,

- (14)  $\exists a \forall b (a + b = b)$                        $\exists a \forall b (a * b = b)$                       *identity*

Again, given that there are such numbers (and that uniqueness follows with other properties), we won't go wrong by assigning them names – and things will be simpler if we do. So, holding with standard practice (!) let us use '0' to refer to the identity element for addition, and '1' to refer to the identity element for multiplication. Then (14) appears in the form,

$$\begin{array}{ll} \text{Any } b \text{ is such that } 0 + b = b & \text{Any } b \text{ is such that } 1 * b = b \\ \forall b (0 + b = b) & \forall b (1 * b = b) \end{array}$$

We require also of 0 and 1 that  $0 \neq 1$ . And, where ' $<$ ' is a two-place relation symbol (which may stand between terms in the way '=' does),

- (15) If  $a$  is less than  $b$  then for any  $c$  the sum  $a + c$  is less than the sum  $b + c$ .  

$$\forall a \forall b \forall c [a < b \rightarrow (a + c < b + c)]$$

The first principles (12) and (13) are properties of *closure* – any real numbers have a sum and product among the real numbers. (14) is among "field" properties of the reals and (15) among "order" properties. But the important point is just that quantifiers of the basic principles already range over real numbers.

These properties of the reals may themselves be derived from the construction briefly mentioned in chapter 1 (p. 11). Additional features are described in the boxed area below. Suppose as in the box, we have derived that  $0 < 1$ . Then we may reason as follows: Consider any  $x$  such that  $0 < x$ . By (15), with 1 in the place of  $c$ , we have that  $0 + 1$  is less than  $x + 1$ . But by (14),  $0 + 1$  is just 1. So 1 is less than  $x + 1$ . So for any  $x$  such that  $0 < x$ , 1 is less than  $x + 1$ . And since  $0 < 1$ , this gives us that 1 is less than  $1 + 1$ ,



### Properties of the Reals

Suppose the sum and product functions, along with the identity elements 0 and 1 are defined as in the text. Then the reals may be characterized by the following fundamental principles.

#### Field Properties

$a + b = b + a$	$a \cdot b = b \cdot a$	<i>commutivity</i>
$a + (b + c) = (a + b) + c$	$a \cdot (b \cdot c) = (a \cdot b) \cdot c$	<i>associativity</i>
$a + 0 = a$	$a \cdot 1 = a$	$0 \neq 1$ <i>identity</i>
$\forall a \exists b(a + b = 0)$	$\forall a[a \neq 0 \rightarrow \exists b(a \cdot b = 1)]$	<i>inverse</i>
$a \cdot (b + c) = a \cdot b + a \cdot c$		<i>distributive</i>

#### Order Properties

$(a < b) \vee (a = b) \vee (b < a)$	$(a = b) \rightarrow \sim(a < b)$
$(a < b) \rightarrow [a + c < b + c]$	$(a < b \wedge 0 < c) \rightarrow (a \cdot c < b \cdot c)$
$[(a < b) \wedge (b < c)] \rightarrow (a < c)$	

**Completeness Property:** Every non-empty subset of real numbers which is bounded above, has a least upper bound in the set of real numbers.

These are all the fundamental properties of the real numbers. But it can take some work to derive results from them! Let's identify the additive and multiplicative inverses of  $a$  as,  $-a$  and  $a^{-1}$ . Now consider the result from the text, that  $0 < 1$ .

First,  $b \cdot 0 = 0$ . By identity,  $b \cdot 0 = b \cdot 0 + 0$ ; by inverse,  $= b \cdot 0 + [(b \cdot 0) + -(b \cdot 0)]$ ; by associativity  $= [b \cdot 0 + b \cdot 0] + -(b \cdot 0)$ ; by distributivity  $= b \cdot [0 + 0] + -(b \cdot 0)$ ; by identity  $= b \cdot 0 + -(b \cdot 0)$ ; by inverse  $= 0$ .

Now, by identity  $0 \neq 1$ ; so from the first order property, either  $0 < 1$  or  $1 < 0$ . Suppose  $1 < 0$ . Then with  $-1$  for  $c$  in the order property for addition,  $1 + -1 < 0 + -1$ ; but by inverse  $1 + -1 = 0$  and by commutivity and identity  $0 + -1 = -1$ ; so  $0 < -1$ . So  $1 < 0 \wedge 0 < -1$ ; so with  $-1$  for  $c$  in the order property for multiplication,  $1 \cdot -1 < 0 \cdot -1$ ; but by commutivity and identity,  $1 \cdot -1 = -1$ , and by commutivity and the above result,  $0 \cdot -1 = 0$ ; so  $-1 < 0$ ; so we have  $0 < -1 \wedge -1 < 0$ ; so by the last order property,  $0 < 0$ ; but  $0 = 0$ ; so with the second order property,  $\sim(0 < 0)$ . Reject the assumption, it is not the case that  $1 < 0$ . So  $0 < 1$ .

This is not a test – you should not worry about it for the main content! It is here for your interest, and the central point about quantification, which you can see without the details.

which is just to say that  $1 < 2$ . So Fred is less than two feet tall. And similarly for related conclusions. So this bit of theory about numbers enables us to draw the right sort results from sentences like (7). Of course, this theory is committed to the existence of numbers.

But we are not *forced* to introduce theory about numbers. Instead of infinitely many principles like (10), we might try some general theorizing about *magnitude properties*, that is, about heights, distances, weights, and the like. Thus, e.g., say we adopt the following general principles. (If you have understood the previous proposal, this one will not seem particularly original!) Let ' $p$ ', ' $q$ ' and ' $r$ ' range over magnitude properties of a given type.

- (12') For any two magnitude properties, there is a unique property that is their sum:  

$$\forall p \forall q \exists r [Spqr \wedge \forall y (Spqy \rightarrow (y = r))]$$

And we'll let  $+$  be a function such that  $x + y$  has as output the unique sum of the magnitude properties assigned to  $x$  and  $y$ . Similarly,

- (13') For any two magnitude properties, some unique property is their product:  

$$\forall p \forall q \exists r [Ppqr \wedge \forall y (Ppqy \rightarrow (y = r))]$$

Again, we'll introduce a function  $(x * y)$  such that with magnitudes  $a$  and  $b$  assigned to the variables, the product of  $a$  and  $b$  is the output of the function. Again the sum and product functions have distinct *identity* elements. Thus,

- (14')  $\exists p \forall q (p + q = q)$                        $\exists p \forall q (p * q = q)$

Given that there are such magnitudes, we won't go wrong by assigning them names – and things will be simpler if we do. Let us use ' $Z$ ' to refer to the additive identity element the property, *having no magnitude* and ' $U$ ' to refer to the other, *having a single unit of magnitude*. Then the identity principles appear as,

$$\begin{array}{ll} \text{Any } p \text{ is such that } Z + p = p & \text{Any } p \text{ is such that } U * p = p \\ \forall p (Z + p = p) & \forall p (U * p = p) \end{array}$$

Require also that  $Z \neq U$ . And, where ' $<$ ' indicates an ordering on the magnitude properties,

- (15') If magnitude  $p$  is prior in the ordering to  $q$  then for any  $r$  the sum  $p + r$  is prior in the ordering to the sum  $q + r$ .  

$$\forall p \forall q \forall r [p < q \rightarrow (p + r < q + r)]$$

And we might continue with other principles as for the reals. Then, as before, it follows that *having one unit of magnitude* is prior in the ordering of magnitudes to *having two*

*units of magnitude.* Then, with some principle according to which if  $p < q$  and the the magnitude of of a thing is  $p$ , then its magnitude is less than  $q$ , we have that Fred is less than two feet tall.

So a bit of theory about properties enables us to reach the same sorts of conclusions. In either case, the total theory puts us in a position to draw the desired conclusions from (7), and does not require that we adopt infinitely many *ad-hoc* principles relating individual height predicates. Of course, this second theory quantifies over properties. So pick your poison. Or maybe try another? At this stage, we begin to see the significance of the suggestion that accepting a metaphysical theory is on a par with accepting a theory of science. It is not *deduced*, or even immediately obvious, what truth condition is right. Various proposals will seem more or less plausible depending on a wide variety of considerations.

Return to the number proposal. From what we have seen, classical mathematics is committed to numbers. But, so far, we have not said anything about whether numbers are abstract objects. Thus far at least, one might hope to retain classical mathematics but escape without appeal to abstract objects. As we have seen in chapter 1, however, this may be no easy matter. So, *without any appeal to Plato's beard*, we seem driven by theoretical pressures into just the sort of debates that arose in chapter 1. And, as Quine argues so forcefully in the latter portions of "On What There Is," we are thus drawn into questions about the relevant data and about theory plausibility.

We now have in place the basic elements of the Quinean approach to ontology: Beginning with some straightforward (or maybe not so straightforward) truths, we ask about their truth conditions. We attempt to express truth conditions in our canonical notation, and are committed at least to whatever has to exist in order to make the canonical expressions true. It is then a matter of theory acceptance *which* account we will accept as the legitimate truth condition. In general, there is pressure to give *some* account in the canonical notation of the truth conditions for the starting sentences, though it is always possible to bite the bullet and reject a starting point. Perhaps it is fair to say that many accept this basic structure. Or, at least, many are willing to work within it. However, as we shall see in the next chapter, the picture is still missing at least one crucial link—where this link has important and surprising metaphysical consequences.

## Chapter Five

### Quine's Second Thesis: Sufficiency and Application

We have assumed that there is a world and that sentences are true or false by virtue of the way it is; developed a canonical notation for expression of truth conditions; considered Quine's account of how the canonical notation commits us to the existence of objects; and seen something of how to use the notation as part of an overall method for metaphysics. But it's not yet clear *which* truth conditions may be expressed in the notation. Ideally, it would be sufficient to express any truth condition. But this is more than Quine, and maybe anyone, claims. Still, it may seem that a certain "extension" of  $\mathcal{U}$  is sufficient to express any truth condition. An immediate consequence of the thesis that this "extended"  $\mathcal{U}$  is sufficient to express any truth condition, is trouble for sentences or theories whose supposed truth conditions have, or seem to have, no expression in the extended  $\mathcal{U}$ . Interestingly, Quine's criterion Q1 may *itself* fall into this category. Thus, in this chapter, we'll ask what truth conditions may be expressed in  $\mathcal{U}$  (and the "extended"  $\mathcal{U}$ ), consider some classes of ordinary expressions whose truth conditions seem to have no expression in the extended language, and take up consequences of Quine's method, as it applies to his own criterion of ontological commitment. In the end, we will have set up a host of interesting and difficult metaphysical problems—problems at the core of contemporary metaphysics.

#### *I. The Sufficiency of the Notation*

We want to know which truth conditions can be expressed in  $\mathcal{U}$ . It will be helpful to say something about *why* this matters. This will lead us to think about the notion of an *extensional* language, and the relation between  $\mathcal{U}$  and extensional languages more generally. Then we will be in a position to develop a response to the main question about *which* truth conditions can be expressed in  $\mathcal{U}$ .

(A) As we have seen, Quine holds that,

- Q1 A theory is committed to those and only those entities to which the bound variables of the theory must be capable of referring in order that the affirmations made in the theory be true.

In the previous chapter, we simply *assumed* that a theory's truth conditions were expressed in the canonical notation, and asked whether expressions in that form commit us to just the entities which must be fed into functions in order for the expressions to be true. The verdict was, in the main, favorable. But this leaves open the question whether every truth condition may be expressed in the canonical notation. And a favorable evaluation of Quine's overall method would seem to require a positive verdict with respect to this second question as well.

To see this, consider the way Quine's method is supposed to work. In applying the method, it is not enough merely to notice that 'Bill is president' has no quantifier, and to conclude that it therefore involves us in no ontological commitments. Rather, the idea is to express its truth condition in the canonical notation, and then to apply Q1. Thus, ' $\exists x([x \text{ is-bill} \wedge \forall y(y \text{ is-bill} \rightarrow x = y)] \wedge x \text{ is president})$ ' includes quantifiers, and commits us, in a straightforward way, to a unique thing that is-bill. So we move from the ordinary statement, to its truth condition in  $\mathcal{U}$ , and then to the ontological commitment. This process requires that it be *possible* to express the original truth condition in  $\mathcal{U}$ . If it is not possible correctly to express some theory's truth conditions in  $\mathcal{U}$ , then the method is incapable of discovering that theory's ontological commitments.

This point has important consequences. Consider some theory  $t$ , and suppose every account of its truth conditions as expressed in  $\mathcal{U}$  involves commitments which we cannot accept. Does it follow that we should reject theory  $t$ ? It does not—or at least it does not if  $t$  has or might have truth conditions inexpressible in  $\mathcal{U}$ ! That is, the argument,

If theory  $t$  is acceptable, then there is an acceptable account of its truth conditions expressed in  $\mathcal{U}$ , or there is some other acceptable account.

Every account of  $t$ 's truth conditions expressed in  $\mathcal{U}$  is unacceptable.

Theory  $t$  is unacceptable.

is invalid—for we may accept the theory along with an alternative account of its truth conditions. Of course, the argument would be valid, if the first premise were simply, "If theory  $t$  is acceptable, then there is an acceptable account of its truth conditions expressed in  $\mathcal{U}$ ." And some of Quine's reasonings seem to require the revised premise. We'll see this as we turn, in this chapter, to consideration of Quine's influential article, "Reference and Modality."<sup>1</sup> In that article, Quine is particularly concerned with quantified modal logic—which we may think of as a theory governing notions of *possibility* and *necessity*. Given resources associated with our notation  $\mathcal{U}$ , Quine considers various attempts to account for statements involving these notions. As he tells the story, each of the attempts either fails accurately to capture truth conditions, or commits us to an unacceptable ontology. He concludes, "so much the worse for quantified modal logic"—which is to say that quantified modal logic is therefore problematic (e33). This move seems to require the premise that truth conditions for statements of quantified modal logic, if there are any such conditions, are in fact expressible in  $\mathcal{U}$ .

---

<sup>1</sup> Quine, "Reference and Modality," revised version in Quine, *From a Logical Point of View* 2nd ed. (Cambridge: Harvard University Press, 1980).

Thus there is second way to object against Quine's criterion. We've seen McX's objection that it commits us to too little. But we might also object that Quine gets ontological commits wrong precisely because he gets *truth conditions* wrong, if he limits himself to the resources of a notation like  $\mathcal{U}$ . Perhaps we accept Quine's criterion in the form: "A theory *expressed in*  $\mathcal{U}$  is committed to those and only those entities to which the bound variables of the theory must be capable of referring in order that the affirmations made in the theory be true"; but, when faced with some ordinary theory, of mathematics or whatever, we deny that its truth condition can be expressed in  $\mathcal{U}$ . In this case, reasoning about commitments of theories expressed in  $\mathcal{U}$  is irrelevant to ontological commitments of the original theory. Of course, having rejected Quine's method in some particular case, it would be nice to see some alternative account of the theory's ontological commitments. But the basis for rejecting the method is undercut to the extent that there is a ground to accept,

TC    Arbitrary truth conditions may be expressed in  $\mathcal{U}$ .

If this is the case, then Quine has a ground for accepting the revised first premise for the above argument—and when faced with some expression whose truth condition seems difficult to express in the canonical notation, it is natural, not to reject Quine's method, but to dig in and work all the harder within it. If there is a way to express the truth condition, then we're faced merely with something difficult—not with something impossible or hopeless. Given TC, Quine might borrow from Russell, claiming that opting out of the method "has many advantages; [but] they are the same as the advantages of theft over honest toil." With Russell, then, we might leave such advantages to others, "and proceed with our honest toil."<sup>2</sup>

As it turns out, however, TC is too strong. Rather than argue for TC, we'll be closer to Quine's own strategy if we introduce the notion of an *extensional* language, and consider the claim,

Q2    The truth condition for any expression (which has a truth value), may be expressed in some extensional language.

The canonical notation is extensional. But it is not the only extensional language, and there is room to hold that TC does not follow from Q2. If there are truth conditions which may be expressed in some extensional language, but not in our canonical notation  $\mathcal{U}$ , then Q2 may be true though TC is not. Clearly, to make progress, we need to say what it is for a language to be extensional.<sup>3</sup>

---

<sup>2</sup> From a related context in Russell, "The Philosophy of Logical Atomism," in *The Philosophy of Logical Atomism*, ed., D. Pears (Lasalle: Open Court, 1985), p. 71.

<sup>3</sup> On these last paragraphs, see Quine, "Logic and the Reification of Universals," *From a Logical Point of View* (Cambridge: Harvard University Press, 1980), p. 105.

(B) A language is *extensional* just in case all of its sentences are extensional. So we need to know what it is for a sentence to be extensional. A sentence is *extensional* if and only if it meets conditions, E1, E2 and E3, described below. An extensional sentence meets all three. A nonextensional sentence fails at least one.

For E1, we need the notion of a *singular term*, where a singular term designates some particular object. We have already encountered such terms as the *subject terms* for **T1**. Thus, e.g., in ‘Bill plays the sax’ and ‘the president plays the sax’, ‘Bill’ and ‘the president’ are singular terms. But similarly, in ‘Bill loves Hillary’ and ‘Bill loves the first lady’, ‘Hillary’ and ‘the first lady’ designate objects and therefore count as singular terms as well. As we’ve seen, it is possible to *avoid* singular terms. But it isn’t *necessary* to avoid them and, at least in cases where existence is unproblematic, one might utilize singular terms as such. Given this, the first condition is easy to state. In an extensional sentence,

- E1 If there is a singular term, switching it for one with the same reference doesn’t alter the truth value of the sentence.

Assume that Superman is real, and that he is as traditionally described. Then ‘Superman’ and ‘Clark’ refer to the same object. But

- (1) Lois believes that Superman can fly.

is true, and

- (2) Lois believes that Clark can fly.

is not—where (2) is like (1) except that ‘Superman’ is replaced by ‘Clark’. Since switching the coreferential singular terms switches the truth value, (1) does not meet condition E1 and is not extensional. Notice that *some* switching may leave the truth value unchanged. Thus, e.g.,

- (3) Lois believes that the man of steel can fly.

is like (1) except that ‘Superman’ is replaced by ‘the man of steel’—and plausibly the truth value of (3) is no different from that of (1). And, similarly,

The best reporter at The Daily Planet believes that Superman can fly.

switches ‘Lois’ for the (plausibly) coreferential ‘the best reporter at The Daily Planet’ but leaves the truth value no different than (1). Insofar as the move from (1) to (2) *does* change the truth value, though, (1) fails condition E1 and isn’t extensional.

On the other hand, if ‘Superman is over six feet tall’ is true, then ‘Clark is over six feet tall’ is true, and ‘The man of steel is over six feet tall’ is true as well. These do not by themselves demonstrate that ‘Superman is over six feet tall’ meets E1. For the requirement is that the truth value remain the same after substitution of *any* coreferring singular term. But it is natural to think that this is the case, and thus that ‘Superman is over six feet tall’ does meet condition E1. Of course, to be extensional, it must meet conditions E2 and E3 as well.

The second condition focuses not on switching singular terms, but on switching predicate terms. We are familiar with predicate terms from the canonical notation. Say predicates are “coextensional” iff they apply to just the same objects—iff their functions return T for exactly the same things. Then, in an extensional sentence,

- E2    If there is a predicate, switching it for another that is coextensional with it does not alter the truth value of the sentence.

Suppose, as might have been the case before modern medicine, that every creature with a heart is a creature with a kidney and every creature with a kidney is a creature with a heart. Then it might be that,

- (4)    Alfred believes that every creature with a heart is a creature with a heart

is true, but

- (5)    Alfred believes that every creature with a heart is a creature with a kidney

is false. But (4) is like (5) except that the second instance of ‘is a creature with a heart’ is replaced by the coextensional ‘is a creature with a kidney’. So (4) fails to meet condition E2 and is therefore not extensional. On the other hand, if

- (6)    Superman is a creature with a heart

is true, and ‘creature with a heart’ and ‘creature with a kidney’ are coextensional, then

Superman is a creature with a kidney

is true as well. Again, this does not *prove* that (6) meets condition E2. But it is plausible that it does meet the condition. Notice that (6) seems to meet *both* E1 and E2.

The final condition focuses not on switching singular terms or predicates, but on switching parts that are themselves sentences—on switching “sentential” parts. Again, we should be familiar with such parts from our discussion of  $\mathcal{Q}$ . (E.g., in ‘ $\sim\exists xFx$ ’, ‘ $\exists xFx$ ’ is a sentential part but ‘ $Fx$ ’ is not—for the first is a sentence and the second is not.) In an extensional sentence,



- E3 If there is a sentential part, switching it with a sentence that has the same truth value doesn't alter the truth value of the (whole) sentence.

Suppose that 'Clark was born on another planet' is true. Then, it is natural to suppose that

- (7) Lois believes that two plus two equals four

is true, but

- (8) Lois believes that Clark was born on another planet

is false. But (7) is like (8) except that the true 'two plus two equals four' is replaced by the true 'Clark was born on another planet'. So switching complete sentences with the same truth value switches the true (7) into the false (8). So (7) does not meet E3 and is not extensional. On the other hand, 'It is not the case that dogs fly' is true. But replacing 'dogs fly' with any other sentence that is false, leaves the truth value intact. Thus, e.g., 'It is not the case that two plus two equals five' is true and, 'It is not the case that cats fly' is true, etc. Again, this does not prove that 'It is not the case that dogs fly' meets E3. But it is plausible that it does meet the condition.

Notice that a single sentence may seem to succeed or fail on all three conditions at once. Thus, e.g., it may be that, 'Lois believes that Superman is a creature with a kidney' fails each of the conditions E1, E2 and E3. Say that Superman has a kidney, but Lois doesn't believe it—though, having been close to Superman and heard his heart beat, she believes that he has a heart. Then, 'Lois believes that Superman is a creature with a kidney' is F, but 'Lois believes that Superman is a creature with a heart', and presumably, 'Lois believes that Clark is a creature with a kidney' are true. Given this, switching 'Clark' for 'Superman' switches truth values, so 'Lois believes that Superman is a creature with a kidney' fails E1; similarly, switching 'is a creature with a heart' for 'is a creature with a kidney' switches truth values, so 'Lois believes that Superman is a creature with a kidney' fails E2; and switching 'Superman has a heart' for 'Superman has a kidney' switches truth values, so 'Lois believes that Superman is a creature with a kidney' fails E3. On the other hand, 'It is not the case that Superman has a heart' has the same truth value as 'It is not the case that Clark has a heart' and the same truth value as, 'It is not that case that Superman has a kidney' and, the same truth value as, 'It is not the case that two plus two equals four'. Thus, 'It is not the case that Superman has a heart' seems to meet all three conditions and thus to be extensional.

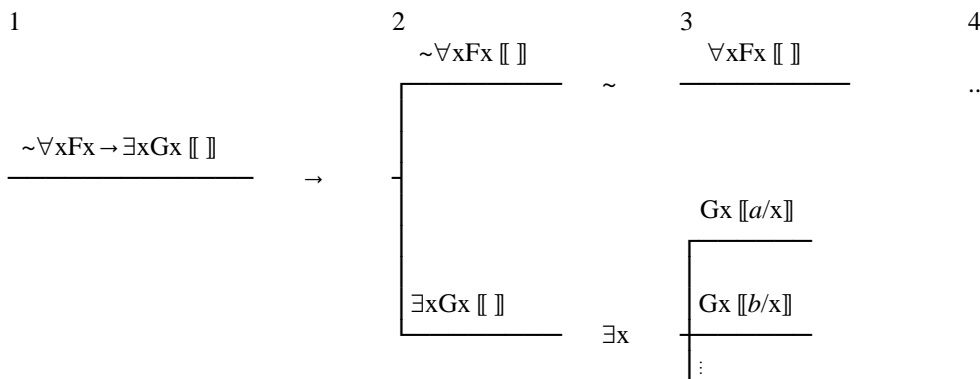
(C) The canonical notation  $\mathcal{U}$  is extensional. Every sentence in  $\mathcal{U}$  meets E1 by default. Since there are no singular terms in  $\mathcal{U}$ , the antecedent of E1, 'if there is a singular term...' is always false. So the conditional (as interpreted by ' $\rightarrow$ ') is always true.

Notice that, insofar as whatever singular terms there might have been in  $\mathcal{U}$  are “transformed” into predicates, the weight of E1 is “shifted” onto E2.

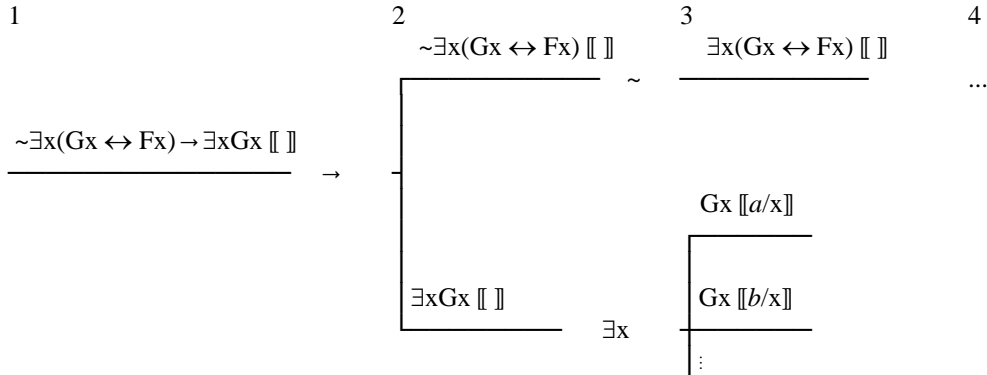
But  $\mathcal{U}$  is sure to meet E2 as well. Consider a tree  $\mathcal{T}$  for some sentence  $\mathcal{S}$ , and a tree  $\mathcal{T}'$  for a sentence  $\mathcal{S}'$  like  $\mathcal{S}$  except that some instance(s) of a predicate  $\mathcal{P}$  in  $\mathcal{S}$  are replaced by a coextensional  $\mathcal{P}'$ . Then, as the trees are constructed in the “forward” direction,  $\mathcal{T}$  is like  $\mathcal{T}'$  except that instances of  $\mathcal{P}$  are replaced by  $\mathcal{P}'$ . Because  $\mathcal{P}$  and  $\mathcal{P}'$  are coextensional, at the branch tips,  $\mathcal{T}$  and  $\mathcal{T}'$  are T and F at just the same locations. But then, as truth is calculated in the “backwards” direction, the trees have to be true and false in the same places as well. The nodes have just the same operators, so the calculations proceed in just the same way. So  $\mathcal{S}$  is true iff  $\mathcal{S}'$  is true. Thus, e.g.,  $\exists xRx$  is sure to have the same truth value as  $\exists xSx$ , if  $R$  and  $S$  apply to exactly the same individuals—and similarly for more complex cases. So sentences in  $\mathcal{U}$  are sure to meet E2.

For E3, consider a tree  $\mathcal{T}$  for a sentence  $\mathcal{S}$  and a tree  $\mathcal{T}'$  for a sentence  $\mathcal{S}'$  like  $\mathcal{S}$  except that some complete sentence  $\mathcal{Q}$  in  $\mathcal{S}$  is replaced by a  $\mathcal{Q}'$  that has the same truth value as  $\mathcal{Q}$ . Then, as it is constructed in the “forward” direction, from the trunk *up to and including* branches on which  $\mathcal{Q}$  and  $\mathcal{Q}'$  appear alone and in their entirety,  $\mathcal{T}$  is like  $\mathcal{T}'$  except that instances of  $\mathcal{Q}$  are replaced by  $\mathcal{Q}'$ . But if  $\mathcal{Q}$  and  $\mathcal{Q}'$  have the same truth value, then  $\mathcal{Q}$  is true on a branch in  $\mathcal{T}$  where it appears alone iff  $\mathcal{Q}'$  is true on the corresponding branch in  $\mathcal{T}'$ . But then, since the trees have the same “shape” up to those branches, the “backwards” calculations *from those points* proceed in parallel; so  $\mathcal{S}$  is true iff  $\mathcal{S}'$  is true. It will be helpful to consider an extended example for this case.

Let’s consider ‘ $\sim\forall xFx \rightarrow \exists xGx$ ’ and replace ‘ $\forall xFx$ ’ with ‘ $\exists x(Gx \leftrightarrow Fx)$ ’. Then  $\mathcal{S}$  is ‘ $\sim\forall xFx \rightarrow \exists xGx$ ’,  $\mathcal{Q}$  is ‘ $\forall xFx$ ’ and  $\mathcal{Q}'$  is ‘ $\exists x(Gx \leftrightarrow Fx)$ ’. So  $\mathcal{S}'$  is ‘ $\sim\exists x(Gx \leftrightarrow Fx) \rightarrow \exists xGx$ ’. Be sure you understand this. We suppose  $\mathcal{Q}$  and  $\mathcal{Q}'$  have the same truth value. Here are starts at the trees for  $\mathcal{S}$  and  $\mathcal{S}'$ . For  $\mathcal{S}$ ,



And for  $\mathcal{S}'$ ,



So far, the trees are exactly the same, except that the second replaces  $\mathcal{Q}$ , that is  $\forall xFx$ , with  $\mathcal{Q}'$ , that is,  $\exists x(Gx \leftrightarrow Fx)$ . The completed trees would diverge in their expansions of the upper branches at (3). *But it doesn't matter.* We are given that  $\forall xFx$  has the same truth value as  $\exists x(Gx \leftrightarrow Fx)$ ; so the upper branches at (3) have the same truth value; and all the other branches at (3) are the *same*; so all the branches at (3) have the same truth value; so the branches at (2) have the same truth value; so the branches at (1) are the same as well—and  $\mathcal{S}$  has the same truth value as  $\mathcal{S}'$ . It doesn't matter *how* we get the truth values for the upper branches at (3); once we get them, supposing that the truth values are the same, the backwards calculations proceed in parallel.

Since every sentence in  $\mathcal{U}$  satisfies E1, E2 and E3,  $\mathcal{U}$  is extensional. But, from this, it does not follow that sentences in  $\mathcal{U}$  have sufficient power to describe *whatever* can be described in an extensional language. To see this, let's concentrate on sentential parts. Consider three sentences,  $\mathcal{P}$ ,  $\mathcal{Q}$  and  $\mathcal{R}$ . These might take on any of the following combinations of T and F:

	$\mathcal{P}$	$\mathcal{Q}$	$\mathcal{R}$
1	T	T	T
2	T	T	F
3	T	F	T
4	T	F	F
5	F	T	T
6	F	T	F
7	F	F	T
8	F	F	F

Say the truth condition for some complex sentence is that  $\mathcal{P}$ ,  $\mathcal{Q}$  and  $\mathcal{R}$  fall either into combination 2 or into combination 6. One way to say that this is so is as follows:

$$(\mathcal{P} \wedge \mathcal{Q} \wedge \sim \mathcal{R}) \vee (\sim \mathcal{P} \wedge \mathcal{Q} \wedge \sim \mathcal{R})$$

(some inner parentheses are omitted for ease of reading). If the first disjunct is T then  $\mathcal{P}$ ,  $\mathcal{Q}$  and  $\mathcal{R}$  fall into combination 2, and if the second disjunct is T then they fall into combination 6; so this sentence is true if  $\mathcal{P}$ ,  $\mathcal{Q}$  and  $\mathcal{R}$  are as in combination 2 or combination 6. There may be simpler sentences that would impose the same condition, but this is of no importance for present purposes. Similarly, we might select, say, combinations 1, 4 or 8 with,

$$(\mathcal{P} \wedge \mathcal{Q} \wedge \mathcal{R}) \vee (\mathcal{P} \wedge \sim \mathcal{Q} \wedge \sim \mathcal{R}) \vee (\sim \mathcal{P} \wedge \sim \mathcal{Q} \wedge \sim \mathcal{R})$$

Given more starting sentences, say,  $\mathcal{P}$ ,  $\mathcal{Q}$ ,  $\mathcal{R}$  and  $\mathcal{S}$ , there are more combinations:

	$\mathcal{P}$	$\mathcal{Q}$	$\mathcal{R}$	$\mathcal{S}$		$\mathcal{P}$	$\mathcal{Q}$	$\mathcal{R}$	$\mathcal{S}$
1	T	T	T	T	9	F	T	T	T
2	T	T	T	F	10	F	T	T	F
3	T	T	F	T	11	F	T	F	T
4	T	T	F	F	12	F	T	F	F
5	T	F	T	T	13	F	F	T	T
6	T	F	T	F	14	F	F	T	F
7	T	F	F	T	15	F	F	F	T
8	T	F	F	F	16	F	F	F	F

But we can do the same sort of thing. We might select combinations 1, 6 or 11 with,

$$(\mathcal{P} \wedge \mathcal{Q} \wedge \mathcal{R} \wedge \mathcal{S}) \vee (\mathcal{P} \wedge \sim \mathcal{Q} \wedge \mathcal{R} \wedge \sim \mathcal{S}) \vee (\sim \mathcal{P} \wedge \mathcal{Q} \wedge \sim \mathcal{R} \wedge \mathcal{S})$$

And, in general, the more starting sentences, the more combinations there are. (Where  $n$  is the number of starting sentences, there are  $2^n$  combinations.) But for any finite number of starting sentences, there is always a way to select, by some sentence of  $\mathcal{U}$ , any combinations whatsoever.<sup>4</sup>

But what if there are infinitely many starting sentences? Though for any sentence of  $\mathcal{U}$ , there is one longer than it, there is no sentence of  $\mathcal{U}$  that is itself infinitely long. And it is not obvious that there are not aspects of the world that would be indescribable in  $\mathcal{U}$ —though it would be describable in an extensional language with infinitely long sentences. This is not so absurd as one might first think. Thus, e.g., we might have predicates corresponding to each integer: ‘is-one’, ‘is-two’, etc, along with an infinite but random set  $S = \{1, 7, 11 \dots\}$ . To specify the membership of  $S$  we would seem to need an infinite expression of the sort,

$$x \text{ is a member of } S \text{ iff } (x \text{ is-one} \vee x \text{ is-seven} \vee x \text{ is-eleven} \dots)$$

---

<sup>4</sup> The proof mirrors the proof of “expressive completeness” given in, e.g., Roy, *Symbolic Logic* §10.2 or Bergmann Moor and Nelson §6.2. This proof is typically taken up in intermediate or advanced courses of logic, but perhaps you are in a position to grasp the strategy already: we can generate a *recipe* for *producing* the desired sentences for the arbitrary case.

Or perhaps there are infinitely many microphysical conditions  $C_1, C_2, \dots$  under which a thing is *red*. Then we might want to say,

$$x \text{ is red iff } (C_1x \vee C_2x \vee \dots)$$

It is certainly possible to *define* a language with infinitely long sentences<sup>5</sup>—though of course it would be difficult to speak or write any of them! Sentences of such a language might be extensional by our definitions. So there may be extensional languages whose expressive power exceeds that of our canonical notation  $\mathcal{U}$ , and there may be some (infinite) aspects of the world that are not describable in  $\mathcal{U}$ . Even so, it should be clear that  $\mathcal{U}$  is a fairly powerful extensional language. We can set up predicates which apply to infinitely many objects. The language includes infinitely many sentences. And it is adequate to express arbitrary finite combinations of the sort considered above. So let's suppose, at least for now, that  $\mathcal{U}$  has all the *extensional* expressive power that we require. Since *all* its expressive power is extensional, we need Q2 to move to the conclusion that the expressive power of  $\mathcal{U}$  is sufficient for our purposes.

(D) Quine's second thesis, (Q2) is that extensional expressive power is enough. Officially, Quine would probably justify Q2 as a theoretical or experimental result: as a matter of fact, when we set out to give truth conditions for the sentences of science or whatever, we find that the sentences of  $\mathcal{U}$ , or at least of some extensional language, are all that is required. As we shall see, this suggestion is, on the face of it, implausible. Or, at least, Quine uses that fact that supposed truth conditions for some sentences have no straightforward account in the canonical language as a reason for questioning the intelligibility of those sentences. But such sentences would seem to be just the sort of evidence that should call an experimental justification for Q2 into question. But there are other ways to motivate Q2—where, at some level, these may have motivated Quine to accept ahead of time that there *must* be some adequate extensional account of the truth conditions for any (intelligible) sentences that have truth conditions.

We have assumed that there is a world and that sentences are true or false by virtue of the way it is. Without prejudicing the question of what *sorts* of things there are, and of the nature of properties or relations, one might think that this assumption amounts to saying that there are things with properties and relations, and that it's the things with their properties and relations that make sentences true or false. Suppose this is right. Then *all we need* to describe the world is the ability to pick out things and to say that they have whatever properties and relations they do. It may be that the truth condition for some sentence is that some complex thing/property/relation combination obtains, or that things fall into some range of thing/property/relation combinations. But, in the final analysis,

---

<sup>5</sup> Indeed, some philosophers have done so. See, e.g., the first part and bibliography of John Barwise, "Infinitary Logics" in *Modern Logic: A Survey* ed., E. Agazzi (Boston: D. Reidel, 1980).

there is nothing to say about the world except that things fall into certain thing/property/relation combinations. This seems to be just the sort of thing that  $\mathcal{U}$  is set up to do, and motivates the emphasis on extensional languages.

Say the world is such that ‘thing  $\alpha$  has property  $\mathcal{P}$ ’ is true. Then, if ‘ $\beta$ ’ picks out the same thing as ‘ $\alpha$ ’, ‘thing  $\beta$  has property  $\mathcal{P}$ ’ should be true as well—for whatever condition makes the first true, makes the second true as well. The *indiscernibility of identicals* is a principle, often thought of as necessary and *a-priori*, according to which for any thing  $a$  and any property  $\mathcal{P}$ , if  $a = b$ , and  $a$  has  $\mathcal{P}$ , then  $b$  has  $\mathcal{P}$  as well.<sup>6</sup> Insofar as the singular terms in a language pick out things, and its sentences say only that the things thus identified have some properties, or stand in some thing/property/relation combinations, the indiscernibility of identicals seems to legitimate E1. So, e.g., Quine opens his paper, “Reference and Modality” as follows:

One of the fundamental principles governing identity is that of *substitutivity*—or, as it might well be called, that of *indiscernibility of identicals*. It provides that, *given a true statement of identity, one of its two terms may be substituted for the other in any true statement and the result will be true* (e21, emphasis his).

Such substitutivity is just what is required by E1.

Immediately following these words, however, Quine observes that there are ordinary examples that seem contrary to this principle. So, e.g.,

(9) Cicero = Tully

and

(10) ‘Cicero’ contains six letters

are true, where,

(11) ‘Tully’ contains six letters

is not.<sup>7</sup> But (9) is a true statement of identity, and (10) is like (11) except that one of the terms in the true statement of identity is substituted for the other. Since (10) is true, by

---

<sup>6</sup> The indiscernibility of identicals is to be distinguished from the much more controversial identity of indiscernibles, according to which if  $a$  and  $b$  have all their properties in common, then  $a = b$ .

<sup>7</sup> Cicero (Marcus Tullius Cicero), had the English byname, ‘Tully’. Of interest for examples from Quine: On October 21 63BC, Cicero denounced Catiline on the floor of the Senate, charging him with treason and conspiring to overthrow the Republic. In the end, evidence of the conspiracy was uncovered, and Catiline’s army was defeated by Rome.

the principle, we expect (11) to be true as well. But it is not. What has gone wrong? Quine does not jump from examples of this sort to the conclusion that substitution authorized by the indiscernibility of identicals is mistaken. Rather, he suggests that in (10), ‘Cicero’ does not function so as to refer to *Cicero*. Rather, in the context of quotation marks, it picks out a *word*, namely the word spelled, ‘C’-‘i’-‘c’-‘e’-‘r’-‘o’. But then ‘Tully’ is not really coreferential with it, and we do not have a counterexample to the substitution principle. Conclusion: insofar as a sentence is saying that a *thing* has some property or stands in some thing/property/relation combination, substituting another name of *that thing* into the sentence does not alter truth value. An appropriate move is from (10) and,

‘Cicero’ = Tully’s other name that begins with ‘C’ and ends with ‘o’

to,

(12) Tully’s other name that begins with ‘C’ and ends with ‘o’ contains six letters

Both (10) and (12) are true; so there is no counterexample to E1. By requiring that our language be extensional, we thus require clarity about what things are said to have what properties. And this is just right when we are concerned about questions of ontology. And once we are clear about the objects, there is no counterexample to the principle that an extensional language is sufficient to express the truth condition.

Justifications for E2 and E3 are related to the justification for E1. It is helpful to focus on *what we need sentences to do*. To express a truth condition, we need to say that some complex thing/property/relation combination obtains, or that things fall into some range of thing/property/relation combinations—or, if we wish to avoid talk of properties and relations as such, that predicates apply to such-and-such combinations of things. But we don’t need to say more than that. Our conditions ensure that we can do this much—but no more. Given E1, all that matters about a singular term is *that it refers to a particular object*. If something else about a singular term matters for truth value, we get the truth value to flip by switching with a singular term that differs with respect to the something else (as ‘Cicero’ and ‘Tully’ have different numbers of letters). Similarly, given E2, the only thing that matters about a predicate is *that it applies to some particular objects*. If something else matters for truth value, we get the truth value to flip by switching with a predicate that differs with respect to the something else. If sentences say only “there are applyings of these predicates to those objects,” then switching coextensional predicates should leave truth values unchanged—for the coextensional predicates apply to just the same objects.

And similarly for E3. A complete sentence puts some condition on the properties and relations of objects. Given E3, the only thing that matters about a sentence which is itself a part of another is *whether its condition is met*. If something else matters for truth

value, we get the truth value to flip by switching with a sentence that differs with respect to the something else. If a complex sentence says just that such-and-such conditions are met, then switching one condition that is met (unmet) for another that is met (unmet) should leave truth values unchanged. Thus the picture according to which the world consists of combinations of things with properties and relations, *and nothing more*, motivates the suggestion that an extensional language is adequate to say whatever there is to say about the world—and thus adequate to give any truth condition.<sup>8</sup>

## II. What the Notation Cannot Accommodate

Every sentence in  $\mathcal{U}$  is extensional. So there is a *prima facie* problem about giving in  $\mathcal{U}$  truth conditions for sentences which seem to be non-extensional. Further, if the supposed truth condition for a sentence resists *all* extensional accounts, and any truth condition has an extensional account, then there is reason to think that the original sentence does not properly say anything about the world at all. This is important insofar as many ordinary sentences seem to be non-extensional. In this section, I outline some types of sentences that seem to be non-extensional, and say something about responses available to the Quinean.

(A) We have already seen that belief contexts may be non-extensional. Thus, e.g., we might imagine a language something like  $\mathcal{U}$  except that it is outfitted with a “belief operator”  $B$  that takes a variable and a sentence into a sentence.  $B(x, \mathcal{S})$  is true iff the object assigned to  $x$  believes that  $\mathcal{S}$ . So, e.g., we might say,

$$(13) \exists x(x \text{ is-alfred} \wedge B(x, \text{the morning star is the morning star}))$$

That is, there is some  $x$  such that  $x$  is-alfred and  $x$  (Alfred) believes the morning star is the morning star. But even with (13) and,

$$\text{the morning star} = \text{the evening star}$$

it may be that

$$\exists x(x \text{ is-alfred} \wedge B(x, \text{the morning star is the evening star}))$$

is false. If this is so, (13) fails E1. It is easy to see that sentences with ‘ $B$ ’ might fail (E2) and (E3) as well (consider (4) and (7)). Since every sentence of the canonical notation is extensional, it follows that no combination of its operators can result in an operation that works like  $B$ . So there is a question about *how* an extensional language might represent

---

<sup>8</sup> This view is related to the logical atomism associated with Russell and the early Wittgenstein. Notice, though, that the above reasoning does not require that there be any special “primitive” facts or things.



truth conditions for a sentence like (13). We'll turn to this question in a moment. But first, let's consider some further examples of nonextensional sentences.

Suppose we let a language have some operators,  $[F]$  'henceforth' (at all future times),  $\langle F \rangle$  'at some future time',  $[P]$  'hitherto' (at all past times), and  $\langle P \rangle$  'at some past time'. Both 'the population of the earth is over 10 billion' and 'the population of the earth is under 3 billion' are (now) false. But

(14)  $\langle P \rangle$ (the population of the earth is under 3 billion)

that is, 'the population of the earth was under 3 billion' is true, and,

(15)  $\langle P \rangle$ (the population of the earth is over 10 billion)

that is, 'that the population of the earth was over 10 billion' is false. So substituting sentences with the same truth value changes the true (14) into the false (15). So (14) fails E3. Similarly, it might be that (x is Bob's sibling) and (x is Bob's brother) currently return T for just the same objects. But then,

(16)  $[F]\forall x((x \text{ is Bob's brother}) \rightarrow (x \text{ is male}))$

is true, though

$[F]\forall x((x \text{ is Bob's sibling}) \rightarrow (x \text{ is male}))$

is false if it will be the case that Bob has a sister. If this is so, (16) fails E2. Against such cases, there is a fairly ready response according to which the substitutions are legitimate only when predicate extensions or truth values are *timelessly* the same. That the brothers are the same as the siblings is true *now*, but not timelessly—and similarly for the population case. But introducing the  $F$  and  $P$  operators requires that 'is' be less than timeless. On this basis, (14) and (16) fail to meet the criteria for extensionality. But, again, every sentence of the canonical notation is extensional. So there is a problem about representing  $F$  and  $P$  by means of operators in the canonical notation.

There are related problems for possibility and necessity. Suppose we include in a language operators ' $\Box$ ' for "necessarily" and ' $\Diamond$ ' for "possibly." Then,

Benjamin Franklin = the inventor of bifocals

and,

(17)  $\Box$ (Benjamin Franklin is Benjamin Franklin)

are true, but,

$\Box(\text{Benjamin Franklin is the inventor of bifocals})$

is false. It could have been that someone else was the inventor of bifocals. So (17) fails (E1) and (E3). Similarly,

(18)  $\Box\forall x[(x \text{ is Bob's brother}) \rightarrow (x \text{ is male})]$

is true, where

$\Box\forall x[(x \text{ is Bob's sibling}) \rightarrow (x \text{ is male})]$

is not. It's necessary that anything that is Bob's brother be male, but it's not *necessary* that anything that is his *sibling* be male. So (18) fails (E2). Like belief and temporal operators, then, these operators for possibility and necessity seem to generate non-extensional contexts, and thus to be unrepresentable by means of operators of  $\mathcal{U}$ .

(B) These are not the only nonextensional contexts, but they are enough for now. There are many responses. In general, the idea is to refine thinking about objects involved. That's how we responded to the “ ‘Cicero’ contains six letters” case (10) above. Similarly, to take another case from Quine, it may be that

(19) Giorgione was so-called because of his size

is true, where

Barbarelli was so-called because of his size

is false—even though Giorgione = Barbarelli. In this case, ‘Giorgione’ refers to *Giorgione*; it's not the *name* that is so-called because of its size, but the *man*. Still, as Quine observes, the name seems to do “double-duty”; it is used to pick out not only the man, but also the name. Thus,

Giorgione was called ‘Giorgione’ because of his size

seems to say the same as (19). And there is no problem with substitution into this based on Giorgione = Barbarelli to reach,

Barbarelli was called ‘Giorgione’ because of his size.

Both are true if either is true. Thus E1 seems met when we get clear about the objects involved. And, in general, the idea is that a context will “go extensional” when we get the objects right.

This strategy seems particularly natural in the case of the belief operator *B*. Perhaps we do not attribute *any* property to the morning star when we say (13)  $\exists x(x \text{ is-alfred} \wedge B(x, \text{the morning star is the morning star}))$ . Rather, ‘the morning star’ is part of a phrase that picks out a *proposition believed*, or *belief state*—where the latter might reside entirely “in the head.”<sup>9</sup> Thus, believing “the morning star is the morning star” is one thing, and believing “the morning star is the evening star” is another. Thus, parallel to Cicero and ‘Cicero’, we treat ‘the morning star’ as something other than a singular term that picks out a celestial body, and deny that ‘the evening star’ is coreferential in the context of *B*—for the substitution results in reference to a different proposition. If the two “singular terms” are not coreferential, then there is no reason to think that the one should substitute for the other without a change of truth value.

For the tense operators *F* and *P*, again, one may think there is a problem about objects—and Quine offers a well-developed and intuitive extensional alternative.<sup>10</sup> The idea is not that we have the *wrong* objects under consideration, but rather that some objects are ignored. Suppose we include in predicates an extra place for a temporal index. Then instead of (14),  $\langle P \rangle$ (the population of the earth is under 3 billion) we may say something like,

$$(20) \quad \exists t[(t \text{ is a time before now}) \wedge (\text{the population of the earth is under 3 billion at } t)]$$

(20) has no proper sentential part, and so no sentential part corresponding to the one that was switched, and got us into trouble, before. (20) is a sentence of  $\mathcal{U}$  and therefore extensional. The same kind of solution works for the other tense case as well. Perhaps,

$$(21) \quad \forall t[t \text{ is a time after now} \rightarrow \forall x(x \text{ is Bob's brother at } t \rightarrow x \text{ is male at } t)]$$

is true, though

$$\forall t[t \text{ is a time after now} \rightarrow \forall x(x \text{ is Bob's sibling at } t \rightarrow x \text{ is male at } t)]$$

is not. But this is no problem for extensionality! For we are assuming that the relations, (x is Bob’s brother at t) and (x is Bob’s sibling at t) don’t apply to the same pairs of objects. They may apply to the same individuals when the time is *now*, but to different individuals at other times. For any person/time pair for which (x is Bob’s brother at t) returns T, the person is male. But there may be person/time pairs for which (x is Bob’s sibling at t) returns T, and the sibling isn’t male. So even though all of Bob’s siblings are

---

<sup>9</sup> But cf., e.g., Putnam, “The Meaning of Meaning,” *Language, Mind and Knowledge*, ed. K. Gunderson (Minnesota: University of Minnesota Press, 1975); reprint in Putnam, *Mind Language and Reality* (Cambridge: Cambridge University Press, 1975).

<sup>10</sup> *Word and Object*, (Cambridge: The MIT Press, 1960), 170ff.

male now, the predicates aren't coextensional. Since (21) is a sentence of  $\mathcal{N}$ , it is extensional. Maybe, then, we have the objects right. Notice, though, that we include *times* among the things we think there are. Perhaps this is acceptable.

More problematic are parallel solutions for possibility and necessity. One option is to try to parallel the solution for beliefs. Perhaps necessity and possibility attach to *sentences* or *propositions*—so that switching 'the inventor of bifocals' for 'Benjamin Franklin' as above switches the propositions under consideration. If this is so, 'Benjamin Franklin' and 'the inventor of bifocals' aren't coreferential. Rather, they are parts of longer designating expressions. Necessity attaches to the proposition expressed by 'Benjamin Franklin is Benjamin Franklin' but not to the proposition expressed by 'Benjamin Franklin is the inventor of bifocals'. One might allow that something about meanings of the terms makes one proposition necessary and the other not – and, at some level, Quine is not opposed to this suggestion.<sup>11</sup> But notice: this solution does not accomodate language-independent properties of possibility and necessity for things – and one may think things do have such properties. Thus Franklin is necessarily human, necessarily a descendent of certain parents, could have entered a different career, and the like – and all this apart from modes of speech. If this is right, the current proposal is not sufficient.

Here's another option: Parallel to the solution times with  $[F]$  and  $\langle F \rangle$ , we claim there are objects that have been ignored. In particular, other *ways the world could have been* have been left out of consideration. The world is some way. But there are other ways that it could have been. If one way the world could have been is such that someone other than Franklin is the inventor of bifocals, then it is *possible* that the inventor of bifocals is not Franklin, and if every way the world could have been is such that Franklin is Franklin, then it is *necessary* that Franklin is Franklin. That is, we may agree that Franklin has to be Franklin—that,

$$(26) \quad \forall w[(w \text{ is a way the world could have been}) \rightarrow (\text{Benjamin Franklin is Benjamin Franklin at } w)]$$

is true, but allow that,

$$\forall w[(w \text{ is a way the world could have been}) \rightarrow (\text{Benjamin Franklin is the inventor of bifocals at } w)]$$

---

<sup>11</sup> But see, e.g., "Three Grades of Modal Involvement," *Proceedings of the XIth International Congress of Philosophy* (Amsterdam: North-Holland Publishing Co., 1953; reprint in Quine, *The Ways of Paradox and Other Essays* revised ed. (Cambridge: Harvard University Press, 1976). See also, "Two Dogmas of Empiricism," *Philosophical Review* (1951); revised in Quine, *From a Logical Point of View* 2nd ed. (Cambridge: Harvard University Press, 1980).

is false. Just as ‘(x is Bob’s brother at t)’ and ‘(x is Bob’s sibling at t)’ aren’t coextensional even though they apply to the same objects now, so ‘(x is Benjamin Franklin at w)’ and ‘(x is the inventor of bifocals at w)’ aren’t coextensional even though they apply to the same object at the way the world actually is. If there are ways the world could have been where Franklin isn’t the inventor of bifocals, then the predicates aren’t coextensional, and there is no problem about E2. Similarly, we may allow that ‘(x is Bob’s brother at w)’ and ‘(x is Bob’s sibling at w)’ don’t apply to the same objects at *every* way the world could have been. So the relations aren’t coextensional, and we may admit,

$$\forall w \forall x [(w \text{ is a way the world could have been} \wedge x \text{ is Bob's brother at } w) \rightarrow x \text{ is male at } w]$$

is true, but deny

$$\forall w \forall x [(w \text{ is a way the world could have been} \wedge x \text{ is Bob's sibling at } w) \rightarrow x \text{ is male at } w]$$

without any problem for extensionality. Of course now we quantify over *ways the world could have been*. But what are these things? and where are they? We may have stumbled into Platonism once again! We’ll return to this issue in a later chapter.<sup>12</sup>

In each case, there are other options, and I don’t claim to have discussed any of these solutions in detail. Rather, I have only developed some examples in order to dramatize the dialectical situation. When faced with a locution that seems not to be extensional, there would seem to be three options: First, we might accept some extensional analysis. This will typically involve some shifting of the original or surface understanding about which things are involved, and might involve some entities that are controversial. Second, if there is no extensional analysis, or there are strong reasons for rejecting every extensional analysis, one might decide to treat the original non-extensional sentences as without truth conditions. That is, one might decide that they don’t say *anything* about the world. As always, this is a desperate expedient. Thus, e.g.,

---

<sup>12</sup> In “Reference and Modality,” Quine doesn’t consider the solution based on “ways the world could have been” (or “possible worlds”)—for it came to prominence only after Quine first wrote. He does consider an attempt to revive the meaning-based solution by treating *meanings* (or concepts) as the very objects to which we refer. The problem, Quine thinks, is a sort of “levels” confusion: though everything we refer to a certain way must be so-and-so, it’s not the case that we must refer to a *way of referring* in just one way. So if ways of referring are themselves objects, related necessities depend in turn on ways of referring to *them*. He concludes that things have modal properties only on some version of “Aristotelian essentialism”—a doctrine rejected by Quine, but typically embraced by those who accept possible worlds. For a highly influential argument against Quine, see Kripke, *Naming and Necessity* (Cambridge: Harvard University Press, 1980). In different places (esp. “Worlds and Modality,” *The Philosophical Review* 102 (1993): 335-361, and “Things and *De Re* Modality,” *Noûs* 34 (2000): 56-84, I claim that something like a meaning-based approach can, in fact, make sense of modal properties.

notions of possibility and necessity are deeply embedded in our notion of scientific law, in the account of argument validity for logic, and—as we shall see in the next section—in Quine’s own criterion of ontological commitment. This is not to say that some original sentences might not have to go. We have seen Quine’s hostility to “possible objects”—and this hostility transfers directly to “ways the world could have been.” Indeed, he would like to do away with possibility and necessity altogether. But there is a third option: We might hold that the picture of the world according to which there are thing/property/relation combinations *and nothing else* is fundamentally flawed. But then we would want to understand the alternative. Insofar as it is not easy to see what this alternative could be, much modern metaphysical discussion has centered around options one and two.

### III. *The Method Applied to the Method*

As others have observed, Quine himself ought to be uncomfortable with Q1. According to Q1, a theory is committed to those and only those entities to which the bound variables of the theory *must* be capable of referring in order that the affirmations made in the theory be true. But this “must” suggests necessity, and raises difficulties akin to those considered above. We can see this by considering how to express the truth condition for Q1 in  $\mathcal{U}$ —by considering what the method has to tell us about the commitments of Quine’s criterion.<sup>13</sup>

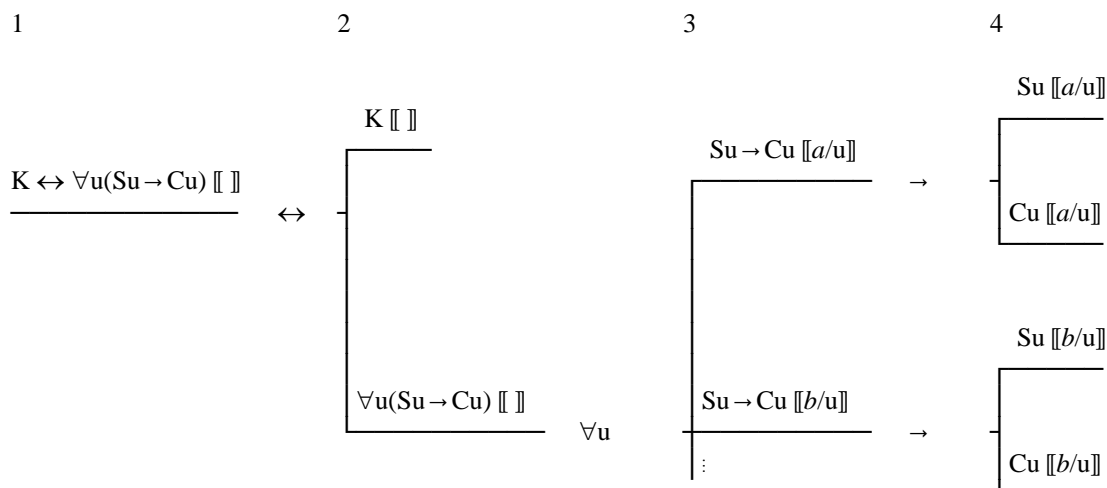
Say a set  $u$  *satisfies* a theory  $t$  if all the sentences in  $t$  remain true when we treat the members of  $u$  as all the things there are. The things in  $u$  are the things fed into functions—and so the things to which variables are capable of referring. Then it may seem natural to recast Q1 as follows,

Q1a Theory  $t$  is committed to entities of kind  $K \leftrightarrow \forall u(\text{if } u \text{ satisfies } t, \text{ then } u \text{ contains } Ks).$

We are committed to  $Ks$  iff the only way to make the sentences of our theory true is to include  $Ks$  in  $u$ . This might get more complicated as we move toward a more official statement in  $\mathcal{U}$ . But it’s enough for now. With natural abbreviations, here’s the tree:

---

<sup>13</sup> This section draws especially on the first part of R. Cartwright, “Ontology and the Theory of Meaning,” *Philosophy of Science* 21 (1954): 316-325. Cf. Campbell, 177-181.



Let's not worry about the existence of sets. Perhaps there is an adequate account along the lines of Goodman's complex objects from chapter 1. Perhaps not. Anyway, given the importance of sets in mathematics and elsewhere, we proceed on the assumption that there is some viable account. Supposing that Q1a is true, both branches at (2) have the same truth value. Thus a theory is committed to Ks when the lower branch at (2) is T, and not committed to Ks when the lower branch at (2) is F. So, carrying on to stage (4), if every  $u$  that satisfies a theory contains Ks, the theory is committed to Ks; if some  $u$  satisfies the theory but doesn't contain Ks, then the theory isn't committed to Ks.

So far, this may seem right. But suppose this is our theory:

$$\exists x(x \text{ is a rabbit})$$

According to Q1a, this theory is committed to rabbits—for any set that satisfies the theory contains a rabbit. And it isn't committed to dogs. For a set with rabbits and no dogs satisfies the theory: for any such set, the upper branch at (4) is T and the lower branch is F; so the corresponding branch at (3) is F; so the universal at the bottom of (2) is F; so, if Q1a is true, the upper branch at (2) is F as well. Good! But we run into trouble if our theory is,

$$\exists x(x \text{ is a unicorn})$$

Then *no* set  $u$  satisfies the theory! Since there are no unicorns, there are no sets with unicorns as members. So all the upper branches at (4) are F; so all the branches at (3) are T; so the lower branch at (2) is T; and, if Q1a is T, the upper branch at (2) is T as well—*no matter what K is*. Is the theory committed to unicorns? According to Q1a, it is. So far, so good. Is it committed to dogs? *According to Q1a, it is*. For *any* kind K, the lower branch at (2) is T; so, according to Q1a, the upper branch at (2) is T as well, and the theory is committed to Ks. But surely this is wrong: Since our theory isn't committed to dogs, Q1a has to be wrong.

Our trouble arose for theories satisfied by no sets whatsoever. So we might try to condition the principle to take such theories into account. Consider, say,

Q1b Theory  $t$  is committed to entities of kind  $K \leftrightarrow [\exists u(u \text{ satisfies } t) \wedge \forall u(\text{if } u \text{ satisfies } t, \text{ then } u \text{ contains } Ks)]$ .

This introduces an *extra condition* which must be met in order to get a T on the lower branch at (2). Unless there is at least one set which satisfies a theory, the branch won't be T. So, according to Q1b, the unicorn theory isn't committed to dogs: Since no  $u$  satisfies the theory, the right-hand side of the triple-bar is false; so if Q1b is T, the left side of the triple-bar is false as well, and the theory isn't committed to dogs. Good! But the theory isn't committed to unicorns either! Since *no* set  $u$  satisfies the theory, the left side of the triple-bar is false *no matter what  $K$  is*. So, given Q1b, our unicorn theory isn't committed to anything. But, again, this is wrong. So Q1b is false as well.

Here's one way out: In both cases, our trouble arose because there weren't any sets with unicorns as members. Because there aren't any unicorns, there are no sets with unicorns as members, and so no sets that satisfy the unicorn theory. But the problem goes away if there are, in some sense, sets with unicorns as members. Let's return to Q1a. If some sets take their members from Wyman's heaven, or from things at different ways the world could be, then there *are* sets with unicorns as members. Any such set which contains no dogs, makes the top branch at (4) T and the bottom F; so the corresponding branch at (3) is F, and the bottom at (2) is F as well. So, according to Q1a, assuming there are such sets, the theory *isn't* committed to dogs. Including sets with unicorns as members among the objects fed into our functions, puts us in a position to say that the theory is committed to unicorns, but not to dogs. This is as it should be.

One might think that problems reappear when our theory is,

$\exists x(x \text{ is a round square})$

Then, even if some sets draw their members from things at different ways the world could be, there are no sets with round squares as members. For there *couldn't* be a round square. So all the upper branches at (4) are F; so all the branches at (3) are T; so the lower branch at (2) is T and, according to Q1a, the theory is committed dogs, rabbits and whatever. But maybe Q1a isn't therefore on the rocks. First, there might *be* round squares in Wyman's heaven. If so, then not all the upper branches at (4) are F, and there is room to deny that our theory is committed to dogs. Alternatively, if we take the members of sets just from ways the world *could* be, maybe we are happy to allow that the theory *is* committed to everything—on the ground that anything follows from a contradiction (see the Appendix). So Q1a might very well do the job. But it does so only insofar as we're involved in commitments with which Quine would be uncomfortable. (!)



Does this mean that we (or Quine) should junk Q1? I think not. Whatever its ontological commitments, the criterion *seems* to make sense. It is *true* that bound variables must be capable of referring to unicorns, if the unicorn theory is true. So Q1 must have *some* truth condition—and we are left with the task of discovering what it is. What we've seen, then, is just how strong pressure on Quine is to find an acceptable account of truth conditions for his own Q1. If we have no alternative to the picture on which there are thing/property/relation combinations and nothing else, and it's unreasonable to reject the necessity in the condition (no matter how much we may want to do so), we are left with Quine's method for finding the truth conditions. In this case, what we need is a viable account of what there is that makes Q1 true. If this involves commitment to Wyman's heaven or other ways the world could be, so be it. Or maybe there is some other account. Adjudicating among ontological commitments of such accounts is precisely what Quine's method is set up to do.

## Part III

### Realism and Truth

## Chapter Six

### Putnam's Anti-Realism

So far, our discussion has been guided by the assumption that there is a world and that sentences are true or false by virtue of the way it is. But this assumption is not above dispute. In this chapter, I take up a line of objection associated with Hilary Putnam's recent defense of "anti-realism."<sup>1</sup> As we have understood it, the assumption that there is a world and that sentences are true or false by virtue of the way it is just amounts to the assumption that realism is right. The anti-realist rejects this assumption. So Putnam's argumentation is directly relevant to our understanding of metaphysics. He argues, negatively, that the sort of realism associated with our assumption is wrong and, positively, that there is a superior alternative.

Perhaps Putnam himself would be willing to say, "there is a world and sentences are true or false by virtue of the way it is." But he would not understand the sentence as we do, and it is important for us to be clear about what Putnam opposes. The assumption naturally divides into two parts. First,

R1) There is a world.

We have assumed that there is a world and have, in effect, understood that it is independent of minds. It is not part of the assumption that the world is any *particular* way. It may consist of concrete objects, abstract objects, spiritual objects, or whatever. But, whatever it is like, either there are rocks or there are not, either there are numbers or there are not, either there are angels or there are not, etc. The brute features of the world are the way they are without respect to what people think or say about them. If there were no minds, the world would continue to exist as it is – except, of course, for the absence of the minds. The second part is correlative with the first.

R2) Sentences are true or false by virtue of the way the world is.

By (R1), the world and its features are "out there." A sentence is then true when the world is as the sentence says it is, and false when the world is not as the sentence says it is. On this view, words "connect" somehow with objects and/or features in the world. "Clinton is in the Oval Office now" is true just in case Clinton and the Oval Office are in the right relation. Or, in the canonical language, we might have, ' $\exists x \exists y (x \text{ clintonizes} \wedge y \text{ ovalofficeizes} \wedge x \text{ is in } y)$ ' so the clintonizer and the ovalofficeizer have to be in the right

---

<sup>1</sup> Putnam calls his view "internal realism" and opposes it to what he calls "metaphysical realism." But "metaphysical realism" is what most call "realism," and his position is at least "anti metaphysical realism."

relation. In either case, for there to be a truth condition at all, the words have to be interpreted in terms of objects or features of the world. After that, the world takes over. The realist thus disconnects *what it is* for a sentence to be true from *how people discover* that a sentence is or is not true. So, e.g., it may be that, “there is, on some planet in the universe, a rock formation that is an exact duplicate of the Venus de Milo” is false even though there is no way for anyone to determine that it is false. And, more dramatically, it seems coherent for a person to suppose that she is a “brain in a vat” even though no test *could* reveal that she is or that she is not a brain in a vat. Putnam rejects realism – that is, he rejects (R1) and (R2) on this understanding – so he rejects what we have assumed throughout.

This chapter takes up in turn Putnam’s chapters, “Is There Still Anything to Say About Reality and Truth?”, “A Problem About Reference,” and “Two Philosophical Perspectives” – which may be seen as developing his attacks on the realist understanding of (R1) and (R2), and then his own positive alternative to realism.

### *I. Realism and Truth*

At first glance, “Is There Still Anything to Say About Reality and Truth?” divides into two main arguments, one from the last section of the article on conceptual schemes, and another from the relation of ordinary phenomena to fundamental physics, from the first sections leading up to that. As we shall see, however, the argument from the first sections may collapse into a version of that from the last.

(A) Putnam begins his chapter with the realist as “seducer,” promising a “fair maiden” ordinary rocks, tables, and chairs, with solidity, colors and the like, which he cannot deliver. So the attack is against the realist understanding of (R1). The central argument in these first sections follows a general pattern, on which ordinary phenomena are said to have no realistic account from the perspective of fundamental physics. This main argument seems to run as follows:

- 1) If there is an objective world, then it consists entirely of concrete material objects.
- 2) If the world consists entirely of concrete material objects, then every feature of it has an account in terms of the “fundamental objects of physics.”
- 3) It is not the case that every feature of the world has an account in terms of the “fundamental objects of physics.”

---

- 4) It is not the case that there is an objective world.

By “objective world,” I mean an “external” world of the sort the realist thinks there is. And here the “fundamental objects of physics” may turn out to be fields – or whatever. As stated, the argument is deductively valid. It falls into the pattern,

If A then B
If B then C
not-C
not-A

If we suppose that A, so that the premises are true and the conclusion not, then with the first premise B; so with the second premise, C; but this contradicts the third premise; so there is no consistent situation in which the premises are true and the conclusion not, and if (1), (2) and (3) are true, (4) must be true as well. Given this, the focus of the debate shifts to soundness, and thus to the question of the truth of the premises. Putnam appears to assume that the realist will accept (1) and (2) and spends most of his time defending (3).

Now, (1) is *not* entailed by the realism of (R1) and (R2). Indeed, our assumption explicitly leaves open that there may be objects in addition to concrete material things. But Putnam seems to think that the contemporary realist is, as a matter of fact, a materialist. So, e.g., at the top e42 he observes that “modern objectivism [realism] has simply become materialism.” In a related article, he titles a section, “Why I focus on materialism,” and begins,

The reason I am going to focus my attack on materialism is that materialism is the only *metaphysical* picture that has contemporary ‘clout’. Metaphysics, or the enterprise of describing the ‘furniture of the world’, the ‘things in themselves’ apart from our conceptual imposition, has been rejected by many.... Today, apart from relics, it is virtually only materialists (or ‘physicalists’, as they like to call themselves) who continue the traditional enterprise.<sup>2</sup>

I don’t know what to say about their being “relics,” – as I expect a platonist or religious believer would certainly object, and many realists would not be happy to call themselves “materialists” or “physicalists.” But, even if I am mistaken about this, I don’t know that Putnam’s assumption affects much of the debate in “Reality and Truth.”<sup>3</sup> That is, abstract or immaterial objects to the side, surely most modern realists accept that at least some

---

<sup>2</sup> Putnam, “Why There Isn’t a Ready-Made World,” in his *Realism and Reason: Philosophical Papers, Volume 3* (Cambridge: Cambridge University Press, 1983), 208.

<sup>3</sup> An exception might be his discussion of intentionality for those who accept that there are non-material minds.

things are concrete material objects. But then we can recast Putnam's argument, and continue.

- 1') If there is an objective world, then it consists at least partly of concrete material objects.
- 2') If the world consists at least partly of concrete material objects, then every physical feature of them has an account in terms of the "fundamental objects of physics."
- 3') It is not the case that every physical feature of concrete material objects has an account in terms of the "fundamental objects of physics."

---

- 4) It is not the case that there is an objective world.

Suppose we give Putnam (1'). The argument remains valid, so the issue shifts to (2') and (3').

Putnam does not argue for (2'). However, again, he thinks the realist will accept it. On e41, he suggests that a fundamental objectivist [realist] assumption is that, "fundamental science – in the singular, since only physics has that status today – tells us what properties things have 'in themselves'." The reasoning seems to be that fundamental science tells us about all the physical properties that are "really there." So if a physical feature of a thing is really there, it must *be* some combination of whatever features fundamental science describes, and therefore must have an account in those terms.

A first worry has to do with the "fundamental objects of physics." Putnam never says what these are. A reference to "finished science" (e35) suggests that maybe nobody knows what they are. Rather than say what the fundamental objects are, he speaks vaguely about "well-behaved function[s] of the dynamical variables" (e37), where these presumably correspond to the formulas of finished physics. As we shall see, his defense of (3') requires that these equations be something like the equations of current physics. Impressed by results in the philosophy of science, according to which the only constant of science is its propensity to change, one might reject this assumption, and Putnam's argument along with it. But I think there is a more fundamental difficulty, so let's give him the adequacy of modern physics. Insofar as we have faith in physics, perhaps we are happy to do so. What is more important, I think, is that Putnam restricts the notion of an "account" for (2') to ones that can be given in *finite* expressions in the language of first-order fundamental physics – or, at least, this seems required for the rest of his argument.

With this much in hand, we can understand Putnam as repeatedly arguing that there are no finite accounts in the language of fundamental physics to be had for certain

obvious and ordinary physical features of material objects. Thus, e.g., he observes that redness may have many different physical realizations and concludes,

There may well be an infinite number of different physical conditions which could result in the disposition to reflect (or emit) red light and absorb light of other wavelengths. A dispositional property whose underlying non-dispositional ‘explanation’ is so very non-uniform is simply incapable of being represented as a mathematical function of the dynamical variables (e36).

And similarly, he observes the difficulty of capturing all the relevant conditions that are relevant to solubility, and concludes,

There is no reason to think that all the various abnormal conditions (including bizarre quantum mechanical states, bizarre local fluctuations in the space-time, etc.) under which sugar would not dissolve if placed in water could be summed up in a closed formula in the language of fundamental physics.

...If the ‘intrinsic’ properties of ‘external’ things are the ones that we can represent by formulas in the language of fundamental physics, by ‘suitable functions of the dynamical variables’, then *solubility* is... not an ‘intrinsic’ property of any external thing (e40).

The second paragraph of the above quotation pulls together themes from each premise of the argument; so it suggests that our interpretation of the argument is roughly correct. So the problem for (3') is that no *finite* formula of physics is going to capture what we have in mind by ‘is red’ or ‘is soluble’. These notions seem to admit of infinitely many realizations at the level of the fundamental particles. (Putnam’s reasoning with respect to intentionality is similar.) Of course, the situation might change if future science is very different from current science. But perhaps this sort of difficulty is likely to remain.

Suppose he is right: It is not the case that for every physical feature of a concrete material object, there is a finite account of it in the language of fundamental science. Does it follow that there is no objective world? I do not think so! The point is already anticipated from chapter 5 (p. 104) where we observed that even Quine does not require that every truth condition has an expression in a finite language like  $\mathcal{U}$ . The requirement of Q2 is rather that every truth condition has an expression in a (maybe infinite) extensional language. And, once we extend to infinite conditions in the language of fundamental physics, Putnam’s observations that no finite expressions will do, are beside the point. The lack of such expressions does nothing to call into question the existence of *real* (microphysical) conditions corresponding to redness, solidity and the like. The charge, then, is that Putnam equivocates: On the one hand, to the extent that we understand the notion of an “account in terms of the fundamental objects of physics” to include expressions of infinite complexity, the realist may very well grant (2'). But, in

this sense, Putnam does nothing to defend (3'). On the other hand, if the "accounts" are limited just to finite expressions, Putnam defends (3'). But there is no reason for the realist to grant (2'). Either way, then, the argument fails.<sup>4</sup> So it may be that the Seducer of Putnam's melodrama is a gentleman unjustly accused – for, again, a condition is no less real for requiring infinite description!

Now, Putnam might insist that his point is sufficient to show that the *specification* of properties such as *being red*, *being soluble*, and the like is not just a matter of physics, but is rather inextricably bound up with human conceptualizing and theorizing. Perhaps this is so. But the relevance of this point to *realism* is hardly clear – and leads us directly to the argument of the last section of his article.

(B) In the last section of "Reality and Truth" Putnam develops a case meant to demonstrate how fundamental features of the world depend on human conceptualization. To the extent that this is so, the independence of the world assumed as background to (R1) seems to fail.

For the case, we are given three simple and distinct objects, **x1**, **x2**, and **x3**, and asked to say how many total objects there are. A natural first answer is, 'Three'. But if objects combine along lines we have associated with Goodman (p. 13), and Putnam associates with certain Polish logicians, then there are seven, **x1**, **x2**, **x3**, **x1 + x2**, **x1 + x3**, **x2 + x3** and **x1 + x2 + x3**. Perhaps the reasoning is as follows.

- 1) From the perspective of conceptual scheme 1, there are three objects.
- 2) From the perspective of conceptual scheme 2, there are seven objects.
- 3) The number of objects is scheme-relative

And this conclusion is taken to be an instance of the anti-realist's thesis. Note that, similarly, given just the primitives from science, one may think that different ways of putting the primitives together would result in different ordinary things – different colors or whatever – thus my suggestion that the cases, if not the arguments, from the sections of the first part feed into this argument from the last.

There are a couple of different reactions a realist might have against this. The first is to insist that at least one of 'there are three objects' or 'there are seven objects' is false. However the world is, there is a fact about the way objects combine, and whatever that

---

<sup>4</sup> Indeed, at least on this account of it, the argument looks like a straw man – and a particularly egregious one insofar as points related to the above may be made in terms of *supervenience* (a notion which we shall meet shortly) and realisms, particularly in philosophy of mind, are *typically* cast in such terms, rather than by the finite reductive notions with which Putnam saddles the realist.



way is, it is such that it is not the case that there are both three and seven objects.<sup>5</sup> This leaves it open what to make of “From the perspective of conceptual scheme *x*...” On the one hand, it might come across like saying, “It is true for me that...” In this case, to the extent that the premises make any sense at all, the first is true if there are three objects, and the second is true if there are seven. So the realist who denies that there are both three and seven objects denies at least one of (1) and (2), and so that the argument is sound. On the other hand, “From the perspective of conceptual scheme *x*...” may be more like a *belief* operator. Then saying that there are three objects from the perspective of one scheme and seven from the perspective of another would be like saying that some people believe there are three objects and others believe there are seven. But, of course, from these premises, the conclusion does not follow – just as different beliefs about the shape of the earth do not show that its *shape* is belief-relative, so different beliefs about how many objects there are do not show that the number of objects is scheme-relative (if you have doubts about this, return to chapter 1, p. 16).

I think Putnam will respond against this that it is *arbitrary* to single out one or the other of (1) or (2) as false. To see this it may help to consider some potential organizational principles.<sup>6</sup> Consider first, *contact*: say objects *x* and *y* compose an object *z* just in case *x* is in contact with *y*. But you in contact with your sock, which is in contact with your shoe, which is in contact with the floor.... So by *contact* the conglomeration of you-sock-shoe-floor, and the rest, turn out to be one big thing. Suppose we try to restrict the range of things by requiring *bonding*: objects *x* and *y* compose an object *z* just in case *x* is bonded to *y*. Now, the nature of the bonds may matter. Perhaps there is no bond between your shoe and the floor. But will we say that “fastening” is sufficient – as your shoe is fastened to your foot by the laces? or a Democrat and a Republican are “fastened” when they shake hands prior to a vicious debate? Or will “cohesion” be required – as, say, some joker smears the inside of your shoe with glue, or glue on the hands of the politicians? Or will we say “fusion” at the wrist of identical twins is a sufficient bond to make them one? Van Inwagen proposes a view on which simples are things, and some *xs* compose a thing iff the activity of the *xs* constitutes a life. So, strictly speaking, on his view, rocks, tables and chairs are not things, though frogs, dogs and people are. In each case, we offer a theory which gives a means of counting things, but the theories may themselves come to seem arbitrary relative to the world.

And surely it is possible to recognize other organizational principles. So, we may recognize snowballs as things. But perhaps some Martians play with *snowbells* (barbell

---

<sup>5</sup> This is the position Van Inwagen defends in “The Number of Things,” in *Realism and Relativism, Philosophical Issues* 12 (2002): 176-196. His own account of what things there are is developed in his *Material Beings* (Ithaca: Cornell University Press, 1990).

<sup>6</sup> Adapted from Van Inwagen, *Material Beings*.

shaped because they have two hands on each wrist).<sup>7</sup> Or perhaps they play with *showbees* (frisbee shaped because they have large flat inflexible palms), or whatever. Each of these is connected. But The Great Wall of China is not. And neither is an installation in an art museum consisting of a tennis shoe dangling by a frayed lace, a baseball bat in the corner, and a dancing raisin under the shoe, with a sign that reads, “Impending Doom” – where critics rave that Impending Doom will revolutionize the art world. Insofar as we allow that these are things, we seem to allow that things may have arbitrary organizing principles. And the possibility of Martians, weird artists, and the like raises the possibility that any combination of things might fall under such a principle. So Putnam may argue that it is not the world, but some merely human set of concerns or interests that make us include some things and reject others under a scheme. As we saw for Van Inwagen above, we can *offer* an account that distinguishes among conditions. But such accounts may themselves appear arbitrary against the background of the full range of accounts available. But, Putnam says, there is then no adequate ground for the realist to divide between ‘there are three objects’ and ‘there are seven objects’ in our case.

Suppose this is right. We already have in place the basis for a second realist reply: We do not so much create as recognize things objectively in the world with our different organizational principles. Thus, the *concept* ‘object’ shifts from one scheme to another. There are objects<sub>1</sub> (objects as identified under counting scheme 1) and there are objects<sub>2</sub> (objects as identified under counting scheme 2). As it happens, there are three objects<sub>1</sub> and seven objects<sub>2</sub>. This observation has no more tendency to show that the total number of objects in the world is scheme-relative than does the observation that there are, say, more cows than horses.

Observe that, on this response, both the objects<sub>1</sub> and objects<sub>2</sub> exist. Thus we are pushed toward what Sosa calls an “explosion of reality” – on which things from *all* the schemes exist. And, as we imagine running the argument against the background of arbitrary schemes, the realist seems pushed toward something like the following. Take the notion of an “occupied” point as primitive, and say,

ER For any set of occupied spatial points, there is a unique thing that occupies those points.

Then we have the objects<sub>1</sub> and objects<sub>2</sub> but there will also be snowballs, snowbells, snowbees, weird art and the like. And if someone sets out to count just the things of one sort or another, that is fine. And, so long as we are willing to accept the “exploded” reality, it is no problem for realism.

---

<sup>7</sup> Now adapting from E. Sosa, “Putnam’s Pragmatic Realism,” *The Journal of Philosophy* 90 (1993): 605 - 626.

But perhaps even (ER) is not enough. Consider the following case, from Cortens, *Global Anti-Realism*.<sup>8</sup>

Suppose that there are two communities, A and B, which are almost exactly alike. In either community, when someone takes a watch to a jeweler to be repaired, the jeweler takes it apart, finds the problem, and puts the pieces back together so they once again make up a watch in good working order. The only difference is that in A, when a person returns to the jeweler, he typically says something like this: “Remember the watch I brought in last week to be repaired? You said that you would have to take it apart to fix the problem. Do you have it ready for me?” In B, a person is more likely to say something like, “Do you have my new watch ready, the one you were going to build out of the parts of the broken watch that I brought you last week?” Loosely expressed, all the members of A say that the recently assembled watch is identical with the watch brought in last week, while all the members of B regard them as two different watches.

So now the argument reappears as follows,

- 1') From the perspective of conceptual scheme A, there is one watch.
- 2') From the perspective of conceptual scheme 2, there are two watches.
- 3') The number of watches is scheme-relative

Our general realist reply has been that the different premises simply talk about different things. But, on (ER) it is not clear that there are different things for the premises to be about. At any given time, in either community, if you look around at all the things there are, all you will see is one watch. The problem is that community B counts things differently across time. Maybe, then, the problem is that we need to “explode” along the temporal, as well as spatial dimension.

ER! For any set of occupied spacetime points, there is a unique thing that occupies those points.

On (ER!) we have not only objects scattered across space, like Impending Doom, but also objects scattered across time. So, there is an object consisting of Napoleon and my left toenail, “Napolenail.” And, likewise, there is an object whose career is that of the one

---

<sup>8</sup> Andrew Cortens, *Global Anti-Realism: A Metaphilosophical Inquiry* (Boulder: Westview Press, 2000), 55-6.

watch on scheme A, and others whose careers are those of the two watches on scheme B.<sup>9</sup> So we recapitulate the response from before. We may have *interests* in talking about one sort of thing rather than another, but that this is not relevant to what things there *are* in the world, which supports them all. Of course, again, on this response, we are committed to the super-exploded view (ER!).<sup>10</sup>

Though his response to this is very brief, Putnam complains, not against explosion as such, but that it is not *neutral*; it is rather *partisan* – as I have developed the reply, his complaint is that the “exploded” scheme is privileged over all the others (e45). But this is not right. Without subscripts, on the first realist line of reply, the different realist notions of ‘object’ are in competition, so that to accept one is to reject the others. This is what Van Inwagen does. But, with subscripts, on the current response, the different accounts do not compete. Grant that (ER!) is the most *general* account of the objects, encompassing all the objects from accounts that are more limited. But, in accepting (ER!) the realist does not *deny* any of the other accounts; that is how she admits (1) and (2) of Putnam’s argument. If there remains any sense in which the resultant view is partisan, it would seem at least that the teeth have been pulled from the complaint that the realist’s view is *arbitrary* – so that it is no longer clear how Putnam’s objection is supposed to have any force.

Observe, though, that the argument seems to leave us with two extreme results. Suppose we reject the first realist reply on which there are primitive (but seemingly arbitrary) principles according to which some principle or another singles out the *things*. Then we are left either with Putnam’s view on which the world supplies none of the things, but rather that things result from the operation of the various conceptual schemes – and so with anti-realism. Or else we are left with the view that the world supplies all of the things, so that they are merely recognized within the different schemes – and so with (exploded) realism. Thus, though Putnam’s argument does not score a fatal blow against the realist position, it does inform its nature. This is an interesting result, from an argument does not achieve its original purpose!

---

<sup>9</sup> I am being deliberately vague about whether these objects are “three-” or “four-” dimensional. The debate here is whether an object whose “career” includes some times should be said to have a *part* at each of those times and so itself be “spread out” over four dimensions.

<sup>10</sup> And we are not done. As we shall observe in the next part, A. Sidelle forwards a version of the argument that seems to push the realist toward a “super-duper” exploded view with *multiple* things in the same spatiotemporal locations (as a statue and the lump of clay of which it is composed). I think this sort of explosion is much more difficult for the realist to swallow than the above. Responding to the issue is a central task of the next (as yet unfinished) part of this text.

## *II. A Problem About Reference*

So far, then, Putnam's attack seems to founder on the reply that there simply *are* features of the objective world corresponding to each of the different schemes. By (R2), sentences are true or false by virtue of the way this objective world is. "A Problem about Reference" along with "Two Philosophical Perspectives" attack the latter claim. By itself, an attack on (R2) might seem to leave (R1) intact. However, an attack on (R2) goes directly at the way we have understood metaphysical method. And realism is surely crippled, if it is left with a world, but no way to talk about it.

The core of Putnam's chapter is an argument to the effect that *no matter how many and how detailed are the sentences we accept, truth values of those sentences do not suffice to determine the reference of our terms to things in an objective world*. We will get to this argument after setting up some background. The argument is clear enough. But then we shall have to consider carefully its proper interpretation and application.

(A) Consider first a traditional picture we have encountered before (p. 29). On this view, ordinary names refer, in the first instance, to ideas. Reference to external objects is only through or by means of the ideas. But just how is this reference between the ideas and external objects supposed to work? Suppose it something like a "picturing" relation. As it turns out, this sort of proposal will not do. First, there are *generality* issues. If 'cat' is associated with a picture, the picture will at once have to be vague enough to let in all the cats, but specific enough to keep out all other animals. This may be quite a trick! A picture would seem by nature to be a *particular*. Similarly it is not clear how, say, a single picture of a triangle could represent all the different kinds of triangles there are. But set such concerns to the side, and consider a case like a personal proper name where the picture model may seem to be the most plausible. A picture, in the most general sense, must represent by virtue of some *mapping* scheme whereby elements of the picture are taken to represent certain features of what is represented. In a color photo, the color on the photo represents the "same" color in the pictured, and the outline represents a similar outline; though the fact that the photo is on paper and two-dimensional do not represent that the pictured is on paper or two-dimensional. In a black-and-white photo, a dark spot does not represent directly that the pictured is that color, but rather that the brightness of the pictured is correlated with the brightness of the photo – bright on the object, bright on the photo. In a negative, we have a representation of the object, but now bright on the negative correlates to dark on the object. In a digital picture, we might have some grid and associated color and intensity codes: the thing represents, but the mapping by which it represents may be idiosyncratic to the particular manufacturer of the device. In short, the mapping for a picture is not part of a picture as such, and so a picture does not *by itself* represent anything. But then, how is a mapping established? If by another picture, that picture will need some interpretation for its

mapping – and the classical proposal is at least not a complete account of reference.<sup>11</sup>

It is natural to think that the picture model can be supplemented or replaced by information, perhaps by what we might call “concepts.” But this model of reference is subject to a famous attack, waged in part by Putnam himself in another place,<sup>12</sup> and rehearsed in the first part of “A Problem About Reference.” We imagine a twin-earth, just like this one, with people like us, etc., except that the oceans are filled with some XYZ otherwise indistinguishable from water instead of H<sub>2</sub>O. Now imagine that we are in the year 1750, prior to the advent of modern chemistry. When J. S. Bach on Earth says, “Give me a glass of water” he refers to water. But, presumably, when Twin-Bach, says “Give me a glass of water” on Twin-Earth he refers to XYZ. But, by hypothesis, Earth-Bach and Twin-Bach are duplicates and, in particular are *mental* duplicates of one another. Thus, by itself, what is “in their heads” – conceptual, pictorial or whatever – seems insufficient to determine reference.

And the same point may be made by related hypotheses. Borrowing from the Stoics, suppose this is a world of two-way eternal recurrence – so everything that is happening now has happened infinitely many times before and will so happen again. (Perhaps there is an infinite series of “big-bangs” and collapses, cycling one after another.) In each cycle, there is a person molecule-for-molecule like me, with thoughts like mine. Still, when I say “I love my wife” I refer to the particular person with whom I live, and when a “Roy” in some other age says it, they refer to the person with whom they live – though the *contents* of our minds are the same, the *referents* of our words are different. As Putnam puts it, “cut the pie any way you like, ‘meanings’ just ain’t in the head!” (“Meaning of Meaning,” p. 227).

(B) The usual conclusion drawn from considerations of this sort (along with considerations of the sort raised in note 8 of chapter 2) is that reference depends on some sort of external connections to things, and in particular that reference requires some specifically *causal* connections with the things to which we refer. Thus, Earth-Bach refers to water because that is the stuff with which he in fact interacts, and Twin-Bach to XYZ because that is the stuff with which he interacts. Similarly, I refer to my wife Rose because she is the person with whom I interact, not ones in other ages.

But in “A Problem About Reference” Putnam takes up a different response, which he calls the “received view,” according to which reference is fixed by certain “operational and theoretical constraints.” These constraints work very much like definitions for our

---

<sup>11</sup> These points about reference and mapping are associated with Wittgenstein, *Philosophical Investigations*, trans. G.E.M. Anscombe (New York: Macmillan, 1953). For the points about generality, see Berkeley, “Introduction” to *Principles of Human Knowledge*.

<sup>12</sup> “The Meaning of Meaning” in Putnam, *Mind, Language, and Reality: Philosophical Papers*, vol. 2 (Cambridge: Cambridge University Press, 1975), 215–71.

terms. But the important point is that, on this approach, reference between our terms and the world is supposed to be fixed by the *truth values* of certain key *sentences*. The main argument of the chapter is that this approach also will not suffice to fix a unique relation of reference between our terms and objects in an objectively existing world. In his chapter, Putnam develops a concrete case involving cats on mats and cherries on trees; the general argument appears only in an appendix. I think that, given our background with the canonical notation, the general argument is easier to understand than the example, and so present a simplified version of the general argument here. I will return to the case of cats on mats and cherries on trees as an application after.

Return to our notions from chapter 3. An interpretation always has a universe of some objects, with some assignments to the predicate letters. These are the *referents* of the predicates (and, given that we are using Russell's way out, that is all we need for reference). In a moment, we shall consider different interpretations with a universe of ordinary concrete objects. But let us begin by considering an abstract *skeleton* or *structure* for these interpretations. Thus, suppose we are given some arbitrary objects,  $o_1$ ,  $o_2$ ,  $o_3$ , and  $o_4$  and an interpretation with a structure  $S$  such that,

$$\begin{array}{lll}
 S: & U & = \{o_1, o_2, o_3, o_4\} \\
 & S(R) & = \{o_1\} \\
 & S(F) & = \{o_2\} \\
 & S(M) & = \{o_3\} \\
 & S(S) & = \{o_4\} \\
 & S(D) & = \{o_1, o_2\} \\
 & S(C) & = \{o_3, o_4\} \\
 & S(P) & = \{\langle o_1, o_3 \rangle, \langle o_1, o_4 \rangle, \langle o_2, o_3 \rangle, \langle o_2, o_4 \rangle\}
 \end{array}$$

We shall see that this can be given some intuitive content. But first, observe that from a structure of this sort we are in a position already to calculate the truth value of sentences involving these predicates. Thus, e.g.,  $\sim \exists x(Dx \wedge Cx)$ ,  $\exists x(Mx \wedge \forall y(My \rightarrow x = y))$ , and  $\forall x \forall y((Dx \wedge Cy) \rightarrow Pxy)$  are all true. Here is a tree for the first.

$$\begin{array}{c}
 \frac{\sim \exists x(Dx \wedge Cx) \llbracket \rrbracket^T}{\sim \exists x(Dx \wedge Cx) \llbracket \rrbracket^F} \sim \quad \frac{\exists x(Dx \wedge Cx) \llbracket \rrbracket^F}{\exists x} \quad \begin{array}{l}
 \begin{array}{l} Dx \wedge Cx \llbracket o_1/x \rrbracket^F \\ \hline Dx \wedge Cx \llbracket o_2/x \rrbracket^F \\ \hline Dx \wedge Cx \llbracket o_3/x \rrbracket^F \\ \hline Dx \wedge Cx \llbracket o_4/x \rrbracket^F \end{array} \wedge \begin{array}{l} \boxed{Dx \llbracket o_1/x \rrbracket^T} \\ \boxed{Cx \llbracket o_1/x \rrbracket^F} \\ \boxed{Dx \llbracket o_2/x \rrbracket^T} \\ \boxed{Cx \llbracket o_2/x \rrbracket^F} \\ \boxed{Dx \llbracket o_3/x \rrbracket^F} \\ \boxed{Cx \llbracket o_3/x \rrbracket^T} \\ \boxed{Dx \llbracket o_4/x \rrbracket^F} \\ \boxed{Cx \llbracket o_4/x \rrbracket^T} \end{array}
 \end{array}
 \end{array}$$

This sentence is true on *any* interpretation with structure  $S$ . Or, generalizing and putting the same point another way, *any interpretations with structure  $S$  must make all the same sentences true*. Be sure you understand this point!

Now consider a pair of concrete interpretations  $I$  and  $J$  with objects matched to  $o_1$ ,  $o_2$ ,  $o_3$  and  $o_4$  as follows,

$I$	Rover	Fido	Morris	Sylvester
	$o_1$	$o_2$	$o_3$	$o_4$
$J$	Rover	Morris	Fido	Sylvester

The interpretations are meant to be specified in relation to the structure  $S$ , so that (since order does not matter) their universes are the same consisting of just Rover, Fido Morris and Sylvester; but e.g.  $I(D) = \{\text{Rover, Fido}\}$  and  $J(D) = \{\text{Rover, Morris}\}$ ; etc.  $I$  is the *natural* or *intended* interpretation. On this account,  $Rx$  is ( $x$  is Rover);  $Fx$  is ( $x$  is Fido);  $Mx$  is ( $x$  is Morris);  $Sx$  is ( $x$  is Sylvester);  $Dx$  is ( $x$  is a dog);  $Cx$  is ( $x$  is a cat) and  $Pxy$  is ( $x$  pursues  $y$ ). Then  $\sim\exists x(Dx \wedge Cx)$ , says that nothing is both a dog and a cat;  $\exists x(Mx \wedge \forall y(My \rightarrow x = y))$ , that some unique thing is-Morris; and  $\forall x\forall y((Dx \wedge Cy) \rightarrow Pxy)$  that any dog pursues every cat. Each of these is true on interpretation  $I$ . It is much less easy to give an intuitive account of the predicate letters on interpretation  $J$ . But *since it has the same structure as interpretation  $I$ , it makes all the same sentences true as interpretation  $I$* . All that matters is the way things are “plugged in” for  $o_1$ ,  $o_2$ ,  $o_3$  and  $o_4$  – so long as things are plugged into the structure consistently, it won’t matter *which* things are plugged into the different slots. But now we have a case where truth values for (all) sentences remain the same, while references for the predicates do not – and so it appears that truth values for the sentences do not suffice to *determine* references for the predicates.

Observe that what we have done is merely to *permute* or shuffle the objects of one interpretation relative to the other. Thus, given objects as for interpretation  $I$  we have a *permutation function*  $P$  which links each object on  $I$  to a “mate” on  $J$ ; thus we have  $P(\text{Rover}) = \text{Rover}$ ,  $P(\text{Fido}) = \text{Morris}$ ,  $P(\text{Morris}) = \text{Fido}$ , and  $P(\text{Sylvester}) = \text{Sylvester}$ . Then interpretation  $J$  can be seen as like  $I$  except that for each object  $x$  in the interpretation of a predicate, instead of  $x$ , we substitute  $P(x)$  in its place. Interpretations which are related in this way are said to be *isomorphic* – and it is a standard theorem from intermediate logic textbooks that isomorphic interpretations make all the same sentences true. But we have already seen the reason why this is so: isomorphic interpretations have the same structure (in effect, the objects of interpretation  $I$  play the role of  $o_1, o_2, \dots$  in the structure, and interpretation  $J$  is specified directly relative to it) and interpretations with the same structure make all the same sentences true. So this is Putnam’s strategy: he



says, “Suppose some specification of truth values for sentences – as many sentences as you like; given any account of reference for these sentences, consider a permutation  $P$  on objects of the domain (indeed, there will be many such permutations), and the corresponding isomorphic interpretations; but these interpretations have all the same sentences true with different references for the predicates. And, if you want to fix truth values at different possible worlds, I can do the same trick at as many possible worlds as you like, so as to alter reference at them all. It follows that truth values do not determine reference.”<sup>13</sup>

The cat/mat example from the main text of Putnam’s chapter does not convey the real force of the argument from the appendix, as described above. In the example, Putnam shows how we can reinterpret ‘cat’ and ‘mat’ with cherries and trees across worlds so that ‘the cat is on the mat’ remains true in just the same cases as before. If you attend carefully, you should be able to see how this works. But the reinterpretation is hardly *general*. Thus though I have in fact *eaten* a cherry (a cat\*) I have not (or so I think) eaten a cat. Since it does not keep all truth values the same, so far as the example goes, it is left unclear whether the seemingly incredible task of altering references without modifying *any* truth values is even possible. The beauty of the permutation strategy is that it reinterprets all the predicates in unison, so that truth values of all the sentences remain the same.

(C) There is much that might be said about this. Given what has gone before, Putnam’s general conclusion is that if truth values are not sufficient to determine reference, then on a realistic account of the world, nothing is sufficient to determine reference. But this is fatal to our assumption (R2) that sentences are true or false by virtue of the way the world is.

The rest of “A Problem About Reference” and much of “Two Philosophical Perspectives” can be seen as a series of replies to objections. He has strong responses to some initial replies. We might say, “Well, the extension of ‘D’ is to be precisely the set of all the *dogs*, and similarly for other predicates; so deviant interpretations like  $J$  are easy to exclude.” But Putnam replies that ‘dog’ is every bit as open to alternate interpretations as is the predicate ‘D’. If this is right, then appeal to ‘dog’ is no help. Similarly, we might think that pointing when we say, “dog” might get around the problem. But pointing is open to reinterpretation as well. So, e.g., my directing the tip of my finger toward Rover and Fido might be understood to single out Rover and Morris – through some understanding of occasional “bent” pointing. The point here is related to one for pictures above: we need a map in order to interpret the map. At this stage, one is tempted

---

<sup>13</sup> This point, made in the appendix of “A Problem About Reference” in terms of permutation and isomorphism, is sometimes made by means of the Löwenheim-Skolem Theorem – another result from formal logic according to which different interpretations may make all the same sentences true (in this case, interpretations with different size universes). See, e.g., Putnam, “Models and Reality,” in his *Realism and Reason: Philosophical Papers, vol 3* (Cambridge: Cambridge University Press: 1983), 1 - 25.

to reply that the deviant interpretations are just too bizarre to be believed. Who would reinterpret pointing in this way, etc.? But, Putnam observes, anyone who understood the terms according to one of the “alternate” interpretations would see our intended interpretation *I* as every bit as “rigged” as we think theirs is. From the perspective of interpretation *J*, *I* mixes dogs and cats in the extension of ‘D’, etc. Putnam thinks the realist has no way out.

But there is a more significant reply. As I suggest above, in response to Twin-Earth and other considerations, certain theorists have held that reference depends on external connections to things, and in particular that reference requires some specifically *causal* connections with the things to which we refer. So the idea is that reference is fixed not merely by what is “inside” our heads, and not merely by the truth values of sentences, but also by these (at least partly causal) external relations between words and things.<sup>14</sup> In “A Problem About Reference” Putnam considers a response along these lines in the mouth of Hartry Field, who proposes that reference is a ‘physicalistic relation’ that is, a “complex causal relation between words or mental representations and objects or sets of objects” (e65). Thus, we are to suppose that there is some relation *R* such that,

- (1) *x* refers to *y* if and only if *x* bears *R* to *y*

where *R* is a physical relation in the sense that it can be defined in the language of science.

Putnam’s first reaction to this is to observe that “If reference is only determined by operational and theoretical constraints, however, then the reference of ‘*x* bears *R* to *y*’ is *itself* indeterminate, and so knowing that (1) is true will not help” (e65). So the reply is parallel to those suggested above. Whatever response the realist makes, Putnam’s suggestion is that *it* is subject to interpretation problems as well. Or, again, in another place Putnam says,

The problem is that adding to our hypothetical formalized language of science a body of theory entitled ‘Causal theory of reference’ is just adding more *theory*.... If ‘refers’ can be defined in terms of some causal predicate or predicates in the metalanguage of our theory, then, since each model of the object language extends in an obvious way to a corresponding model of the metalanguage, it will turn out that, *in each model M*, *reference*<sub>M</sub> is definable in terms of *causes*<sub>M</sub>; but, unless the word ‘causes’ (or whatever the causal predicate or predicates may be) is already

---

<sup>14</sup> For replies along this line, see, e.g., Michael Devitt, *Realism and Truth*, 2nd ed. with new afterward (Princeton: Princeton University Press, 1997).

glued to one definite relation with metaphysical glue, this does not fix a determinate extension for ‘refers’ at all.<sup>15</sup>

But this reply misses the point. The realist argues not that she has some *sentences* that determine reference, but that certain aspects of the *world* determine reference. The most a theory can do is explain what it is that the world does (how reference is determined). The theory, which explains how reference relations are established, is itself made to refer via those factors in the world which it describes. Given supposed problems of reference, Putnam finds it puzzling how even to *express* what Field wants to say (e65). But this is question-begging. He is not entitled to *assume* the failure of reference as part of an argument to show there is a problem about reference – else the argument would be all-too easy!

But, in “A Problem About Reference” at least in the context of this response to Field, Putnam seems to recognize the inadequacy of this initial reply. Granting that the response is inadequate, and supposing (1) is true, now Putnam wants to know what *makes* (1) true. Given that there are many physical relations between words and things, each of which would satisfy operational and theoretical constraints, what singles out the particular relation *R*? He thinks this must be metaphysically unexplainable, a magical theory of reference (e65-66). It is, unfortunately, not entirely clear what the problem is supposed to be. I separate three possible objections in order of increasing plausibility and force:

(i) In one place, Putnam nicely sets up the causal response, setting aside the “just more theory” response considered above.

A realist does not claim that reference is fixed by the conceptual connection (i.e., the connection in our theory) between the *terms* “reference”, “causation”, “sense impression”, etc.; the realist claims that reference is fixed by causation itself.<sup>16</sup>

In this case, he responds that the idea that “*nature itself* determines what our words stand for... is totally unintelligible. At bottom, to think that a sign-relation is *built into nature* is to revert to medieval essentialism.” It is not at all clear why he thinks that the causal theory requires essentialism – the view that things have certain properties in every possible world in which they exist, that necessarily if they exist then they have the properties. As stated, the causal theory would not seem committed to any more essentialism than is the science required to account for relation *R* – where this is surely open to debate. But let the realist suppose whatever is required about the world for the corresponding science. If the objection presumes the rejection of *realism* (if realism itself

---

<sup>15</sup> Putnam, “Models and Reality,” p. 18. [N.B. Read ‘interpretation’ for his ‘model’.]

<sup>16</sup> Putnam, “Introduction” to *Realism and Reason*, xi.

is “metaphysically unexplainable,” “magical”), then the objection must miss the point. For the point of “A Problem About Reference” is to show a problem for (R2) given the supposition that there *is* an objective world of the sort the realist thinks there is. So let us grant the realist whatever features of the world she requires for relation *R*. Given that all the many physical relations between words and things are by hypothesis *different* from one another, there is presumably some real difference between *R* and the other relations not a matter of magic or mystery.

(ii) In asking what singles out the particular relation *R*, Putnam could be asking the realist for a *theory* that would say what this difference between *R* and the other relations is supposed to be – that we are left with a matter of mystery until we find out what the difference is. This is not the place to enter into full-fledged debates in philosophy of language! Of course, it would be nice to supply such a theory. But grant that the realist is not in a position to do so, beyond the observation that the *R* is at least partly causal or the like. Is a failure to supply such theory a problem about reference? It is not. Again, by hypothesis, we do not refer *by means* of the theory, but by the concrete relation. Consider again Twin-Earth cases. There was reference to water prior to 1750, apart from an adequate chemical theory. And the theory did not change the reference of ‘water’, but rather put us better in a position to say what it is we had been talking about all along. Similarly, if there is a causal referential relation between words and the world, the causal theorist would seem to refer to this relation – even though she is not now in a position to say in detail just what it is she is talking about. We may admit that, say, the timing of earthquakes is mysterious. But we do not therefore deny that earthquakes exist, and occur at particular times! Similarly, a mystery about the nature of the difference between *R* and other relations is not a reason to deny it. And again, by hypothesis, the unexplained difference between *R* and the other relations *does* exist.

(iii) But I think that Putnam would want to deny that the supposed difference, whatever it may be, could be one that *makes* a difference. That is, as in the case of the argument from conceptual schemes, the idea would be that any relation between words and things (so long as it respects operational and theoretical constraints) is as good as another. It is therefore *arbitrary* to suggest that some one physical relation *R* is reference and the other candidate relations are not. Now, in the case of “explosion” in response to the argument from conceptual schemes, the realist agrees that there are many things, just as there are many relations in this case. But there are practical and theoretical reasons why we are interested in snowballs, rather than snowbells and, though its parts may be interesting, why Napolenail taken as a whole is not. That is, of all the many things there are, certain ones are of particular interest to us because of the practical and theoretical roles they play. And similarly for properties and relations. So the realist need not deny that there are all the many candidate relations. She only insists that *reference* is a relation of particular theoretical and practical interest because of the roles it plays in our lives. And it is thus distinguished from all the many other relations that would keep truth values of sentences the same.

At this stage, Putnam's objection must be that there *is* no role for reference beyond that of settling truth values for sentences. If this is the beginning and end of the role for reference, then his argument is sufficient to knock the pins from under an attempt to meaningfully distinguish it from other relations. But the causal theorist insists that reference is bound up precisely in causal relations between particular people and things. While we may talk about the "butterfly effect" and allow that "everything is causally connected to everything else" there is a *difference that matters* between the particular causal relations between me and my wife, and the ones between me and someone on the other side of the globe or on another planet. So, on the causal theory of reference, at least, there is every reason to hold that there are roles for referential relations bound up with the way we interact by means of language with things in the world, that distinguish *R* from other candidate relations.

Thus the problem about reference does not seem sufficient to dislodge realism. So far as this argument goes, then, the realist retains (R2) and the corresponding method for metaphysics. Again, though, the attacks on realism have informed us about its nature. In this case, the realist seems driven to an externalist picture of reference on which it is the *world* that does the work of reference, not the mind.

### III. Two Philosophical Perspectives

In this short section, I develop the positive alternative to realism that Putnam offers in "Two Philosophical Perspectives," and raise some worries about it.

(A) In this chapter, Putnam allows that realism is a very natural approach, and suggests that Kant is the first philosopher to put forward an "internalist" view of the sort he defends. As a first approximation, we are to understand this view as a generalization of the doctrine of secondary properties. On the traditional doctrine (associated with Locke, but predating him as well), certain properties, *primary* properties, are features of objects as they are in themselves. These include *mass*, *position*, *motion* and, on contemporary physics, *charge* and the like as well. Other properties, *secondary* properties, are not features of objects as they are in themselves. These include *color*, *warmth*, and *taste*. Such properties are rather *powers* of objects to affect us. On traditional accounts, objects have their secondary properties as a result of complex arrangements of primary properties. Thus, e.g., a thing is *red* as a result of some complex arrangement of primary features with the result that it reflects light in certain ways and causes certain reactions in us.<sup>17</sup> Putnam's "Kantian" proposal is that *all* observed properties are secondary – powers of things as they are in themselves to affect us in

---

<sup>17</sup> But distinction may seem to collapse: for then a thing *does* have the property of *being red* as a complex arrangement of primary properties. What the thing lacks is a certain experienced quality of *redness* which is nothing like the arrangement described by physics. This is supposed to be different from the case of the primary properties where the experienced properties somehow match up with the real ones.

certain ways (for the relation to Kant, see the lengthy quotation in n6 on e77). “It follows,” Putnam says, “that *everything* we say about an object is of the form: it is such as to affect *us* in such-and-such a way. *Nothing at all* we say about any object describes the object as it is ‘in itself’, independently of its effect on *us*, on beings with our rational natures and our biological constitutions” (e78).

Actually, this is only a “first approximation” of the view. For we are not to suppose that when we say a chair is brown, made of pine, and so forth that we are attributing powers to a single object. All we can say is that the *world* is such as to affect us in such-and-such way (e79). And even this is problematic. Kant does speak of the mind-independent reality of things as they are in-themselves (the *noumenal* world). But, given the logic of the view as described above, “we can form no real conception of these noumenal things” and, Putnam suggests, “today the notion of a noumenal world is perceived to be an unnecessary metaphysical element in Kant’s thought” – though this is immediately qualified parenthetically, “but... perhaps we can’t help thinking there is *somehow* a mind-independent ‘ground’ for our experience even if attempts to talk about it lead at once to nonsense” (e78).

On this view, there is no access to a notion of truth as correspondence to mind-independent things. At the same time, we do not want to say that truth is just what people believe, or that truth is simple rational acceptability. For truth is supposed to have a sort of stability that mere belief or rational acceptability may not have (as it was once rational to believe that the earth is flat, but is no longer rational so to believe). But, on Putnam’s view, truth *is* a refined sort of rational acceptability: “truth is an *idealization* of rational acceptability” (e73). A true statement “is a statement that a rational being would accept on sufficient experience of the kind that it is actually possible for beings with our nature to have. ‘Truth’ in any other sense is inaccessible to us and inconceivable by us” (e80).

(B) Say you are presented with a theory according to which, (R) “Everything is relative to X.” There is a natural class of objections about the status of (R) and the entities X. Putnam says in the “Preface” to *Reason Truth and History*, in which “A Problem About Reference” and “Two Philosophical Perspectives” appear, that the mind does not *copy* the world and neither does it *make up* the world; rather, “if one must use metaphorical language... the mind and the world jointly make up the mind and the world” (xi). Here I raise some concerns about how the mind and the world interact to make up the mind and the world, and about the status of the claim that they do so.

(i) It is part of Putnam’s view that, “internalism is not a facile relativism that says, ‘Anything goes’. Denying that it makes sense to ask whether our concepts ‘match’ something totally uncontaminated by conceptualization is one thing; but to hold that every conceptual system is therefore just as good as every other would be something else” (e72, cf., “Is There Still Anything to Say About Reality and Truth,” e46). This rejection of relativism follows the Kantian picture: There is a world as it is in-itself, with powers to

affect us in certain ways; though it matters how we react, the nature of our experience is constrained by the way the world is. Similarly, “*We* cut up the world into objects when we introduce one or another scheme of description” (e71, cf., “Is There Still Anything to Say About Reality and Truth,” e45-46); again, though schemes matter, the world as it is in itself acts as an anchor against rampant relativism. So far, so good.

But it is supposed to *follow* from this picture that we cannot say anything about objects as they are in themselves. To the extent that we follow this to its supposedly inevitable conclusion that “today the notion of a noumenal world is perceived to be an unnecessary metaphysical element in Kant’s thought,” we seem have pulled the rug out from under the view itself. That is, without the noumenal world, we are left with the *assertion* that “internalism is not a facile relativism” without the explanation of why this is so. As Devitt remarks, “to say that our construction is constrained by something beyond reach of knowledge or reference is whistling in the dark;” it comes to a sort of mysterious dogmatism.<sup>18</sup> Indeed, the very notion of a “construction” is rendered mysterious without the world of things-in-themselves as material for the construction.

Thus, in the spirit of charity, it may be best to understand Putnam along the lines of his parenthetical remark according to which “perhaps we can’t help thinking there is *somehow* a mind-independent ‘ground’ for our experience.” But this much would have to be more than “nonsense” if internalism itself is to make sense as a positive view. Observe that, though it is very thin (what Devitt calls “fig-leaf” realism) we are now doing metaphysics of the traditional sort. And once we are in this position, there is room to wonder how strong is the implication from the Kantian considerations to the conclusion that we can know nothing of the world as it is in itself.

(ii) In the interaction between minds and the world, insofar as minds are associated with brains and brains are physical objects, it is natural to think that minds would be objects internal to physical schemes. So Putnam says that “objects *and* signs are alike *internal* to the scheme of description” (e71). This is crucial for the way things turn out to match our concepts, and we are able to refer to them. But the internal positioning of the minds which cut the world into objects is mysterious. To bring this out, consider first a movie about the making of a movie. There is no problem about this. But an on-screen lens does not project anything – it is only the real lens and camera that matter for what shows up on screen. Similarly, if the mind together with its concepts are themselves part of the *construction*, it is hard to see how they can themselves be fundamental to the construction process – it looks like Putnam should say no more than that the mind-in-itself (or rather the world as a whole) is such as to result in experiences of such-and-such sort. But this not only runs against what he explicitly says, but removes much of the work minds and schemes are supposed to do.

---

<sup>18</sup> Devitt, *Realism and Truth*, 230.

Thus, in the spirit of charity, it may be best to offer Putnam a realism about minds (or conceptual schemes).<sup>19</sup> On this view, there are in the world these minds or conceptual schemes, the very ones we use to cut up (the rest) of the world into things. Again, this raises the question, how strong is the implication from the Kantian considerations to the conclusion that we can know nothing of the world as it is in itself.

(iii) Putnam holds that we have no access to a notion of truth as correspondence to mind-independent things; truth is rather an idealization of rational acceptability. Consider some claim, say, ‘Snow is white’ and let us suppose that snow is white. Then, on this view, the statement is rationally acceptable in some idealized sense – it is one that a rational being would accept on sufficient experience of the kind that it is actually possible for beings with our nature to have. Having made “charitable” adjustments to internalism as above, a *correspondence* theorist might understand this as follows, “There are minds and the world, and in fact the minds are such as to accept on sufficient experience that snow is white.” But *this* is not available to Putnam. In a discussion of Dummett, he observes that *reductionist* philosophers can be *realists* (e74). Thus Berkeley was a realist, not about physical objects, but about minds. In this case, our correspondence theorist is a realist about minds and the world, and observes that they stand in certain relations. But Putnam wants to say that “a truly non-realist view is non-realist all the way down.” It is not at all clear to me what this comes to.<sup>20</sup> He might say that a rational being would accept that snow is white on sufficient experience of the right sort just in case *this* – that a rational being would accept that snow is white on sufficient experience of the right sort – is something that a rational being would accept on sufficient experience of the right sort. But this is the way to regress. Somehow, then, the notion of truth seems to require a correspondence component – if only to the world insofar as it includes the way we think about it.

(iv) For the sake of charity, we have made moves in the direction of realism with respect to the world and minds, and in the direction of correspondence with respect to truth. I doubt that Putnam would take kindly to such “kindness.” But it is not easy to see how the view holds together apart from it. In the face of such moves, it is natural to ask whether anti-realism is really the *best* conclusion to draw from his Kantian premises. Suppose there are different conceptual schemes. As we saw in our discussion of “Is There Still Anything To Say About Reality and Truth” it simply does not follow that

---

<sup>19</sup> This is the strategy of C. Elder, “Why The Attacks On The Way The World Is Entail There Is A Way The World Is,” *Philosophia* 16 (1986), 191-202.

<sup>20</sup> Along with certain other realist commentators, I am not sure how well I understand Putnam’s views. Perhaps this is precisely because worries of the sort I raise are so close to the surface and remain unaddressed. Particularly insofar as this text remains in manuscript form, I welcome attempts to help clarify my views! Cf., D. Lewis, “Putnam’s Paradox,” *Papers in Metaphysics and Epistemology* (Cambridge: Cambridge University Press, 1999), 57; and P. Van Inwagen, *Metaphysics 2nd ed.* (Cambridge MA: Westview Press, 2002), 76, 81.



nothing we say about any object describes the object as it is ‘in itself’. From the facts that all thinking about the world is conceptualized, and that there are different conceptualizations, it does not follow that thinking does not represent the world. Similarly, given our concessions, the Kantian premises do not block *everything* we might say about objects as they are in themselves. Further, suppose we have made the concessions and allow that there are minds and the world, and that the world is such as to affect us in such-and-such ways. Is it really the *best* understanding of the situation that this is *all* we can hope to say about things as they are in themselves? That is, it may be that quite sensible theories could include accounts of the way things are in themselves so as to *explain* the interactions between people and things. One might say that the common-sense and scientific theories of things are precisely theories of this sort. Thus, it is hard to see how it *follows* from the Kantian premises according to which the world is such as to affect us in such-and-such ways, that we can say no more than this about it.

Putnam comes close to considering such responses toward the end of “Two Philosophical Perspectives” (e86). His response is just to raise again problems about reference. He thinks that our theories will not map onto the world so as to be *about* the world. Unfortunately, in that context, he talks about abstract “maps” of our terms to the world in a context apart from the causal theory of reference, as if the maps were totally unconstrained by the world. And when he does talk about the causal theory in this chapter (e81), he falls back on the question-begging “just more theory” response, which was rightly bypassed from the reply to Field in “A Problem About Reference.”

#### IV. Conclusion

Putnam’s writing is filled with strong language. Realism is associated with ‘incoherence’, ‘disaster’, ‘seduction’, ‘magic’ and ‘mystery’. This is suggestive of an interesting and powerful thesis. But his thesis is no more powerful than corresponding argument. And arguments that we have managed to tease out do not stand up. Of course, the writing is not entirely clear, and this may give reason to think that something has been missed. Perhaps so. But the unclarity will not count as a virtue, and we will want to see both the positive view and the arguments worked out before associating them with anything like the significance suggested by this language. Indeed, on the face of it, it appears that the proposed alternative, anti-realism, is to be charged with mystery and incoherence.

There are, of course, many anti-realist positions, and we have not begun to examine them all. However, I close with some brief remarks on the significance of the issue – which the reader may take for what they are worth. In chapter one (p. 16) we rejected saying a thing is “true for me but not for you” against the background of an assumed realism and correspondence approach to truth. But Putnam offers an account on which, if correct, there might be different truths against the background of different conceptual schemes, with the result that, “the possibility of a certain pluralism is opened up” (e87).

Thus he offers at least a limited vindication for the saying. We have not objected against differences among conceptual schemes as such. And we have had nothing to say about epistemology. But the slide from such relativisms into metaphysical anti-realism is no simple matter. Observe that metaphysical anti-realism would seem to *require* the other relativisms – for the different realities are bound together with different concepts, beliefs and truths. Metaphysical anti-realism thus has far-reaching methodological and practical consequences, and is not to be accepted lightly. Yet such views seem to be background to much common thought (as reflected in the saying, “true for you but not for me” – insofar as it is not merely confused misstatement for something about belief) and in portions of academia.<sup>21</sup> If our conclusions are right, the view is indefensible. To the extent that starting points are misguided, conclusions are called into question as well. So it is important to get starting points right – and it is no small thing that the view should be presupposed.

---

<sup>21</sup> An anecdote: After the 1992 Los Angeles riots, I was watching a talk show in which a panel of “experts” were trying to say what we should do to prevent such situations in the future. I recall being floored when one said, not that we should improve police procedures, family conditions, educational opportunities, economic opportunities, and the like, but that we should “find different ways to describe the situation.” I suppose his point was that our concepts make the reality, and that we need to reconceptualize. I would have said the reality of Rodney King’s bruises were not going to be reconceptualized away!

# Appendices

## Appendix

### Validity and Soundness

Our concerns require the evaluation of arguments, and validity and soundness are tools for argument evaluation. In this appendix, we introduce these notions, with some informal methods for working with them, and apply them in some simple cases. Methods for manipulating these notions are the *topic* for a course in logic. However, a basic grasp of what validity and soundness *are* should be sufficient for our purposes. And that is all that is attempted here.

An *argument* is a set of sentences, one of which (the *conclusion*) is taken to be supported by the remaining sentences (the *premises*). In a bad argument, premises taken to support a conclusion do not actually do so. *Logical validity* and *soundness* correspond to different ways an argument can go wrong. Consider the following two arguments:

Only citizens can vote
Hannah is a citizen
Hannah can vote

All citizens can vote
Hannah is a citizen
Hannah can vote

These arguments go wrong in different ways. The premises of the first are true; as a matter of fact, only citizens can vote, and Hannah (my daughter) is a citizen. But she can't vote. She isn't old enough. Thus, in the first argument, the relation between the premises and the conclusion is defective. Even though the premises are true, there is no guarantee that the conclusion is true as well. We'll say that this argument is *logically invalid*. The second argument is logically valid. If its premises were true, the conclusion would be true as well. So the relation between the premises and conclusion isn't defective as in the first argument. But the premises of the second argument aren't true—not all citizens can vote. So this argument is defective, but in a different way. We'll say that it is *logically unsound*. After some preliminary notions, we'll turn to official definitions of logical validity and soundness, and to some consequences.

#### I. Consistent Stories

Given a certain notion of a *possible* or *consistent* story, it is easy to state definitions for validity and soundness. So I begin by identifying the kind of stories that matter. Then we'll be in a position to state the definitions, and apply them in some simple cases.

Let's begin with the observation that there are different sorts of possibility. Consider, say, "Hannah could make it in the WNBA." This seems true. She is reasonably athletic, and if she were to devote herself to basketball over the next few

years, she might very well make it in the WNBA. But wait! Hannah is only a kid—she rarely gets the ball even to the rim from the top of the key—so there is no way she could make it in the WNBA. So she both could and couldn't make it. But this can't be right! What's going on? Here's a plausible explanation: Different sorts of possibility are involved. When we hold fixed current abilities, we are inclined to say there is no way she could make it. When we hold fixed only general physical characteristics—and allow for development—it is natural to say that she might. The scope of what is possible varies with whatever constraints are in play. The weaker the constraints, the broader the range of what is possible.

The sort of possibility we're interested in is *very* broad, and constraints are correspondingly weak. We'll allow that a story is *possible* or *consistent* so long as it involves no *internal* contradiction. For this, it may help to think about the way you respond to ordinary fiction. Consider, say, *Bill and Ted's Excellent Adventure*. Bill and Ted travel through time in a modified phone booth collecting historical figures (including Socrates) for a history project. So far, so good (excellent). But, late in the movie, Bill and Ted have a problem breaking the historical figures out of jail. So they go back in time tomorrow to set up a diversion for today. The diversion goes off as planned, and the day is saved. Somehow, then, tomorrow has to happen for things to be as they are today. But today has to happen for tomorrow. Perhaps today and tomorrow have always been repeating in an eternal loop. But, according to the movie, there were times before today and after tomorrow. So you say, "how can *this* be?" Notice: your objection doesn't have *anything* to do with the way things actually are—with the nature of actual phone booths and the like; it has rather to do with the way the movie hangs together internally.<sup>1</sup> Similarly, we want to ask whether stories hold together *internally*. If a story holds together internally, it counts, for our purposes, as consistent and possible. If a story doesn't hold together, it isn't consistent or possible.

Consider some examples: (a) The true story, "Everything is as it actually is." Since no contradiction is actually true, this story involves no contradiction; so it is internally consistent and possible. (b) "All dogs can fly: over the years, dogs have developed extraordinarily large and muscular ears; with these ears, dogs can fly." It's bizarre, but not obviously inconsistent. If we allow the consistency of stories according to which monkeys fly, as in *The Wizard of Oz*, or elephants fly, as in *Dumbo*, then we should allow that this story is consistent as well. (c) "All dogs can fly, but my dog Fido can't; Fido's ear was injured while he was chasing a helicopter, and he can't fly." This *isn't* internally

---

<sup>1</sup> In ordinary cases of time travel (in the movies) time seems to move in a sort of 'Z' so that after yesterday and today, there is *another* yesterday and *another* today. So time doesn't return to the very point at which it first turns back. In this case, however, time seems to move in a sort of "loop" so that a point on the path to today (this very day) goes through tomorrow. (With this in mind, it is interesting to think about, say, the *Terminator* and *Back to the Future* movies and, maybe more consistent, *Groundhog Day*.) Even if I'm wrong, and this movie is internally consistent, the overall point should be clear. And it should be clear that I'm not saying anything serious about time travel.

consistent. If *all* dogs can fly and Fido is a dog, then Fido can fly. You might think that Fido remains a flying *sort* of thing. In evaluating consistency, however, we require that *meanings remain the same*. If “can fly” means just “is a flying sort of thing,” then the story falls apart insofar as it says both that Fido is and isn’t that sort of thing; if “can fly” means “is himself able to fly,” then the story falls apart insofar as it says that Fido himself both is and isn’t able to fly. So long as “can fly” means the same in each use, the story falls apart insofar as it says both that Fido can and cannot fly. (d) “Germany won WWII: the United States never entered the war; after a long and gallant struggle, England and the rest of Europe surrendered.” It didn’t happen; but the *story* doesn’t contradict itself. For our purposes, then, it counts as possible. (e) “ $1 + 1 = 3$ ; the numerals ‘2’ and ‘3’ are switched (‘1’, ‘3’, ‘2’, ‘4’, ‘5’, ‘6’, ‘7’...); so that taking one thing and one thing results in three things.” This story *doesn’t* hang together. Of course, numerals can be switched. But switching numerals doesn’t make *one* thing and *one* thing *three* things! We tell stories in our own language (imagine that you are describing a foreign-language film in English). According to the story, people can say correctly ‘ $1 + 1 = 3$ ’, but this doesn’t make it the case that  $1 + 1 = 3$ . Compare a language like English except that ‘fly’ means ‘bark’; and consider a movie where dogs are ordinary, but people correctly assert, in this language, “dogs fly”: it would be wrong to say, in *English*, that this is a movie in which *dogs fly*. And, similarly, we haven’t told a story where  $1 + 1 = 3$ .

As with a movie or novel, we can say that different things are true or false *in* our stories. In *Bill and Ted’s Excellent Adventure*, it is true that Bill and Ted travel through time in a phonebooth, but false that they go through time in a DeLorean. Of course, in some other story, it may be that a DeLorean goes through time. In the real world, of course, it is false that phonebooths go through time, and false that DeLoreans go through time. But there are a couple of contexts where caution is in order. The first is when some story we tell doesn’t include enough information to determine whether a sentence is true or false. In *The Wizard of Oz*, it is true that Dorothy wears red shoes. But neither the book nor the movie have anything to say about whether she likes Wheaties. By themselves, then, neither the book nor the movie give us enough to tell whether “Dorothy likes Wheaties” is either true or false in the story. Similarly, there is a problem when stories are inconsistent. Suppose, according to some story,

- (i) All dogs can fly
- (ii) Fido is a dog
- (iii) Fido can’t fly

Given (i), all dogs fly; but from (ii) and (iii), it seems that not all dogs can fly. Given (ii), Fido is a dog; but from (i) and (iii) it seems that Fido isn’t a dog. Given (iii), Fido can’t fly; but from (i) and (ii) it seems that Fido can fly. The problem isn’t that inconsistent stories say too little, but rather that they say too much. When a story is inconsistent, we’ll simply refuse to say that it makes any sentence (simply) true or false.

Again, let’s consider some examples. Consider the sentences,

- (i) If you are in San Francisco, then you are in California
- (ii) You are in California
- (iii) You are in San Francisco

And consider the stories, (a) “San Francisco is in California, and you are there.” This is consistent. And it makes all the sentences true. Since San Francisco is in California, it is true that if you are in San Francisco, then you are in California. And since, according to the story, you are in San Francisco, (ii) and (iii) are true as well. (b) “San Francisco and San Bernardino are in California, but they are in different parts of the state, and you are in San Bernardino.” This is consistent. As before, it makes (i) true. And it makes (ii) true as well, for you are in San Bernardino, and San Bernardino is in California. But (iii) is false. Since you are in San Bernardino, you aren’t in San Francisco. (c) “Aliens have carted San Francisco, with you in it, to another state.” This is odd, but it isn’t obviously inconsistent.<sup>2</sup> In this case, (i) isn’t true; if you are in San Francisco, you are in another state—not California. And (ii) is false as well; you are in that other state. However, (iii) is true—for you remain in San Francisco. (d) “Aliens have carted San Francisco to another state; you are in San Francisco, but have remained in California.” This story isn’t consistent. If San Francisco is in another state, and you are there, then you aren’t in California. Since the story isn’t consistent, it doesn’t make (i), (ii) or (iii) either true or false. (e) “San Francisco is in California, and you are baffled by readings required for a philosophy class.” There is nothing inconsistent about this story. It makes (i) true. But it has nothing to say about *where* you are. So far as what we have said then, (ii) and (iii) are neither true nor false. Clearly, we can construct stories that make (i), (ii) and (iii) true and false in different combinations. And the definition of validity depends on just this sort of thing.

## II. Definitions

The definition of validity depends on what is true and false in consistent stories. The definition of soundness builds directly on the definition of validity. Note: in offering these definitions, I *stipulate* the way the terms are to be used; there’s no attempt to say how they are used in ordinary conversation; rather, we say what they will mean for us in this context of logic and philosophy.

- LV An argument is *logically valid* if and only if there is no consistent story in which all the premises are true and the conclusion is false.

---

<sup>2</sup> At least if we agree that moving the buildings with some soil moves *the city* (as the London Bridge is now in Arizona). If we say that, after the buildings are moved, *San Francisco* is a hole in the ground, then there is room to reject the story as inconsistent.

- LS An argument is *logically sound* if and only if it is logically valid and all of its premises are true in the real world.

Logical (deductive) validity and soundness are to be distinguished from *inductive* validity and soundness. For the inductive case, it is natural to focus on the *plausibility* or *probability* of stories—where an argument is relatively strong, when stories that make the premises true and conclusion false are relatively implausible. Logical (deductive) validity and soundness are thus a sort of limiting case, where stories that make premises true and conclusion false are not merely implausible, but impossible. In a deductive argument, conclusions are supposed to be *guaranteed*; in an inductive argument, conclusions are merely supposed to be made probable or plausible. Let's set the inductive case to the side, and focus on deductive arguments.

(A) If an argument is logically valid, there is *no* consistent story that makes the premises true and conclusion false. So, to show that an argument is invalid, it is enough to *produce* even one consistent story that makes premises true and conclusion false. Perhaps there are stories that result in other combinations of true and false for the premises and conclusion; this doesn't matter for the definition. However, if there is even one story that makes premises true and conclusion false then, by definition, the argument isn't logically valid—and if it isn't valid, by definition, it isn't sound. We can work through this reasoning by means of a simple format for validity. So consider the argument,

Eating Brussels sprouts results in good health  
 Ophilia has good health  
 —————  
 Ophilia has good health

Format:

Example:

a. List premises and negation of conclusion.

In a story with premises true and conclusion false,

- (1) Eating brussels sprouts results in good health.
- (2) Ophilia has good health.
- (3) Ophilia has not been eating brussels sprouts.

b. Produce a consistent story in which statements from (a) are all true

Story: Eating brussels sprouts results in good health, but eating spinach does so as well; Ophilia is in good health but has been eating spinach, not brussels sprouts.

c. Apply the definition of validity.

This is a consistent story that makes the premises true and the conclusion false. So, by definition, the argument isn't logically valid.

d. Apply the definition of soundness.

Since the argument isn't logically valid, by definition, it is not logically sound.



We begin, in step (a), with the target for our story. To show that the argument is invalid, we produce, in step (b), a story that hits the target, and so makes the premises true and conclusion false. Steps (c) and (d) apply the definitions to get the final results. For practice, you should use the format to show that the following arguments are not logically valid and not logically sound:

- (1) If Joe works hard, then he will get an 'A'  
Joe will get an 'A'  
——  
Joe works hard
- (2) Harry had his heart ripped out by a government agent  
——  
Harry is dead
- (3) Everyone who loves logic is happy  
Jane doesn't love logic  
——  
Jane isn't happy
- (4) The car won't run unless it has gasoline  
The car has gasoline  
——  
The car will run
- (5) Only citizens can vote  
Hannah is a citizen  
——  
Hannah can vote

(B) For a given argument, if you can't find a story that makes the premises true and conclusion false, you may begin to suspect that it is valid. However, mere *failure* to demonstrate invalidity doesn't *demonstrate* validity—for all we know, there might be some tricky story we haven't thought of yet. So, to show validity, we need another approach. If we could show that *every* story which makes the premises true and conclusion false is *inconsistent*, then we could be sure that no *consistent* story makes the premises true and conclusion false—and so we could conclude that the argument is valid. Again, we can work through the reasoning by means of a format, this time, for validity. Consider the following argument,

No car is a person  
 My mother is a person  
 —————  
 My mother is not a car.

Format:

Example:

a. List premises and negation of conclusion.

In a story with premises true and conclusion false,

- (1) No car is a person.
- (2) My mother is a person.
- (3) My mother is a car.

b. Expose the inconsistency of such a story.

c. Apply the definition of validity.

d. Apply the definition of soundness.

In any such story,

Given (1) and (3),

4) My mother isn't a person.

Given (2) and (4),

5) My mother is and isn't a person.

So any such story is inconsistent. So there is no consistent story that makes the premises true and the conclusion false. So, by definition, the argument is logically valid.

Since in the real world no car is a person and my mother is a person, the premises are true in the real world. Thus, by definition, the argument is logically sound.

Again, we begin, in step (a), listing the premises and negation of the conclusion. Before, in step (b), we would have produced a consistent story to hit this target. This time, we demonstrate that any attempt to do so must collapse into inconsistency. This puts us in a position to conclude at (c) no attempt to hit the target with a consistent story can succeed – so that the argument is logically valid. In this case, soundness depends on whether the premises are true. If one or more were false in the real world, it still would not be sound. Notice that only one format applies in a given case. If we can *produce* a consistent story according to which the premises are true and the conclusion is false, then it is not the case that *no* consistent story makes the premises true and the conclusion false. Similarly, if *no* consistent story makes the premises true and the conclusion false, then we won't be able to *produce* a consistent story that makes the premises true and the conclusion false.

In this case, the most difficult steps are (a) and (b)—where we say what is the case in *every* story that makes the premises true and the conclusion false. For an example, consider the following argument:

Some collies can fly
All collies are dogs
—
All dogs can fly

It's invalid; we can tell a story that makes the premises true and conclusion false—say, one where collies and German Shepherds are dogs, Lassie is a collie who can fly, and Rin Tin Tin is a German Shepherd who cannot. But suppose we proceed with the validity format as follows,

- a. In a story with premises true and conclusion false,
  - 1) Some collies can fly
  - 2) All collies are dogs
  - 3) No dogs can fly
- b. In any such story,
 

Given (1) and (2),

  - 4) Some dogs can fly

Given (3) and (4),

  - 5) Some dogs can and cannot fly
- c. So any such story is inconsistent. So there is no consistent story that makes the premises true and the conclusion is false. So, by definition, the argument is logically valid.

- d. Since in the real world no collies can fly, not all the premises are true in the real world. So, though the argument is logically valid, it is not logically sound.

The reasoning at (b), (c) and (d) is correct. Any story which contains (1), (2) and (3) is inconsistent. You should be able to follow each step. But something is wrong—there is a mistake at (a): It's not the case that *any* story which makes the premises true and conclusion false makes (3) true. The negation of "All dogs can fly" isn't "No dogs can fly," but rather "Not all dogs can fly" ("Some dogs can't fly"). We have indeed shown that every story of a certain sort is inconsistent—but haven't shown that every story which makes the premises true and conclusion false is inconsistent. In fact, as we have seen, there *are* consistent stories that make the premises true and conclusion false. Similarly, in step (b) it is easy to get confused if you consider *too much* information at once. Ordinarily, if you focus on sentences singly or in pairs, it will be clear what must be the case in every story including those sentences. It doesn't matter which sentences you consider in what order, so long as you reach a contradiction in the end.

With this in mind, use the validity format to show that the following arguments are logically valid and to decide (if you can) whether they are logically sound:

- (6) If Bill is president, then Hillary is first lady  
       Hillary isn't first lady  
       \_\_\_\_\_   
       Bill isn't president
- (7) Only fools find love  
       Elvis was no fool  
       \_\_\_\_\_   
       Elvis didn't find love
- (8) If there is a good and omnipotent god, then there is no evil  
       There is evil  
       \_\_\_\_\_   
       There is no good and omnipotent god
- (9) All sparrows are birds  
       All birds fly  
       \_\_\_\_\_   
       All sparrows fly

- (10) All citizens can vote  
       Hannah is a citizen  
       —  
       Hannah can vote

Again, the failure of one format doesn't, by itself, demonstrate that the other will work—for the failure may be due to some lack of creativity or imagination on our part. In courses on formal logic, powerful methods are developed for identifying validity and invalidity. Still, working with the formats should help us to see what validity and soundness *are*—and so what the methods of logic are designed to accomplish.

## II. Some Consequences

(A) First, a consequence we want. The conclusion of every sound argument is true in the real world. To see this, suppose we have a *sound* two-premise argument, and consider the true story. The premises and conclusion must fall into one of the following combinations of true and false:

1	2	3	4	5	6	7	8
T	T	T	F	T	F	F	F
T	T	F	T	F	T	F	F
—	—	—	—	—	—	—	—
T	F	T	T	F	F	T	F

If the argument is sound, it is valid; so no consistent story makes the premises true and the conclusion false. But the true story is a consistent story. So we can be sure that the true story doesn't result in combination (2). So far, the true story might fall into *any* of the other combinations. Thus the conclusion of a valid argument may or may not be true in the real world. But if an argument is sound, its premises are true in the real world. So, for a sound argument, the true story doesn't result in any of the combinations (3) - (8). (1) is the only combination left: if an argument is sound, its conclusion is true in the real world: If there is *no* consistent story where the premises are true and the conclusion is false, and the premises are in fact true, then the conclusion must be true as well. Notice: we don't need that the conclusion is true in the real world in order to say that an argument is sound; rather, in discovering that an argument is sound—by discovering that it is valid and its *premises* are true—we *establish* that its conclusion is true. And this is just what we want.

- (B) Another consequence seems less welcome. Consider the following argument:

Snow is white  
 Snow isn't white  
 —

All dogs can fly

It's natural to think that the premises aren't connected to the conclusion in the right way—for the premises have nothing to do with the conclusion—and that this argument therefore shouldn't be valid. But if it isn't valid, by definition, there is a consistent story that makes the premises true and the conclusion false—and in this case, there is no such story, for *no consistent story makes the premises true*. Thus, by definition, this argument is logically *valid*. The format applies in a straightforward way. Thus,

(a) In a story with premises true and conclusion is false,

- 1) Snow is white
- 2) Snow isn't white
- 3) Some dogs can't fly

(b) In any such story,  
Given (1) and (2),

- 4) Snow is and isn't white

(c) So any such story is inconsistent. So there is no consistent story that makes the premises true and the conclusion false. So, by definition, the argument is logically valid.

(d) Since in the real world snow is white, not all of the premises are true in the real world. So, though the argument is logically valid, it is not logically sound.

This seems bad! Intuitively, there is something wrong with the argument. But, on our official definition it is logically valid. One might rest content with the observation that, even though the argument is valid, it isn't sound. But this doesn't remove the general worry. For this argument,

Grass is green

---

$1 + 1 = 2$

has all the problems of the other, and is logically *sound* as well. (Why?) One might, on the basis of examples of this sort, decide to reject the (classical) account of validity with which we have been working. Some do just this.<sup>3</sup> But let's see what can be said in defense of the classical approach. (And the classical approach *is*, no doubt, the approach you have or will encounter in any introductory course on logic.)

---

<sup>3</sup> Especially the so-called "relevance" logicians. Relevant systems are best taken up after a course in classical logic. Those with an introduction to truth tables might find §1.3 and §16.3 of J. Nolt, *Logics* (Belmont: Wadsworth Publishing Co., 1997) an accessible introduction. Cf., e.g., Routley, Meyer, et. al., *Relevant Logics and Their Rivals I*, (Atascadero: Ridgeview Publishing Co., 1982).

As a first line of defense, one might observe that the conclusion of every sound argument is true, and ask, “What more do you want?” We use arguments to demonstrate the truth of conclusions. And nothing we have said suggests that sound arguments don’t have true conclusions: an argument whose premises are inconsistent, is sure to be unsound—for at least one premise won’t be true in the real world; and an argument whose conclusion can’t be false, is sure to have a true conclusion! So soundness may seem sufficient for our purposes. Even though we accept that there remains *something* about argument goodness that soundness leaves behind, we can insist that soundness is useful as an intellectual *tool*. Whenever it is the truth or falsity of a conclusion that matters, we can profitably employ the classical notions.

But one might go farther, and dispute even the suggestion that there is something about argument goodness that soundness leaves behind. Consider the following two principles:

(A)  $\mathcal{P}$  or  $\mathcal{Q}$ , not- $\mathcal{P}$   


---

  
 $\mathcal{Q}$

(B)  $\mathcal{P}$   


---

  
 $\mathcal{P}$  or  $\mathcal{Q}$

According to principle (A), if you are given that  $\mathcal{P}$  or  $\mathcal{Q}$ , and that not- $\mathcal{P}$ , you can conclude that  $\mathcal{Q}$ . If you have cake or ice cream, and you don’t have cake, you have ice-cream; if you are in California or New York, and you aren’t in California, you are in New York; etc. Thus (A) seems hard to deny. And similarly for (B). Where “or” means “one or the other or both,” when you are given that  $\mathcal{P}$ , you can be sure that  $\mathcal{P}$  or anything. Say you have cake; then you have cake *or* ice cream, cake *or* brussels sprouts, etc.; if grass is green, then grass is green or pigs have wings, grass is green or dogs fly, etc.

Return now to our problematic argument. As we’ve seen, it is valid according to the classical definition. We get a similar result when we apply principles (A) and (B):

- |    |                                   |                       |
|----|-----------------------------------|-----------------------|
| 1. | Snow is white                     | premise               |
| 2. | Snow isn’t white                  | premise               |
| 3. | Snow is white or all dogs can fly | from 1 with (B)       |
| 4. | All dogs can fly                  | from 2 and 3 with (A) |

If snow is white, then snow is white or anything. So snow is white or dogs fly. But if snow is white or dogs fly, and snow isn’t white, then dogs fly. If we want to deny the concision—if we want to reject the validity of this argument and so our classical notion of validity—we have to reject one of our principles (A) or (B). But it’s not obvious which

should go.<sup>4</sup> If we have intuitions according to which (A) and (B) should stay, and also that the definition of validity should go, then we have conflicting intuitions. Thus our intuitions might, at least, be sensibly resolved in the classical direction.

The issues are complex, and are the subject for courses in advanced logic, and philosophy of logic. It is enough for us to treat the classical approach as a useful tool: it is useful in contexts where what we care about is whether conclusions are true. Also, as should be apparent from the main body of the text, this approach to argument evaluation—depending, as it does, on *possibility* and *necessity*—raises interesting and challenging *metaphysical* questions.

Use our formats to say whether the following are logically valid or invalid, and sound or unsound. Hint: You may have to do some experimenting to decide whether the arguments are logically valid or invalid—and so which format applies.

- (11) If Bill is president, then Hillary is first lady

Bill is president

\_\_\_\_\_

Hillary is first lady

- (12) Most professors are insane

TR is a professor

\_\_\_\_\_

TR is insane

- (13) The earth is (approximately) round.

\_\_\_\_\_

There is no round square

- (14) Some dogs have red hair

Some dogs have long hair

\_\_\_\_\_

Some dogs have long red hair

---

<sup>4</sup> The relevance logicians reject (A).



(15) All dogs can fly  
Fido is a dog  
Fido cannot fly

\_\_\_\_\_

I am blessed

For further thought: Should we accept the classical account of validity? Explain your position, with special reference to difficulties raised above.

### Further Reading

This is an extremely select bibliography, meant only to suggest good places to start. For other sources, the best place to look is in their footnotes and/or bibliographies. This is particularly effective if a work is recent. Another excellent resource is *The Philosopher's Index*. The *Index* is a *Reader's Guide*-type tool for philosophers. It is available in the reference section of most academic libraries, often in an electronic format.

(A) *Things and parts*. Recent books are Mark Heller, *The Ontology of Physical Objects* (Cambridge: Cambridge University Press, 1990); Peter van Inwagen, *Material Beings* (Ithaca: Cornell University Press, 1990); and Michael Jubien, *Ontology, Modality, and the Fallacy of Reference* (Cambridge: Cambridge University Press, 1993). There's an accessible introduction to Jubien's view in chapter 9 of his text, *Contemporary Metaphysics* (Malden: Blackwell Publishers, 1997). Recent articles with good bibliographies are Theodore Sider, "Four Dimensionalism," *The Philosophical Review* 106 (1997): 197-231; and Michael Rea, "Temporal Parts Unmotivated," *The Philosophical Review* 107 (1998): 225-260.

(B) *Possibility and necessity*. Standard works are, Saul Kripke, *Naming and Necessity*, (Cambridge: Harvard Univ. Press, 1980); David Lewis, *On The Plurality of Worlds*, (New York: Basil Blackwell, 1986); Alvin Plantinga, *The Nature of Necessity*, (Oxford: Oxford Univ. Press, 1974); and the essays in M. Loux, ed., *The Possible and the Actual*, (Ithaca: Cornell Univ. Press, 1979). Again, there's a chapter in Jubien, *Contemporary Metaphysics*. If you are interested in my own views, see Roy, "Worlds and Modality," *The Philosophical Review* 102 (1993): 335-361; "In Defense of Linguistic Ersatzism," *Philosophical Studies* 80 (1995): 217-242; and "Things and *De Re* Modality," *Noûs* 34 (2000): 56-84.

(C) *Properties and universals*. Good places to start are Mellor and Oliver, eds. *Properties* (Oxford: Oxford University Press, 1997); and David Armstrong, *Universals: An Opinionated Introduction* (Boulder: Westview Press, 1989). See also the chapter on Platonism in Jubien, *Contemporary Metaphysics*; and Hilary Putnam, "On Properties," in his *Mathematics Matter and Method: Philosophical Papers Vol I*, 2nd ed. (Cambridge: Cambridge University Press, 1979): 305-322. In addition, you may find the essays in M. Loux, ed., *Universals and Particulars* (Garden City: Anchor Books, 1970), to be helpful. That collection contains an essay by A. Donagan in which he develops some themes from your Russell reading; B. Aune replies in, "On Postulating Universals," *Canadian Journal of Philosophy* 3 (1973): 285-94.

(D) *Events*. Donald Davidson is a "lightning rod" for much discussion of events; of particular importance are essays 6-8 in his *Essays on Actions and Events*, (New York:

Oxford University Press, 1980). These are not as easy to follow as one might like, but you can give them a try. Other sources are chapter 4 of R. Chisholm, *Person and Object* (London: George Allen and Unwin, 1976); along with the articles in *Actions and Events*, ed., LePore and McLaughlin, (New York: Basil Blackwell, 1985)—which has a nice bibliography; and *Essays on Davidson*, ed., Vermazen and Hintikka, (New York: Oxford University Press, 1985).

(E) *Mathematical truth*. Many discussions of mathematical truth can be read only after extensive training in mathematics and logic. Still, a good place to start is *The Philosophy of Mathematics*, ed. Hart (Oxford: Oxford University Press, 1996). Relatively accessible books are Penelope Maddy, *Realism in Mathematics*, (Oxford: Clarendon Press, 1990); and John Bigelow, *The Reality of Numbers* (Oxford: Oxford University Press, 1988). Paul Benacerraf, “What Numbers Could Not Be,” *Journal of Philosophy* 70 (1973): 661-680, develops a classic objection to realism. See also Hilary Putnam, “Philosophy of Logic,” in his *Mathematics Matter and Method: Philosophical Papers Vol I*, 2nd ed. (Cambridge: Cambridge University Press, 1979); and, for a response to Putnam, Bas van Fraassen, “Critical Notice,” *Canadian Journal of Philosophy* 4 (1975): 731-743.

(F) *Fictional objects*. Amie Thomasson, *Fiction and Metaphysics* (Cambridge: Cambridge University Press, 1999) is a particularly clear and accessible place to start. For other approaches, see the chapter in Jubien, *Contemporary Metaphysics*; David Lewis, “Truth in Fiction,” in his *Philosophical Papers Vol. I* (Oxford: Oxford University Press, 1983); and Terence Parsons, “A Meinongian Analysis of Fictional Objects, *Grazer Philosophische Studien* 1 (1975): 73-86—and for an extended treatment of his view, *Nonexistent Objects* (New Haven: Yale University Press, 1980). R. Bertolet, “Reference, Fiction, and Fictions,” *Synthese* 60 (1984): 413-37, and H. Wettstein, “Did the Greeks Really Worship Zeus?” in the same journal: 439-449, combine to put pressure on Parsons in the direction of something like Lewis. Edward Zalta, *Abstract Objects* (The Netherlands: Reidel, 1983) develops another (relatively technical) Meinongian approach.

(G) *Reality and truth*. This isn’t easy stuff. (Depending on how things go, I may be able to help with at least one chapter for this edition of *About What There Is*; but we will have to see how things go.) You might get started with chapter 4 of Peter van Inwagen’s text, *Metaphysics* (Boulder: Westview Press, 1993); and/or chapter 5 of Jubien, *Contemporary Metaphysics*. For original sources on realism, see William Alston, *A Realist Conception of Truth* (Ithaca: Cornell University Press, 1996); Alston, “Yes Virginia, There is a Real World,” *Proceedings and Addresses of the American Philosophical Association* 52 (1979): 779-808; David Lewis, “Putnam’s Paradox,” *Australasian Journal of Philosophy* 62 (1984): 221-236; and Michael Devitt, *Realism and Truth*, (Princeton: Princeton University Press, 1984). For antirealism, see Hilary Putnam, “Why There Isn’t a Ready-Made World,” in *Realism and Reason: Philosophical Papers vol. 3*, (New York: Cambridge University Press, 1983); the first three chapters of Putnam,

*Reason Truth and History*, (Cambridge: Cambridge University Press, 1981); and the first two chapters of Putnam, *The Many Faces of Realism*, (LaSalle: Open Court, 1987). Other prominent anti-realist positions are Nelson Goodman, *Ways of Worldmaking*, Indianapolis: Hackett Publishing Company, 1978); and Richard Rorty, *Philosophy and the Mirror of Nature*, (Princeton: Princeton University Press, 1979).

## Assignment Schedule

### *Chapter 1*

Read over the entire chapter, noting your questions and comments the margin and on blank facing pages. Then,

A1 Read Roy, 2-6.

1. In your own words, what is the “general theoretical difficulty” for the classical approach to metaphysics? Do you think this difficulty is decisive?
2. In your own words, what is the “general theoretical difficulty” for the Kantian approach to metaphysics? Do you think this difficulty is decisive?
3. Given **T1**, why is sentence (4) problematic? What problems might there be for the suggestion that the *Fanfare* is marks on paper? that it is some (particular) sound waves in the air?

A2 Read Roy, 6-11.

1. What is the second objection against Platonism? How does it contrast with the first? Do you think it is successful? That is, how strong should our bias in favor of the concrete be? Explain.
2. On p. 10, Roy suggests that appeal to mind-independent ideas, or to ideas in the mind of god, might help with problems of conceptualism. What is the problem, and how is this appeal supposed to help? In what sense, if any, does it tend toward Platonism?
3. On p. 11, Roy suggests that appeal to mind-independent ideas, or to ideas in the mind of god, might help with a different problem of conceptualism. What is the problem, and how is this appeal supposed to help? Does it? Explain.

A3: Read Roy, 11-16.

1. Give a full nominalistic expansion of (5). Hint: You will want to say that *the dwarfs* has seven parts which are, individually, dwarfs. So your answer will look like, “There are seven dwarfs in the forest iff the dwarfs is in the forest and has a dwarf-part *a*....”

2. Suppose the universe consists of just four “ordinary” things *a*, *b*, *c* and *d*. Given this, what things does Goodman think there are? *How many* things does he think there are? Now explain, in your own words, why Goodman’s approach to groups doesn’t solve the problem about infinity.
3. Suppose a deductivist wants to understand the numerals ‘1’, ‘2’, ‘3’, ‘4’, ‘5’ etc. as standing in for referents to certain *inscriptions*,

/, //, ///, ////, /////, etc.,

and wants to understand ‘+’ as concatenation; so that, e.g.,  $3 + 2 = 5$ ,

/// + // = /////

Given that “for any number there is one greater than it” is a premise for ordinary arithmetic, how does this *deductivist* have a problem about using algorithms from ordinary arithmetic to reach the conclusion that  $245 + 117 = 362$ ? Hint: How many inscriptions are there? will the premise be *true*?

A4: Read Roy, 16-23.

1. Suppose you are in a religious discussion and someone says, “I don’t care what is true for you, it is true for me that god does not exist.” Taking literally and seriously this language about truth, given the correspondence theory, how is there a problem about truth?
2. Suppose you are in a religious discussion and someone says, “I don’t care about your reality, in my reality there is no god.” Taking literally and seriously this language about reality, how do you think Roy would respond? Do you think he is right?
3. Is your reaction to the metaphysical project, “how fascinating” or “it’s uninteresting”? Explain. (Feel free to advocate either view!)

EX1: In a paper of about four pages, *attack* one of the following claims. Be sure that your paper responds to issues raised in the text.

The classical and Kantian approaches to metaphysics are fatally flawed.  
 Platonism, as a theory of numbers, is fatally flawed.  
 Nominalism, as a theory of numbers, is fatally flawed.  
 Conceptualism, as a theory of numbers, is fatally flawed.  
 Deductivism, as a theory of numbers, is fatally flawed.  
 No proposition is true for some people and false for others.

## *Chapter 2*

Read over the entire chapter, and Quine “On What There Is,” at least through the second paragraph on e5, noting your questions and comments the margin and on blank facing pages. Then,

A5: Read Quine OWTI, through the second full paragraph on e3, and Roy, 25-30.

1. How does the “beard” argument work? Do you see any way out of the conclusion that there is some thing to which “Pegasus” refers? Explain.
2. How is it that McX postulates systematic predicate ambiguities? Do you think these ambiguities undercut the plausibility of McX’s view?
3. Roy suggests, on Quine’s behalf, that McX misunderstands the force of the “beard” argument. Explain. Do you think this is right? Why?

A6: Read Quine OWTI, through the first sentence of the first full paragraph on e4, and Roy, 31-34. Respond to the following:

1. Quine says that “Wyman... is one of those philosophers who have united in ruining the good old word ‘exist’.” Explain.
2. In what sense is Wyman’s view supposed to be superior to McX’s? Do you think it is? Explain.
3. What are the different rationales proposed for Occam’s razor? How is the second supposed to cut closer than the first? Which, if any, do you think is right? Explain.

A7: Read Quine OWTI, through the first two full paragraphs on e4, and Roy, 34-40. Respond to the following:

1. Roy claims claim there is at least a 3-way ambiguity about “criteria of identity.” Explain each of the different senses.
2. The Zeus case, with the reincarnation model, is supposed to be a problem for the anti-Wyman argument on p. 37. How does that case undercut the argument?
3. The Zeus case, with the reincarnation model, is supposed to support the anti-Wyman argument on p. 39. How does that case support the argument?

A8: Read Quine OWTI, through the first two full paragraphs on e5, and Roy, 40-43. Respond to the following:

1. Roy suggests a response for Wyman to the “cupola” case different than the one Quine gives him. What are these two responses?
2. Quine rejects the doctrine that whatever is contradictory is meaningless. Explain and evaluate one of his two main reasons.
3. Roy suggests that the combination of the “Parthenon” and “fat-man” objections is fatal to **T1**. How is this supposed to work?

EX2: In a paper of about four pages, *attack* one of the following claims. Be sure that your paper responds to issues raised in the text.

Occam’s razor establishes a bias in favor of “desert landscapes.”  
 The “Parthenon” objection is fatal to McX’s use of the “beard” argument.  
 The “fat-man” objection is fatal to Wyman’s use of the “beard” argument.  
 A statement may be meaningful, though contradictory.

### *Chapter 3*

Read over the entire chapter, and Quine “On What There Is,” at least through the first full paragraph on e8, noting your questions and comments the margin and on blank facing pages. Then,

A9: Read Roy, 44-52. Respond to the following (no need to type):

1. For each of the expressions below, say whether it is a formula. If not, say why. If so, demonstrate with a diagram.
2. For each of the expressions that is a formula, what is its main operator? what are its immediate subformulas?
3. For each of the expressions that is a formula, use underlines to indicate the scope of each quantifier, and say if it is a sentence. If it is not a sentence, say why.



- |  |   |
|--|---|
| a) $(\exists x)Fx$   | g) $\sim[Fx \wedge Gy] \wedge Hz$                                 |
| b) $(\sim Fx \wedge Gy \wedge Hz)$                           | h) $\forall x \sim[Fx \wedge Gy] \wedge Hz$                       |
| c) $\forall z \exists x (Fzx \leftrightarrow Kxz)$           | i) $\forall x (\sim[Fx \wedge Gy] \wedge Hz)$                     |
| d) $\forall z \sim \exists w (Fzx \leftrightarrow Kxz)$      | j) $\forall x (\sim[Fx \wedge Gy] \wedge \exists z Hz)$           |
| e) $(\mathcal{P}x \wedge \mathcal{Q}y)$                      | k) $\forall x (\forall y \sim[Fx \wedge Gy] \wedge \exists z Hz)$ |
| f) $\forall z \exists x (\exists z Fzx \leftrightarrow Kxz)$ | l) $\forall x \forall y (\sim[Fx \wedge Gy] \wedge \exists z Hz)$ |

A10: Read Roy, 53-66. Respond to the following (no need to type):

1. Assume  $a$ ,  $b$  and  $c$  are the only things in the universe; assume further that  $a$  is a dog but  $b$  and  $c$  are not; that  $a$  and  $b$  are lucky but  $c$  is not; that  $a$  barks at  $b$  and  $c$  but that there are no other barking relations, and let predicate letters be as in the text. That is, consider an interpretation  $I$  with  $U = \{a, b, c\}$ ,  $I(D) = \{a\}$ ,  $I(L) = \{a, b\}$  and  $I(B) = \{\langle a, b \rangle, \langle a, c \rangle\}$ . On this interpretation, for each of the sentences below, use the tree method to say whether the sentence is T or F.
2. Setting aside the specific assumptions from problem (1)—but keeping trees in mind—on the model of the text, explain the range of conditions under which each of the sentences below is true. Thus, e.g., the requirement for (a) is that not everything is a dog.

- |                              |  |
|------------------------------|--|
| a) $\sim \forall x Dx$       | f) $\forall y \exists x Bxy$                                 |
| b) $\sim \forall x \sim Dx$  | g) $\exists y \forall x Bxy$                                 |
| c) $\sim \exists x \sim Dx$  | h) $\forall x (Dx \vee \sim Lx)$                             |
| d) $\forall x \exists y Bxy$ | i) $\sim \exists x (Dx \wedge Lx)$                           |
| e) $\exists x \forall y Bxy$ | j) $\exists x \exists y ([Dx \wedge Dy] \wedge \sim(x = y))$ |

A11: Read Quine OWTI, through the first full paragraph on e8, and Roy, 66-72.

1. In your own words, carefully say why  $\exists x (\mathcal{P}x \wedge \forall y (\mathcal{P}y \rightarrow x = y) \wedge \mathcal{Q}x)$  satisfies Russell's three conditions.
2. Roy suggests that Quine's name-to-predicate suggestion is open to the objection that "derived predicates themselves require objects of reference." Suppose they do require objects of reference. Why is this supposed to be a problem?
3. What is it to say that reference is *sufficient* for meaning? What is it to say that reference is *necessary* for meaning? What does Frege's example seem to establish? What does the present-king-of-France example seem to establish? Explain.

EX3: In a paper of about four pages, *attack* one of the following claims. Be sure that your paper responds to issues raised in the text.

A successful mode of regimentation need not “translate” ordinary language.  
 Quine, with Russell, successfully shaves Plato’s beard.  
 “Derived” predicates are meaningful apart from objects of reference.

#### Chapter 4

Read over the entire chapter, and Quine “On What There Is,” through the end, noting your questions and comments the margin and on blank facing pages. Then,

A12: Read Quine OWTI, through the second full paragraph on e10, and Roy 73-77.

1. For each of the following “theories,” say what entities it is committed to (according to Q1). Assume an ordinary understanding of ( $x$  is a parent of  $y$ ). In each case, explain your answers.
  - a.  $\exists x(x \text{ is a dog})$
  - b.  $\forall x(x \text{ is a dog} \rightarrow \exists y(y \text{ is a dog} \wedge y \text{ is a parent of } x))$
  - c.  $\exists x(x \text{ is a dog}), \forall x(x \text{ is a dog} \rightarrow \exists y(y \text{ is a dog} \wedge y \text{ is a parent of } x))$
2. What is a universal? Given this, explain how the argument at the top of p. 77 is charged with equivocation.
3. Do you think there are universals? Explain.

A13: Read Russell, “The World of Universals,” e16-e20, and Roy, 77-88.

1. What generates the resemblance series Roy claims is vicious? Why think that it is a “vicious” regress? Do you agree? Explain.
2. How does Quine block the regress proposed by McX (Russell)? Do you find his position plausible? Explain.
3. Roy suggests that “whatever may be the outcome of this debate about properties, again, Quine seems to have found a “neutral” mode of expression which makes the debate possible.” Say this is right. How is it important for the evaluation of Quine’s criterion of ontological commitment?

A14: Read Quine OWTI, through the end, and Roy 88-95.

1. Think a bit about a truth condition for ‘some zoölogical species are cross-fertile’. What do you think it should be? Either give and defend a condition, or discuss difficulties associated with doing so.
2. Use the principles developed in the text to show that  $2 < 3$  (that  $2 < 1 + 2$ ). Then see if you can combine this with transitivity (the last of the order principles from the boxed page) to show that  $1 < 3$  – and so that Fred is less than three feet tall.
3. In the last pages of Quine’s article, do you think he says that all metaphysics is myth making, or allows that metaphysical theorizing aims at truth? Explain. What is your view? Explain.

EX4: In a paper of about four pages, *attack* one of the following claims. Be sure that your paper responds to issues raised in the text.

To be is to be the value of a variable.  
 The first resemblance series discussed by Roy isn’t vicious.  
 The second resemblance series discussed by Roy is vicious.  
 Quine successfully deflects McX’s argument for properties.  
 We should include numbers among the things we think exist.

### Chapter 5

Read over the entire chapter, and the parts of Quine, “Reference and Modality” in e21-e34, noting your questions and comments the margin and on blank facing pages. Then,

A15: Read Roy 96-105.

1. Why might someone “opt out” of Quine’s method? Given this, how does TC make opting out have the virtues of “theft over honest toil”?
2. Produce sentences of your own which fail E1, E2 and E3. In each case, explain how your sentence fails the condition.
3. Construct trees for ‘ $\forall x(Fx \rightarrow \exists yGxy)$ ’ and ‘ $\forall x(Fx \rightarrow \exists yHxy)$ ’. On the assumption that ‘G’ and ‘H’ are coextensional, how can you be sure that the sentences have the same truth value?

A16: Read Quine through e30, and Roy 105-108

1. How do E1, E2 and E3 *restrict* what sentences say about the world? Why might one think that sentences meeting this restriction are sufficient to say all we need to say about the world?
2. Quine thinks truth value switches when his (15) is changed to his (18), but he doesn't say there is a related switch for (30). What, then, does he say about (30), and how are the singular term (§1) and quantifier (§2) cases related?
3. Do you accept the following argument?

Any truth condition may be expressed in an extensional language.  
No truth condition for 'possibly I shall win the lottery' has expression in an extensional language.

\_\_\_\_\_  
'possibly I shall win the lottery' has no truth condition.

Explain your position, with attention to validity and soundness.

A17: Read Quine through e34, and Roy 108-117

1. What do you think Quine has in mind by "Aristotelian Essentialism"? How does this approach to necessity differ from the one based on analyticity?
2. Explain how Q1 and Q2 combine to pressure us to allow *times* into our ontology. Do you think we should therefore do so?
3. Construct a tree for the following condition (with appropriate abbreviations):

Q1c  $\exists u(u \text{ satisfies } t) \rightarrow [\text{theory } t \text{ is committed to entities of kind } K \leftrightarrow \forall u(\text{if } u \text{ satisfies } t, \text{ then } u \text{ contains } Ks)]$

Assuming no set contains unicorns, how is this condition problematic?

EX5: In a paper of about four pages, *attack* one of the following claims. Be sure that your paper responds to issues raised in the text.

We should accept Q1.  
We should accept Q2.  
A context "goes extensional" when we get the objects right.  
We should accept that *Franklin* has *being possibly a janitor*.  
We should accept that there are "ways the world could have been."

*Chapter 6*

A18: Read Putnam, “Is There Still Anything to Say About Reality and Truth” and Roy, 119 - 128, noting your questions and comments the margin and on blank facing pages.

1. Consider the case of intentionality discussed by Putnam but mostly passed over in the by Roy. Does it fit into the argument schema discussed by Roy? Do his criticisms apply? Explain.
2. Carefully explain the solution of the watch case given ER! Just what things are under consideration in the different premises? Why is this solution unavailable given ER alone?
3. Consider a clay statue, along with one culture that thinks exclusively in terms of things as lumps of clay and another that thinks in terms of things as statues. Observe that a lump may be remoldable where a statue is not so that, by the indiscernability of identicals, they are not the same thing differently named. Now in some context, we have the argument,

- 1 From the perspective of conceptual scheme 1, there is a statue.
- 2) From the perspective of conceptual scheme 2, there is a lump.
- \_\_\_\_\_
- 3) What things there are is scheme-relative

Given discussion from the text, how might the realist reply? Explain how the response defuses the argument. Do you see any special problems for this response compared to ER and ER!?

A19: Read Putnam, “A Problem About Reference” and Roy, 129 - 137, noting your questions and comments the margin and on blank facing pages.

1. Consider the interpretation  $I$  for A10.1 with  $U = \{a, b, c\}$ ,  $I(D) = \{a\}$ ,  $I(L) = \{a, b\}$  and  $I(B) = \{\langle a, b \rangle, \langle a, c \rangle\}$ . Use the method of permutations to produce an isomorphic interpretation  $J$  that has different reference for the predicates, but keeps truth values for all sentences the same. Use trees to show that sentences (a) and (i) from that exercise have the same truth values on the two interpretations. (If you still have your work from A10 – done correctly! – you may simply xerox trees for interpretation  $I$  into this exercise, otherwise those trees should be redone here.)

2. Consider Putnam's discussion of intentions. Why does he hold that the having of intentions presupposes the ability to refer? Relate this to the point about mappings from the first part of this section of the text.
3. On (e81) of "Two Philosophical Perspectives" Putnam says, "If I say 'the word "horse" refers to objects which have a property which is connected with my production of the utterance "There is a horse in front of me" on certain occasions by *a causal chain of the appropriate type*', then I have the problem that, if I am able to specify what *is* the appropriate type of causal chain, I must *already* be able to refer to the kinds of things and properties that make up that kind of causal chain. But how did I get to be able to do this?" Taking this in its context, how would Roy see it as problematic as a response against the causal theory?

A20: Read Putnam, "Two Philosophical Perspectives" and Roy, 137 to the end, noting your questions and comments the margin and on blank facing pages..

1. Consider what Putnam says about "conceptual contamination" at the top of e73. With our discussion from "Is There Still Anything To Say About Reality And Truth?" in mind, is there anything there with which a realist should disagree? Explain. In what sense (if any) does a realist think there is One True Theory?
2. In the text Roy suggests that Putnam's view requires modifications in the direction of realism with respect to minds and the world, and in the direction of correspondence with respect to truth. Explain how these count as objections about the status of (R) and the entities *X* with respect to a theory of the sort (R), "Everything is relative to *X*."
3. Kant's view is sometimes explained by the analogy of rose-colored glasses: though everything looks red through our glasses, we are not entitled to conclude that all is red. Under what conditions do you think we would be entitled to make *no* conclusions about color? Similarly, what relations between minds and the world are required for it to follow that we are entitled to no conclusions about the world? Explain.

EX6: In a paper of about four pages, *attack* one of the following claims. Be sure that your paper responds to issues raised in the text.

- Putnam's attack on (R1) from fundamental physics fails.
- Putnam's attack on (R1) from conceptual schemes fails.
- Putnam's attack on (R2) from the problem about reference fails.
- Putnam's internalist view requires "charitable" modifications as in the text.

In paper of about four pages, critically discuss some instance of metaphysical antirealism in everyday thought or in academia outside philosophy as such.

*Review, parts I - III*

A21 The assignment is NOT to answer each of the following questions! Rather, you are to produce at least three written (content) questions for class discussion in review of chapters 1 - 6. In the end, you should be prepared to respond to any any of the following:

1. Compare and contrast the classical, Kantian and contemporary approaches to metaphysics (as in ch. 1). What pressures have led to the contemporary approach? Do you think it is the appropriate way to go? Explain.
2. Set up the (Platonist/Conceptualist/Nominalist/Deductivist) approach to mathematics and critically discuss one objection to it.
3. Explain the sense in which ordinary claims are supposed to generate metaphysical discussion and to count as data for it. Critically discuss the question whether this makes metaphysics a science.
4. Critically discuss Quine's main, reference switching, objection against McX. Do you think it is successful? For this, you will have to expound McX's position and the reference switching objection; then you will be in a position to evaluate.
5. Critically discuss the question whether Quine's questions about the fat man in the doorway generate a problem about criteria of identity for unactualized possibles. For this, you will have to say what unactualized possibles are and expound the way Quine's questions are supposed to apply; then you are in a position to evaluate. You should demonstrate awareness of Roy's position.
6. Critically discuss the question whether Quine's questions about the fat man in the doorway generate a problem about reference for unactualized possibles. For this, you will have to say what unactualized possibles are and expound the way Quine's questions are supposed to be a problem; then you will be in a position to evaluate.
7. Assume that  $a$ ,  $b$  and  $c$  are the only things in the universe, that  $a$  loves  $a$ ,  $b$  and  $c$ , that  $b$  loves  $b$  and  $c$ , and that  $c$  loves  $c$ , and that there are no other loving relations. On these assumptions, use the tree method to evaluate the

truth of the following: i)  $\forall x \exists y (x \text{ loves } y)$ ; ii)  $\exists y \forall x (x \text{ loves } y)$ . What are the general conditions under which these expressions are true?

8. How would Quine say 'Pegasus is winged' (using his canonical notation)? Carefully explain the conditions under which this function is true or false. How would Quine say 'Pegasus does not exist' (using his canonical notation)? Carefully explain the conditions under which the function is true or false.
9. Critically discuss the question whether Quine (Russell) has solved Plato's beard. For this, you will want to say what Plato's beard is, why one might think that it has been solved, and consider reasons to worry.
10. What is Quine's criterion of ontological commitment? Why might one think that it *doesn't* commit us to enough stuff? Do you think Quine's position is right? Explain. This is partly a question about "Plato's stubble" and whether we should be committed by predicates.
11. What is Quine's criterion of ontological commitment? How might it help us to understand and think about ontological commitments corresponding to a claim like, "some zoological species are cross-fertile"? Do you think the resultant conclusions are right? Explain. You should demonstrate awareness of discussion from text and lecture.
12. What is Quine's criterion of ontological commitment? How might it help us to understand and think about ontological commitments corresponding to a claim like, "Fred is one foot high"? Do you think the resultant conclusions are right? Explain. You should demonstrate awareness of discussion from text and lecture.
13. What is Quine's criterion of ontological commitment? What is an *extensional* language? Is it plausible to think that any aspect of the world is describable in such a language? Explain. How is this discussion relevant to evaluation of Quine's criterion of ontological commitment?
14. What is Quine's criterion of ontological commitment? How does it put pressure on us to include *times* within our ontology? Do you think this is right? Explain.
15. What is Quine's criterion of ontological commitment? How does it put pressure on us to include *ways the world could have been* within our ontology? Do you think this is right? Explain.
16. Critically discuss Putnam's attack on (R1) from fundamental physics.



17. Critically discuss Putnam's attack on (R1) from alternative conceptual schemes, together with the plausibility of the "exploded reality" response.
18. Critically discuss Putnam's attack on (R2) from the problem about reference. How effective are his responses to objections? Critically discuss.
19. Critically discuss Putnam's "internalist" view, and especially the question whether it requires modifications to allow for minds and the world along realist lines, and for correspondence truth.
20. What do you think there is? Critically discuss with special attention to the question of abstract and concrete objects.

