



W. V. QUINE

Confessions of a Confirmed Extensionalist
and Other Essays

EDITED BY DAGFINN FØLLESDAL AND DOUGLAS B. QUINE

Confessions of a Confirmed Extensionalist
and Other Essays

Confessions of a Confirmed Extensionalist

and Other Essays

W. V. Quine

Edited by

Dagfinn Føllesdal

Douglas B. Quine

Harvard University Press
Cambridge, Massachusetts
London, England
2008

Copyright © 2008 by the President and Fellows of Harvard College

All rights reserved

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Quine, W. V. (Willard Van Orman)

Confessions of a confirmed extensionalist and other essays / W. V. Quine ; edited by Dagfinn

Føllesdal, Douglas B. Quine.

p. cm.

Includes bibliographical references and index.

ISBN-13: 978-0-674-03084-8 (alk. paper)

1. Philosophy. I. Føllesdal, Dagfinn. II. Quine, Douglas B., 1950- III. Title.

B29.Q55 2008

191—dc22 2008010061

*To the students of Quine:
past, present, and future*

Contents

Acknowledgments	<i>xi</i>
-----------------	-----------

Introduction	<i>1</i>
--------------	----------

I Previously Unpublished Articles

1	Nominalism (1946)	<i>7</i>
2	On the Notion of an Analytic Statement (1946)	<i>24</i>
3	Lectures on David Hume's Philosophy (1946)	<i>36</i>
4	The Importance of Logic for Philosophy (1947)	<i>137</i>
5	Where Logic Is Going (1947)	<i>148</i>
6	Animadversions on the Notion of Meaning (1949)	<i>152</i>
7	The Entangled Philosophies of Mathematics (1950)	<i>157</i>
8	Meaning (1959)	<i>163</i>
9	The Way the World Is (1986)	<i>166</i>
10	Pressing Extensionality (1992)	<i>172</i>
11	The Innate Foundational Endowments (1996)	<i>176</i>
12	The Growth of Mind and Language (1997)	<i>182</i>

II Previously Published Articles

13	Relations and Reason (1939)	<i>195</i>
14	On the Reasons for Indeterminacy of Translation (1970)	<i>209</i>

15	Methodological Reflections on Current Linguistic Theory (1970)	215
16	On Empirically Equivalent Systems of the World (1975)	228
17	Mind and Verbal Dispositions (1975)	244
18	The Nature of Natural Knowledge (1975)	257
19	Facts of the Matter (1977)	271
20	Cognitive Meaning (1979)	287
21	Grammar, Truth, and Logic (1980)	303
22	Ontology and Ideology Revisited (1983)	315
23	Relativism and Absolutism (1984)	319
24	States of Mind (1985)	323
25	The Sensory Support of Science (1986)	327
26	Panel on Reference (1986)	338
27	Indeterminacy of Translation Again (1987)	341
28	Mind, Brain, and Behavior (1989)	347
29	The Elusiveness of Reference (1990)	352
30	The Phoneme's Long Shadow (1990)	364
31	Three Indeterminacies (1990)	368
32	Preface to <i>The Logic of Sequences</i> (1990)	387
33	Two Dogmas in Retrospect (1991)	390
34	Structure and Nature (1992)	401
35	Commensurability and the Alien Mind (1992)	407
36	In Praise of Observation Sentences (1993)	409
37	Truth (1994)	420
38	Promoting Extensionality (1994)	438
39	Indeterminacy without Tears (1994)	447

40	Assuming Objects (1994)	449
41	Naturalism; Or, Living within One's Means (1995)	461
42	Progress on Two Fronts (1996)	473
43	The Flowering of Thought in Language (1997)	478
44	I, You, and It: An Epistemological Triangle (1999)	485
45	Three Networks: Similarity, Implication, and Membership (2000)	493
46	Confessions of a Confirmed Extensionalist (2001)	498
	Credits	507
	Index	513

Acknowledgments

We gratefully acknowledge the help of the following people in transcribing the material and locating hard to find pieces: Juliet Floyd, Roger Gibson, Paolo Mancosu, Michael Sukale, and Giancarlo Zanet. We thank Lindsay Waters and Phoebe Kosman of the Harvard University Press for their support of the publication of these books, and we thank the copy editors Liz Duvall and Cheryl Lincoln for their meticulous review of the manuscript. Finally, we wish to thank Alexander Quine, the third generation Quine engaged in this effort, for his painstaking work in decoding manuscripts dating back half a century and helping to bring them to the light of day.

Confessions of a Confirmed Extensionalist
and Other Essays

Introduction

Willard Van Orman Quine played a crucial role in shaping philosophy during the twentieth century. Early encyclopedias classified him as a logician, but he soon came to be regarded as a general philosopher—first as a philosopher of logic and language, but eventually as a metaphysician, whose radical thoughts about ontology, epistemology, and communication have repercussions within all major areas of philosophy.

His early work, from his dissertation in 1932 to 1943, was mainly in logic, with the eleven-page article “New Foundations for Mathematical Logic” in *American Mathematical Monthly* in 1937 as the most important contribution. This little article has inspired a large number of further contributions and is even today the subject of intensive research and discussion.

However, from the very beginning, Quine’s work in logic was philosophically motivated, and gradually he focused more and more on philosophical issues. Questions of ontology were the first to make their appearance, in “Ontological Remarks on the Propositional Calculus” (*Mind* 1934), and ontology always remained one of Quine’s key philosophical concerns. He sharpened the ontological issues (“To be is to be the value of a variable”) and discussed ontological commitment. He always had a Spartan bent, and his book *Set Theory and Its Logic* (1963) is basically concerned with keeping the ontology as minimal as possible. Together with Nelson Goodman he explored the possibility of being a nominalist. Unlike Goodman, he settled for a platonic realism. However, in his later work, this realism took an intriguing new turn, not yet fully explored in the secondary literature, towards indeterminacy of reference.

Another main theme that came up early in Quine’s work and grew to

become his main contribution to philosophy started as a skepticism towards meaning and other related notions, such as analyticity and modality. This skepticism grew into a major revamping of previous philosophical views on communication and the relation of language to the world. The first glimmer of this appeared in “Truth by Convention” (1936). From 1943 it found expression in a number of articles directed against modal notions, such as necessity and possibility. In 1951 Quine sketched an alternative view on meaning in “Two Dogmas of Empiricism.” This view was worked out in *Word and Object* (1960), where Quine also clinched his criticism of the modalities by arguing that quantification into modal contexts leads to a collapse of modal distinctions.

Quine’s objections against quantifying into modal contexts can be resolved by giving up traditional semantics in favor of a semantics where so-called “genuine” singular terms, or “rigid designators,” refer to the same object in all possible worlds where that object exists. This basic idea in the so-called “new theory of reference” arose as a response to Quine’s critical objections and owes much to his acute analysis of the underlying problems.

Quine’s main concern—his criticism of the notions of logical necessity and possibility, analyticity, and traditional views on meaning—remains vibrant. In a number of books and articles, Quine developed further the view on meaning that he set forth in *Word and Object*. The central idea is one that Quine shares with most other philosophers and linguists, viz. the public nature of language. Quine’s major contribution is that he took this idea seriously and followed it out with great persistence, to consequences that many philosophers find it difficult to accept. One of these consequences, indeterminacy of translation, is the idea of Quine’s that has been most widely discussed. However, it is only a consequence of more fundamental ideas concerning the public nature of language that Quine refined and partly revised in his later writings.

In particular, Quine, from the very first pages of *Word and Object*, stressed that what we perceive and what we take others to perceive plays a crucial role in language learning and language use. This is a key point in Quine: semantics and epistemology are intimately intertwined. His epistemology is naturalistic: it is contained in natural science, as a chapter of empirical psychology, and yet it is epistemology that provides an account of the evidential bases of natural science, including empirical psychology itself. In the study of meaning and communication, a key problem is to get

insight into what others perceive without imputing to them our own view of the world and our own ontology. In *Word and Object* Quine endeavored to do this in terms of stimulus and response. However, although stimuli and responses are empirically accessible, they are not publicly accessible. The evidence we build on in language learning and language use must be accessible to the members of the community in their daily lives. During the thirty-five years that separated *Word and Object* from Quine's latest book, *From Stimulus to Science* (1995), Quine again and again sought to find a way of dealing with what others perceive without begging the questions of meaning and translation. This enterprise involves the whole range of Quine's philosophical insights: his views on epistemology and ontology and on causality, natural kinds, time, space, and individuation.

Quine has created a new way of looking at these eternal questions of philosophy and their interconnections. He left a transformed philosophical landscape for new generations of philosophers to explore.

Five times in his life Quine selected some of his best essays for collections:

1953: *From a Logical Point of View*

1966: *Ways of Paradox and Other Essays* (2nd edition, 1972)

1966: *Selected Logic Papers* (2nd edition, 1995)

1969: *Ontological Relativity and Other Essays*

1981: *Theories and Things*

Quine remained productive until the end of his life, in 2000. His last essay, which gives this collection its name, was published in 2002. During the last twenty years of his life, he wrote a number of impressive essays where he deepened and, in part, modified his position on many central philosophical issues. This volume collects his main essays from this period and includes also some notable earlier essays that were not included in the earlier collections, although they contain illuminating discussions and are quite often referred to by other philosophers and also by Quine himself in his later writings. The volume also contains several unpublished manuscripts that have been brought to light by Quine's son, Douglas Quine, who has collected Quine's manuscripts and notes and organized a web site, www.wvquine.org, which gives access to this information.

The essays are arranged chronologically. It would be tempting to group them according to topics; for example, meaning and modality, translation, mind and the mental, reference and ontology, naturalism, theory, and evi-

dence. However, while some essays fall neatly into such a grouping, others combine themes from several groups and show how these themes are interconnected. This interconnectedness is characteristic of Quine's holism. The index is constructed in such a way as to make it easy to locate the places where Quine discusses particular topics.

— I —

Previously Unpublished Articles

— 1 —

Nominalism

1946

Introduction, Transcription, Notes, and Bibliography
by Paolo Mancosu

Introduction: The origin of the manuscript
and of the publication project.

Quine's lecture "Nominalism" transcribed below was made available by his son, Dr. Douglas Boynton Quine. The manuscript was listed as one of a series of lectures on Dr. Quine's web site dedicated to his father (www.wvquine.org). I happened to peruse the site at some point in 2006 while doing work on Quine's nominalism. The title of the lecture obviously piqued my curiosity, and I then got in touch with Dr. Quine. Once in possession of a PDF of the original lecture, I quickly convinced myself that the lecture contained the most systematic exposition of Quine's views on nominalism I had ever seen and asked for permission to publish a transcription of the lecture, which Dr. Quine generously granted. The lecture greatly improves our understanding of Quine's engagement with nominalism.

But how did the manuscript of the lecture reach us? In answer to this question, Dr. Quine wrote the following in an email communication dated April 7, 2007:

After my father died, I found in his office a small two drawer (14 inch deep) metal filing cabinet for 3 inch by 5 inch note cards. He typically lectured from verbatim lecture notes since the nature of his work made

carefully formed phrases an essential element of his art. The drawers were filled with his note cards ranging from brief remarks presented at memorials and retirement celebrations to complete lectures (such as this set ["Nominalism"]) and even an entire course (613 note cards): his Harvard University Summer School course on Hume from 1946. I posted details of each of these sets of cards under the heading "unpublished manuscripts and verbatim lectures" at the wvquine.org web site that I had first created for him in 1996. Over the years, a number of scholars have expressed interest in various items. When Prof. James Buickerood expressed interest in the Hume lectures, I scanned them for him and he subsequently transcribed and published them as:

W. V. Quine, "Lectures on David Hume's Philosophy," ed. James G. Buickerood in *Eighteenth-Century Thought* 1 (2003): 171–254.

More recently, you expressed interest in the "Nominalism" lecture so I scanned the cards and provided them to you for transcription and publication. I've also tried to help decode the handwriting when my years of experience with his script and familiarity with his "voice" provided clues despite a lack of deep philosophical training.

I would like to thank Dr. Quine for scanning the lecture, for helping decode some difficult passages, for granting permission to publish, and for his kindness and generosity throughout the process. Meanwhile, Professor Zimmerman had gotten wind of the existence of the lecture and suggested publishing the essay in a special volume dedicated to "Nominalism" with Quine's article as the opening essay. I would like to thank him for providing such an ideal venue for publishing Quine's original essay and for his advice throughout the process. For a detailed historical analysis of Quine's engagement with nominalism, see P. Mancosu, "Quine and Tarski on Nominalism," in Dean W. Zimmerman, ed., *Oxford Studies in Metaphysics*, vol. 4 (Oxford: Oxford University Press, 2008), 22–55.

Editorial Remarks

Quine's text is written on eighty-seven cards (3 inch by 5 inch) numbered C0 to C85, with a card numbered C82a spliced between C82 and C83. I have not transcribed the passages that have strike-throughs for they do not add any important modifications to the content of the text. Articles and

books mentioned by Quine are given a bibliographical entry in bold angle brackets. The entry refers to the bibliography at the end of the transcription. Finally, all notes to the text are mine.

NOMINALISM

Presented at Harvard Philosophical Colloquium,
Emerson B, 3:00 P.M., March 11, 1946

i. Introduction

This isn't going to be a defense of nominalism. Nor is it going to be a refutation of nominalism. Nor is it going to be a positivistic argument that the whole problem of universals—the issue of nominalism—is meaningless because metaphysical.

On this latter point, at least, I'm going to adopt and defend a positive—if non-positivistic—position: that the problem of universals has, or can be given, an important meaning, and that the nominalistic issue is a real one. If, after I have tried to formulate the issue, I seem to be on the fence with regard to it, it will not be because of indifference. It will be, rather, for the same sort of reason that has prevented doctors from producing a specific for colds: difficulties of the problem. Difficulties, in the case of the nominalistic issue, which become clearly visible only after progress has been made in sharpening the issue and analyzing some of the problems that prove subsidiary to it. My purpose in this talk is to report some progress in this regard. But, as a token of good faith, and in compensation for not coming to a conclusion, I'll put my cards on the table now and avow my prejudices: I should like to be able to accept nominalism.

ii. Particulars

What I mean by nominalism—to give it a rough formulation preparatory to further analysis—is the view that there are only particulars. Just this simple doctrine. There are no universals, as opposed to particulars. This doctrine has a good historical claim, I think, to the name of nominalism, but I'm not interested today in pressing that claim. What I am talking about today is, in any case, this doctrine. And clearly it depends on the meaning of 'particular'.

The likely ways of construing particulars are of two sorts, corresponding to the opposition between idealism and realism. This term ‘realism’ is confusing, of course, in that it has two senses both of which are relevant to present discussion: realism as opposed to nominalism and realism as opposed to idealism. I may speak of the two ways of construing particulars simply as the *mental* and the *physical*.

Mental version: particulars are simple experiences: concrete, specific mental events.

Physical version: particulars are concrete spatial objects, or better physical events, i.e., stretches of varying lengths in the histories of physical objects.

So we may have nominalism (rejection of universals) with either an idealism of particulars or a realism of particulars. The general problems raised by nominalism are much the same in either case; but, in order to be as specific as possible, I’ll make an arbitrary choice regarding particulars—choosing the physical. The doctrine to be considered, then, is a nominalism of universals combined with a realism (as opposed to idealism) of particulars.

It is the doctrine that the only things there are, are spatio-temporally extended physical objects. What these objects are like, in detail, it is the business of physics to say; the philosophical content of the doctrine is merely (a) that it *is* the business of physics to say, and (b) that there is nothing other than these things.

According to current physics, these things are made up of quanta of energy, each of which is an approximation to a point-event. We may for convenience regard every aggregate of such quanta as a physical object—a *particular*, in the present sense—no matter how scattered its parts may be, no matter how intermingled with extraneous quanta. But an aggregate in the sense not of a class but of a heap of stones: a total concrete object of which the constituent quanta and all aggregates of them are *parts*—spatial parts, really spatio-temporal parts.

This gives us a lot of things, but, according to current physics, only finitely many. Eddington has computed the total number of quanta in the entire extent and duration of the universe: if we call this number k , then the total number of things in the universe—the total number of particulars, in the adopted sense—can be shown by a familiar mathematical principle to be 2^k .

iii. Nominalism is not extensionalism, nor conceptualism

So nominalism is distinct from the doctrine known in modern logic as extensionalism. The main point of the latter doctrine is rejection of properties, or attributes, in favor of classes. But classes are universals equally with attributes, and nominalism in the defined sense rejects both. The fact that classes are universals is something obscured by calling them “mere aggregates”, and feeling that this likens them to heaps, which are indeed particular. A *heap* of stones *is* a particular, bigger than but no less particular and concrete than any single stone which occurs in it as a part. But the class of stones in the heap cannot be identified with the heap.

For, if it could, then by the same token the class of molecules of stones in the heap could be identified with the heap; and then there would be no distinguishing between the two classes. This would be unacceptable, since we want to say that the one class has some 150 members while the other has some trillions.

Classes are as universal and as abstract as attributes. Indeed their only difference—and one which does not concern us here, is that classes having exactly the same members are the same class, while attributes which are true of exactly the same things are sometimes distinguished.

Equally, nominalism is not the same as the doctrine that universals are mental entities rather than Platonic ideas subsisting independent of man.

This doctrine—conceptualism—has a certain kinship to nominalism, notably in the fact that both doctrines are incompatible with classical mathematics; for it can be proved from the assumptions of classical mathematics not merely that there are universals, namely classes, but that there are classes doomed never to be apprehended by the mind of man. But the nominalist is not pacified by conceptualism. If the attribute of *roundness* is taken to be a mental pattern, or pattern of behavior, or type of mental event, so that different mental events are instances of the *same* roundness, then it is still universal rather than particular, and its subjectivity does not excuse it.

iv. On the struthionism of certain positivists

There is a tendency among *logical positivists* to regard the whole issue over universals, and in particular the doctrine of nominalism, as meaningless.

This tendency is not universal among them; Hans Hahn, a mathematician of the Vienna Circle, came out in favor of nominalism in his pamphlet “Überflüssige Wesenheiten” <Hahn 1930>; though to my regret I can’t find satisfaction in the arguments which he advances.

Other logical positivists such as Carnap have occasionally been misconstrued, notably by Prof. Urban in his *Language and Reality* <Urban 1939> under the head of ‘neo-nominalism’, as embracing not merely nominalism but an absurdly extreme form of it: repudiation of everything, concrete and abstract, except words.

Properly understood, however, the attitude of Carnap and his followers is rather that there is no such issue: that the affirmation or denial of nominalism, along with the rest of what the Vienna Circle refers to rather vaguely as metaphysics, is meaningless.

I object to this idea, however, as *struthionism*—by which I mean that it involves the *struthionic* fallacy: that of burying one’s head in the sand. The derivation is from *στρουθίων*, ostrich.

There is no *a priori* ground for throwing out the statement “There are only particulars” as meaningless; for I have explained in intelligible terms what is to be meant by particulars (there are alternatives, indeed); and the other words of the sentence—‘there are’ and ‘only’—are very decidedly parts of the positivist’s own scientific vocabulary, and that of math. itself. Rejection of the compound would appear to be utterly *ad hoc*.

There is indeed room for clarifying the meaning and implications, still, of the nominalist thesis; and this I favor doing.

v. Use of universal terms no renunciation of nominalism

The most commonplace objections to nominalism clearly reflect the need of clarifying and analyzing the thesis. I suppose the objection most commonly felt is that we *do* have ideas of roundness, of populousness, of independence, etc., and that we entertain these ideas when we understand the words or translate them into another language.

Now in a sense the nominalist would agree that we “have an idea of roundness”, “of independence”, etc., but he would say that this whole phrase is merely a condensed or metaphysical way of saying that we understand the term ‘round’ or a synonym, or that we behave in such-and-such distinguishing ways on being presented with any round object, and so on; and he would claim that this description can be formulated finally in thus

presupposing no universal objects. To maintain the opposite, and that the so-called “understanding of roundness” involves apprehension of a universal object, is—in the absence of some further argument—only a flat assertion, and a *petitio principii*.

The nominalist may protest equally against Russell’s conclusion (*An Inquiry into Meaning and Truth*, 436f) <Russell 1940>; he has just finished arguing that universals like ‘yellow’ can, at least, be eliminated in favor of similarity of particulars. “I conclude, therefore, though with hesitation, that there are universals, and not merely general words. Similarity, at least, will have to be admitted . . . It will be observed that the above argument only proves the necessity of the word ‘similar’, not of the word ‘similarity’. . . . There are occurrences which require for their verbal description sentences of the form ‘*a* is similar to *b*’ . . . The word ‘yellow’ is necessary because there are yellow things; the word ‘similar’ is necessary because there are pairs of similar things. And the similarity of two things is as truly a non-linguistic fact as the yellowness of one thing.”¹

Here the nominalist might object, the alleged irreducibility of the word ‘similar’ is no ground for claiming a corresponding universal object. The nominalist may grant that *the similarity of two things* is a non-linguistic *fact*—by which he may mean merely that the statement to the effect that

1. The full passage in Russell reads as follows:

I conclude, therefore, though with hesitation, that there are universals, and not merely general words. Similarity, at least, will have to be admitted; and in that case it seems hardly worth while to adopt elaborate devices for the exclusion of other universals.

It should be observed that the above argument only proves the necessity of the word “similar,” not of the word “similarity.” Some propositions containing the word “similarity” can be replaced by equivalent propositions containing the word “similar,” while others cannot. Suppose, for example, I say “similarity exists.” If “exists” means what it does when I say “the President of the United States exists,” my statement is non-sense. What I can mean may, to begin with, be expressed in the statement: There are occurrences which require for their verbal description sentences of the form “*a* is similar to *b*.” But this linguistic fact seems to imply a fact about the occurrences described, namely the sort of fact that is asserted when I say “*a* is similar to *b*.” When I say “similarity exists,” it is this fact about the world, not a fact about language, that I mean to assert. The word “yellow” is necessary because there are yellow things; the word “similar” is necessary because there are pairs of similar things. And the similarity of two things is as truly a non-linguistic fact as the yellowness of one thing.

(Russell 1940, 436–37)

two things are similar does not follow from the rules of language. But this is not to say that there is a universal object “similarity”.

vi. The plea of syncategorematicity

The nominalist’s main point, in such a defense, is that we may use words, which are *meaningful* in the sense of figuring in true and false statements, without thereby implying that there are objects—universals, e.g.—which the words are names of. Words can, he urges, be syncategorematic, and this there is hardly an opponent of nominalism who will deny. Words like ‘of’ or ‘which’, etc. Even nouns: e.g., ‘sake’. Who ever heard of a *sake*, even as an abstract object. Do people really have *sakes*? When I speak of my sake, asking you to do something for it, it is only a manner of speaking. And the same may be said of universals generally, according to the nominalist. Even when I use a universal term as if it named an object—saying e.g., ‘I have an *idea of roundness*’—it can still be taken as a mere idiom or metaphysical manner of speaking, no more ontological than the idiom ‘for my sake’. But this is surely *too* easy. It is as though one could talk about anything he likes and yet deny that he is talking about it, simply by pleading that the key words of his discourse are to be viewed as syncategorematic.

The nominalist would have altogether too easy a time of it. He could avail himself of any terminology and any conceptual scheme devised by the most ardent realist, and enjoy whatever conveniences it might have, and yet disavow all consequences having to do with presupposition of unwelcome entities. Clearly all content would vanish from questions of what there is and is not.

vii. To be is to be the value of a variable

But the notion of existence is too fundamental to be dismissed thus easily. The idiom ‘there is’ is so ubiquitous in science and everyday discourse that if we find “What is there?” turning out empty or meaningless we shall do well to suspect our analysis.

It is in fact *not* possible, I maintain, for one to make use of any and every terminology or conceptual scheme that he likes and still remain innocent of ontological commitments. He may declare various of his nouns syncategorematic, and indeed we *can’t* convict him of ontological commitments on the score merely of the nouns he uses; but we can catch him on his *pronouns*.

This may seem a queer distinction to make but I think it is fundamental. What I have in mind here when I speak loosely of pronouns corresponds in the notation of modern logic to the so-called bound variables or apparent variables: the variables of quantification. The 'x' which we use when we say 'there is something x such that'. If someone says 'there is something x such that it is a prime number and exceeds ten million,' no appeal to the syncategorematic status of any of his nouns will absolve him from implying that there are such things as numbers. It is no use to appeal that 'x' itself is syncategorematic: in a sense it *is* syncategorematic, but the point is that the idiom 'there is something x such that' is by its very meaning a flat assertion of existence, insofar as this idiom and the term 'existence' (or 'being') mean anything at all.

I won't dwell further on this now, for it is a view that I have set forth in "Notes on Existence and Necessity" (*Journal of Philosophy*, early '43) <Quine 1943> and in other places there cited. My conclusion is that in order to decide what entities are presumed to exist, by a given discourse, we are to examine the *variables* used in that discourse, and consider what *values* they are presumed to take on. The variable, not the constant name, is the decisive thing. In fact, as I have shown in §27 of *Mathematical Logic* <Quine 1940>, there is an artificial but quite straightforward procedure whereby all use of names, or nouns, can be eliminated from use, so that the whole burden of reference to reality, reference to objects, comes to be carried quite adequately by variables of quantification. To be, then, from the point of view of any given written or spoken doctrine, is to be the value of a variable. The thesis of nominalism becomes this: discourse adequate to the whole of science can be so framed that nothing but particulars need be admitted as values of the variables. Occam's Razor does the rest.

If Russell wants to show that the assumption of a universal object *similarity* is necessary, he must, according to the formulation now arrived at, show that similarity has to be admitted as a value of a variable. This he has not shown.

viii. Commitments of mathematics

Let it not be felt, however, that I have fortified nominalism against Russell and others by redefining it as an unassailable doctrine. The version which I have given to nominalism is, I think, a distillation of the essential content of what has vaguely been intended as nominalism all along. Certainly it does justice to my own nominalistic impulses, and I think more can be

claimed for it than that. And it is *not* an unassailable doctrine; on the contrary, it proves to stand in opposition to the whole force of classical mathematics.

For, it is impossible to get classical mathematics without admitting universals—classes—as values of bound variables.

It is impossible even to express certain principles of elementary arithmetic without admitting certain universals—viz., whole numbers—as values of bound variables.

Elementary logic is compatible with nominalism: truth functional theory and entire theory of quantification. Can also throw in the whole theory of identity.

Portions even of the theory of classes. This seems a contradiction in terms—that part of the theory of classes should be compatible with nominalism—but it is a fact. The point is that portions of the so-called theory of classes can be constructed on the basis of logic of truth functions and quantification and identity, through notational conventions, so that we seem to be talking about classes, but it is only a manner of speaking. This is what I call the *virtual* theory of classes, as opposed to the real. It can be made to include the Boolean algebra of classes and part of the theory of unit classes. But the rest of class theory can't be reconstructed on such a virtual basis; it needs assumptions of universals as genuine entities—values of bound variables.

The situation is similar with relations. Much of the theory of relations can be developed as a virtual theory, in which we seem to talk of relations, but can explain our notation in terms of just the logic of truth-functions, quantification and identity. This portion of relation theory includes the bulk of the relation theory of *Principia Mathematica*, but not all; an example of what cannot be got on this basis is the theory of the ancestral, which some of you know about. This and other parts of relation theory must await a basis where universals are admitted as genuine values of bound variables. Most of this on virtual theory is set forth in my Portuguese book <Quine 1944>, 218–223.

ix. Consequences for nominalism

A formulation of nominalism that I reached earlier was this: Discourse adequate to the whole of science can be so framed that nothing but particulars need be admitted as value of the variables.

Now surely classical mathematics is part of science; and I have said that universals have to be admitted as values of its variables; so it follows that the thesis of nominalism is false. What has the nominalist to say to this?

He need not give up yet; not if he loves his nominalism more than his mathematics. He can make his adjustment by repudiating as philosophically unsound those parts of science which resist his tenets; and his position remains strong so long as he can persuade us that these rejected parts of science are neither intrinsically desirable as ends nor necessary as means to other parts which *are* intrinsically desirable.

One reasonable touchstone of intrinsic desirability of a part of science, as an end rather than merely as a means to other parts, is its efficacy in predicting experience. By this standard, mathematics has no intrinsic desirability; so that the defense of nominalism, so far as its incompatibility with mathematics is concerned, must consist in showing that the portions of mathematics which go by the board are *dispensable* as a *means* to those parts of science which *are* effective in prediction. It must consist in showing how the service of classical mathematics as an auxiliary to the natural sciences could be performed, adequately though more clumsily, by those fragments of mathematics or logic which are still constructible from a nominalistic point of view. The added clumsiness needn't matter; for, once the possibility is established, the nominalist can wink at the use of short-cuts which draw on the less clumsy resources of classical mathematics. He can wink at this because the short-cuts will have been justified as mere abbreviations of procedures which could be expected within what he calls a philosophically sound methodology. This is how he might try to square himself with mathematics; but I don't know that he can do it.

x. Motives for nominalism

Now, why should the nominalist want to go through all this trouble? Why does he prefer particulars to universals in the first place?

I think the motives must be distinguished according as our nominalist adopts a mental or physical version of particulars; though there is a sameness of general spirit regardless. In the mental case (!), his motive may be an extreme sensationalism: what we are presented with are sensory events, and it is unphilosophical to assume entities beyond them, *in particular universals*.

In the physical case, his mentality is likely to be that of Lord Kelvin, who

insisted that he did not understand a process until it was reduced to terms of impact of bodies like billiard balls. The nominalist would out-Kelvin Kelvin in avoiding also the universals which Kelvin no doubt accepted without giving them a thought; for the nominalist feels that these are even more unaccountable than the principles of entelechy or gravitation which Kelvin himself found unsatisfactory.

Modern physics may seem to have cut the ground from under this physical type of nominalist, in abandoning even Kelvin's billiard balls. Tangibility is abandoned as a standard of reality even of particulars. Physical ontology itself takes on a nebulousness which might well discourage the nominalist in his efforts to tidy up reality.

But the nominalist spirit is capable of surviving this. He can defer to the physicist regarding the question what is real by physical standards, and insist nevertheless that there is to be nothing else. The physicist in his verbal expositions of his theory borrows from the general as well as from his own field, and in so doing he may very well seem even to countenance universals; but the nominalist reserves the right to refurbish this conceptual scheme of common-sense, which is not the physicist's responsibility—and to produce a substitute conceptual scheme which, while still theoretically adequate to the physicist's purposes, will not countenance any entities beyond those whose existence it is within the *physicist's* professional competence to assert.

Besides this initial impetus to nominalism, there are *a posteriori* considerations which confirm the nominalist in his prejudices. For one thing there are the logical paradoxes, the simplest of which is Russell's of the class of all classes not members of themselves. These paradoxes are attributable to admitting universals. They do not arise in that part of logic that can be developed in a *virtual* theory of classes and relations, by not taking classes and relations as values of bound variables. And though the paradoxes *can* be avoided by *ad hoc* restrictions, wholly lacking in intuitive justification, they can be avoided *only* thus; and the nominalist says "I told you so".

Another *a posteriori* impetus to the nominalist comes from Gödel's theorem of the incompleteness of elementary arithmetic. Since arithmetic is *a priori*, rather than empirical, truth in arithmetic would seem, by normal epistemological standards, to consist in demonstrability; and accordingly the proof by Gödel that there is no possible consistent system of axioms for elementary arithmetic which will not leave some truths of elementary

arithmetic indemonstrable, is a blow. But Gödel's theorem presumably does not hold for that portion of elementary arithmetic, or reasonable facsimile thereof, which can be achieved within the bounds of nominalism.

xi. Fusions as surrogates for universals

In executing nominalism (executing in the sense of the executive, not the executioner), the constantly recurring problem is of course that of finding particulars such that reference to them can be made to serve various of the purposes for which we ordinarily resort to reference to universals. In the case of many universals, such proxy particulars come readily to hand; namely, those particulars which, in the terminology of Dr. Goodman's thesis **<Goodman 1941>**, are called the *fusions* of the universals in question. The fusion of a class of particulars is the smallest particular which has all the original particulars as parts. If our class has as members the slices of a sliced loaf of bread, the fusion of the class is the sliced loaf itself—which, unlike the class, is a particular. The fusion of the class of all gold objects is the scattered particular which is the spatio-temporal composite of all atoms of gold; it is still a particular, for all its scatteredness; scatteredness is merely a certain complexity of shape. The fusion of the class of stones in a heap, to recur to my earlier example, is the heap itself; and the fusion of the molecules of the heap is again the heap.

Now it happens that the purposes of many universals, notably *gold* and *red*, can be served quite adequately by their fusions. Where the realist (or let us say Platonist, meaning, for the present, realist as opposed to nominalist) would say: 'the *attribute gold* is a metal, and is an *attribute* of the coating of my watch', the nominalist can say: 'the *particular gold* (meaning the described fusion) is a metal and *contains* the coating of my watch as a part.' Where the Platonist would say 'The attribute *red* is a color and is an attribute of Emerson Hall', the nominalist can say 'The particular *red* (meaning the fusion) is a color and contains the coating of Emerson Hall as a part.'

What is more to the point technically, *gold* and *red* in these particular senses—the fusions—can, unlike classes, be admitted as values of bound variables without exceeding nominalism.

But this device enables us to supplant only certain favored universals by particulars. It will not work for the class of stones in the heap, for the reasons I indicated earlier when I argued that classes are universals. And it will

not work for such universals as *man* or *round*. These limitations have been discussed by Dr. Goodman, though not specifically in connection with nominalism, in his thesis. But even where a fusion won't serve outright to supplant a universal, it is often helpful along with other devices in special contexts.

xii. Expressions as surrogates for universals

Another nominalistic surrogate for universals, and a more traditional one, is the sheer word or phrase—the *flatus vocis*. Instead of speaking platonically of *universals* which are *attributes* of this and that particular, the nominalist might speak, it would seem, of the corresponding *expressions*, and speak of them as *denoting* this and that particular; and in general much the same net purposes would be served. This device is, in fact, the etymology of the name 'nominalism'.

We could never hope by this method to get an equivalent of the rich world of universals entertained by classical mathematics; for it can be proved by the principles of classical mathematics that the realm of linguistic expressions even though infinite (supposing expressions can be longer and longer without limit), is yet of a lower order of infinity than the total realm of universals; the realm of classes; even the realm of real numbers. We might, however, hope to get an equivalent of a big enough slice of the world of universals to get by with.

But the predicament to be anticipated here is that names, or words generally, are themselves universals: for they are not particular inscriptions, mounds of ink or neon tubes, but general *forms* of inscriptions. So that words as a surrogate for universals are a reversion to universals. This argument against nominalism was advocated by Uno Saarnio, the Finn, in his *Untersuchungen zur symbolischen Logik* <Saarnio 1935>.

Actually, however, this is one obstacle that nominalism can surmount. For it turns out on investigation that many at least of the important constructions involving reference to expressions, i.e., classes of inscriptions, can be carried out equally well by talking about inscriptions themselves, and bringing them into geometrical similarity with one another. The similarity itself does not (as Russell might suggest) survive as a universal, because all we use is a syncategorematic predicate 'is similar to', without admitting similarity as a value of bound variables; hence without assuming any such *entity* as similarity.

xiii. Number in nominalism

Since classical elementary arithmetic is incompatible with nominalism, an important proving ground for nominalistic constructions is in the applications of number to commonplace situations. Dr. Goodman and I worked together over the example ‘Dogs are more numerous than cats’ at some length. But I can explain the outcome of that example better if I start with a simpler one: ‘there are at least nine planets.’ This can be translated into nominalistic terms as follows: ‘there are objects $r, s, t, u, v, w, x, y,$ and z such that r is not s , nor t , nor u etc., and s is not t , nor u , nor v etc., and so on through the list, and r is a planet, and s is a planet, and so on.’ The idiom ‘is a planet’ here does not imply a universal object, because it can be taken as a syncategorematic constant; and similarly for any other predicate of particulars, so long as it occurs only in predicative position. Universals are invoked only when they come in as values of variables and this does not happen here.

In the same way we can define ‘there are at least ten planets’ or any other fixed number. And then we can define ‘there are exactly nine planets’ to mean ‘there are at least nine planets and there are not at least ten planets’.

This gives us a paradigm for the nominalist translation, in general, of ‘there are at least n so-and-sos’ and ‘there are exactly n so-and-sos’, where in place of ‘ n ’ we have any specified numeral, and in place of ‘so-and-so’ any specified general term denoting particulars.

In short, we are able nominalistically to define *any* specific Arabic numeral *in the context* of application to any specific predicate which applies to particulars; we are able to do thus without implying that there are numbers or other universals.

The case ‘there are more dogs than cats’ can now be handled, curiously enough, *if* we know Eddington’s number k (the number of quanta in all space-time) and hence the number of particulars in the universe (which will be 2^k).

For, we can then explain ‘there are more dogs than cats’ as meaning ‘there are at least one dog and not at least one cat or at least two dogs and not at least two cats or at least three dogs and not at least three cats or [and so on up to 2^k].’

Of course a lower limit than 2^k could be used, so long as we knew it exceeded the number of cats; but the point of 2^k is that it would work for all examples of ‘there are more so-and-sos than such and suches’.

Other related idioms, e.g., ‘there are more than twice as many dogs as cats’, ‘one tenth the Frenchmen live in Paris’, etc., can be handled in ways closely related to this example.

But now we encounter an odd problem, which may metaphorically be called that of the ink shortages.

I haven’t written out the definition of ‘there are more dogs than cats’, I’ve only sketched it. It starts with the clause ‘there is at least one dog and not at least one cat’²; next there is ‘or’ and the clause ‘there are at least two dogs and not at least two cats’; and so on for 2^k clauses. But 2^k is the number of particulars in all space-time, so there can’t be 2^k clauses in the sense of inscriptions; there isn’t room in the world throughout its past and future history. So *is* there a definition of ‘there are more dogs than cats’ after all?

This difficulty can be handled by means of a series of conventions of definitional abbreviations, whereby the whole expression can be compressed to manageable proportions. But it is necessary to recognize that definition takes on a new status in this connection: definitional abbreviation ceases to be a mere convenience, and becomes a theoretical necessity; and moreover the expansions are non-existent and impossible, because of the limits of the universe. We cannot say the expanded expression exists as a *form* without inscribed instances, because we have forsworn universals. If under these circumstances the admission of so-called definitional abbreviations seems a concession to nominalism, then the need of this concession is one of nominalism’s weaknesses.

Going back to the example ‘there are more dogs than cats’ another point that has occurred to all of you is that it is ridiculous to make the statement depend on the number of quanta of energy in space-time. This I grant; and if another translation cannot be found which is free of this dependence, I weep for nominalism.

Certainly all such troubles can be avoided by admitting ‘and so on’ as an irreducible part of the language, and maybe this is the best course; though the same puristic or Spartan attitude which inspires nominalism might lead one to feel that ‘and so on’ is indefensible and meaningless in cases where it is impossible in principle to expand it.

Moving to more abstract matters, it is interesting to see how arithmetic itself fares, as distinct from its applications. We have seen how to define each of the specific numerals, in context. It proves possible also to define,

2. Reading ‘cat’ for the original’s ‘dog’.

for the same contexts, the notations of arithmetical equality, sum, product and power—but only as applied to specific numbers. It turns out that the identities of arithmetic, e.g., ‘ $7^2 + 1 = 5^2 \times 2$ ’, all turn out on expansion of definitions to boil down to self-identities of the form ‘ $x = x$ ’; so that the arithmetic of constants becomes *analytic* in a very extreme sense. How about algebra or arithmetic using numerical *variables*? Something closely resembling it can be got in the *syntax* of the arithmetic of constants; the extent of its formal divergence from classical patterns has yet to be explored, but it is bound to diverge appreciably.

I have had to omit mention of some of the less easily conveyed results that have come of my discussions with Dr. Goodman on these problems, because of limitations of time. We have become somewhat clearer, in ways which it wouldn’t be convenient to set forth here, on the nature of the technical problem of making a go for nominalism. But we can’t claim much; the issue of nominalism is still an open problem, though a clearer problem than it used to be to me.

I feel sure that nominalism can be executed, but I don’t know in which sense.

References

- Goodman, Nelson. 1941. *A Study of Qualities*, Dissertation, Harvard University. Reprinted in 1990 by Garland (New York), as part of its Harvard Dissertations in Philosophy series.
- Hahn, Hans. 1930. “Überflüssige Wesenheiten” (Occams Rasiermesser), *Veröffentlichungen des Vereines Ernst Mach*, Vienna: Verlag Artur Wolf. Reprinted in H. Hahn, *Empirismus, Logik, Mathematik* (Frankfurt: Suhrkamp, 1988), 21–37.
- Quine, W. V. 1940. *Mathematical Logic*. New York: Norton.
- . 1943. “Notes on Existence and Necessity.” *Journal of Philosophy* 40, no. 5: 113–127.
- . 1944. *O Sentido da Nova Logica*. Sao Paulo: Martins.
- Russell, Bertrand. 1940. *An Inquiry into Meaning and Truth*. London: Allen and Unwin.
- Saarnio, Uuno. 1935. *Untersuchungen zur symbolischen Logik. I. Kritik des Nominalismus und Grundlegung der logistischen Zeichentheorie (Symbologie)*, Acta Philosophica Fennica, Fasc. I, 154pp.
- Urban, W. M. 1939. *Language and Reality: The Philosophy of Language and the Principles of Symbolism*. London: Allen and Unwin.

On the Notion of an Analytic Statement

University of Pennsylvania, December 18, 1946

Transcribed and edited by Giancarlo Zanet

Has long been customary in philosophy to distinguish true statements into two kinds, analytic and synthetic. Analytic statements are those whose truth follows simply from the meanings of the words; e.g., “No bachelor is married”. Other true statements are synthetic.

Couched in substantially these words, the distinction has been customary since Kant; but there are plenty of hints of it also in earlier philosophers. Obviously the distinction is important for theory of knowledge. For, analytic statements are the ones which make no assumption, tell us nothing about world; and involve no epistemological problems.

Beginning with Kant, the distinction between analytic and synthetic has come to be associated with one of the basic issue of philosophy, viz., the question whether all synthetic statements depend on experience for their partial or complete verification; or whether on the other hand, some truths can be known independently of experience even though synthetic. Kant held the latter; Hume, and also the logical positivists, the former.

Since the notion¹ of analytic statement (along with its complement, the notion of synthetic statement) has become such a key concept for philosophy, it is remarkable if the notion is not even moderately clear in anybody's mind. But I venture to say that it is not. Today I want to dwell on the ob-

1. In the handwritten text, Quine crossed out the word “copt.” (concept) preferring to it the word “notion”. It's also interesting that the term “notion” is that chosen for the title of the conference. In this way he probably avoids to classify “analytic” as a concept (or something related to the mental), and considers it as something that needs a linguistic definition.

scurities involved, and convey some idea of the ramified character of the problem of defining *analytic statements*.

Prima facie the problem of clarifying the notion of analytic statement is twofold. Analytic statements are those which follow from the meanings of the words. Then what is *follow*? and what are *meanings*?

Let's go back to our example:

No bachelor is married.

Follows from meaning of "bachelor". Meaning of "bachelor" is "man not married". How does the statement *follow*? Logically. What is the logical principle?

No man not married is married.

Put "bachelor" for the phrase of same meaning, viz., "man not married", in this logical principle, and you have your analytic statement.

This suggests a definition of analytic somewhat more explicit than our first one.

A statement is analytic if it can be got by putting one expression for another with the same meaning in a (logical principle, or) statement which is logically true.

So let's defer *sameness of meaning*, and see what *logical truth* is. Example was:

No man not married is married.

Its logical truth consists in its having a form:

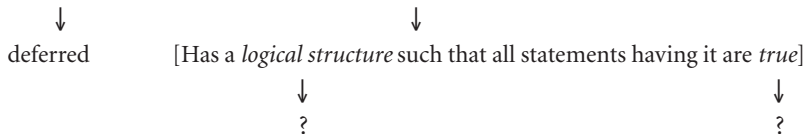
No α not β is β ,

its *logical structure*, such that *all* statements having that structure are true. It is true by virtue of its structure and independently of its content 'man', 'married'. Its structure is (Whitehead) a *validating* form.

Some logical structures are validating forms and some are not. 'No α are β ' is not; for, some statements having this form are false, e. g. 'No lawyers are honest'. The statement 'No stars are inhabited', though true, is not logically true, because its logical structure is not a validating form; there *are* statements of the same logical form which are false.

How far have we come? Analytic:

[Can be got by putting one expression for another of
same meaning in a statement which is *logically true*]

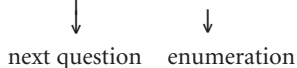


What about logical structure, or logical form? This part is easy. Just two things matter to it:

- (1) the way *logical words* turn up: “no”, “some”, “all”, “is”, “not”, “if”, “and”, “or”, etc.;
- (2) the repetition pattern of the remaining constituents.

So we can redefine logical truth: <card n. 12>

A statement is logically *true* if it is true and remains true under
all uniform changes of the *non-logical* words, (Rid of “logical structure”).



What is truth? Our attempt to get clear on analytic seems now to involve, among other things, getting clear on the general idea of truth. This is often thought of as an unsolved and controversial philosophical topic; thus we hear, e.g., of a pragmatist theory of truth; also we hear disparaging remarks about absolute truth, or truth with a capital T, as if the whole concept of truth were a pretty dubious assumption.

These doubts on the subject of truth have probably developed in part on confusion between truth and belief; or between absolute truth and absolutely certain knowledge. The dictates of the most elementary parts of logic would seem rather to favor absolute truth: any statement you like is either absolutely true or else its denial is absolutely true; we merely may not happen to know which.

In part, far from being a remote and obscure notion, the truth of *any statement* is as clear as the statement itself. For, the statement ‘ $5 > 3$ ’ is true $\equiv 5 > 3$; the statement ‘Philadelphia is on the Delaware’ is true $\equiv$ Philadelphia is on the Delaware. We need have no more hesitation in attributing *truth* absolutely to the statement ‘Philadelphia is on the Delaware’ than in attributing a position on the Delaware to Philadelphia. And the one is, in this particular application, as clear as the other.

We have (as Tarski has emphasized) this quite general characterization of truth:

Any result of putting any one statement in the two blanks of ‘——’ is true
if and only if——

is true.

This is not quite a definition, because circular; but it *can* be shown that any concepts of truth fitting it would be equivalent to each other.

Moreover, I think everyone—whether he thinks he espouses a pragmatist theory of truth, or believes in revealed truth, or anything else—would still accept this paradigm. If he held that ‘The moon is made of green cheese’ was true by his theory of truth, he would hold also that the moon was made of green cheese. Otherwise, his alleged theory of truth simply wouldn’t be a theory of *truth*.

But then these supposedly divergent theories of truth aren’t really different ways of conceiving truth at all; for, as I have said, any two concepts of truth fitting the paradigm are equivalent. The supposed divergence in theories of truth is really a divergence in theories of knowledge, or a divergence in theories of the *world concerning which* true and false statements can be made.

If our paradigm uniquely determines the concept of truth, in the sense that any concepts of truth conforming to it are equivalent to each other, then it would seem—despite its circularity—to be about as good as definition.

And then suddenly we find ourselves trapped in a self-contradiction! Consider the statement:

‘does not yield a true statement when appended to its own quotation’
yields a true statement when appended to its own quotation.

[Show.] Surely ‘true’ is to blame here; everything else is beyond suspicion.

This is an object lesson in circular definition. Our circular definition *isn’t* as good as a genuine definition. Yes, any concepts of truth conforming to the paradigm are equivalent to each other, but the catch is that there aren’t any. A definition determines one concept. Our paradigm determines *at most* one concept, and in part, as the contradiction shows, none. For, the contradiction can be rigorously proved from the paradigm.

The concept of truth, if such there be, fulfills the paradigm; but then the

paradox shows there is no truth. No such concept. Those who scoffed at absolute truth, truth with a capital T, turn out to have been undeservedly right.

This is a blow, because our discourse is badly crippled without the concept of truth; the obstacle raised to defining analytic is really the least of the disaster. Actually, we can get along still with a concept of near-truth, which fits the paradigm except in certain cases of the kind which would lead to paradox.

Roughly the makeshift is this:

Any result of putting any one statement, lacking the word ‘true’, in the two blanks of ‘———’ is true if and only if———

is true.

This is a partial characterization of a concept of near-truth *approximating* to an impossible ideal concept; agreeing with it as applied to statements not themselves speaking of truth.

We can approximate yet more closely to our impossible ideal by introducing a *hierarchy* of truth—concepts, *true*₀, *true*₁, *true*₂, . . . , such that each *true*_i agrees with the impossible *true* as applied to statements not involving *true*_j except for *j* < *i*. The series approaches the completely general but untenable concept of truth as a limit.

But there is more to the truth-problem than this. Thus, let’s go back to the first makeshift, forgetting the hierarchy for the sake of simplicity. The stipulation “lacking the word ‘true’” is really not enough; for we must exclude also all *synonyms* of ‘true’, since these could lead equally to paradox.

So need arises for a notion of *sameness of meaning*—a problem which arose earlier in connection with trying to define analytic, and was deferred. Let’s defer it some more. There are further complications still over this truth problem, but let’s move along. Let’s resume uncritical use of the word ‘true’; imagining that some convention has been agreed on by way of so limiting it as to avoid paradox. For, if we can get a clear definition of analytic relatively even to a provisional concept of truth we shall have clarified our notion of analytic a good deal.

We have², as a stage toward defining analytic, defined logical *truth* in terms of truth. Unfortunately this definition leaves us without a practical *criterion* of logical truth; for, we can’t make all the innumerable changes of

2. See here <Card n.> 12. [p. 26: “So we can redefine logical truth . . . enumeration”]

non-logical components and see if each of the results, one by one, is still true.

Absence of a practical criterion for applying a definition does not necessarily mean, in general, that the definition is not clear. But in the present case it leaves us dissatisfied, because we know that in practice we *can* do better: we *can* establish logical truths, and in doing so we *don't* vary the non-logical components and scrutinize each of the particular results. We see the logical truth of

No man not married is married

directly from its form:

No α not β is β ,

without trying new substitutions at all. And in more complex cases, where we don't see directly, we get our result by deduction from simpler cases; under no circumstances do we resort to varying the non-logical components and trying new examples.

These reflections suggest that we might define logical truth by specifying the rules of deduction by which logical truth is established. This course would have double advantage:

- (1) the definition of logical truth would then correspond directly to our practical method of detecting logical truth, and
- (2) the definition of logical truth would not make use of a general concept of truth.

This new course may or may not be said to be possible, depending on how narrowly or broadly we construe the scope of logic itself. Logic in the broad sense has 3 parts:

1. truth-function theory,
2. quantification theory (all, some, and variables),
3. theory of class-membership.

In this third you can construct relation theory and even arithmetic and mathematics generally. But the third has ontological presupposition from which the other two are free. Viz., it presupposes that, in addition to ordinary concrete individuals, there are abstract entities: classes, or attributes.

This assumption is congenial enough to some philosophers, but not to others; and in any case it marks a philosophically very important bound-

ary within logic. For, the logic below the boundary is independent of any decision on the score of abstract objects; the logic above is not. There isn't time here to show just why this is, but it can be made quite clear. (See "Notes on E. N."³)

For this reason the boundary between first and second and third part of logic might well be viewed as a boundary between logic and mathematics—keeping logic thus neutral as regards ontology. Now it happens that logic in the narrower sense—truth-function theory and quantification theory—admits of a *complete deductive system*, in this sense:

We specify certain logical structures which happen to be of the validating kind—e.g.,

$$\begin{array}{l} \text{If } p \text{ and } q \text{ then } p \\ (y) (\text{If } (x) (\dots x \dots) \text{ then } \dots y \dots) \end{array}$$

and certain rules of inference of the kind that lead from validating forms only to validating forms—e.g., *modus ponens*:

$$\begin{array}{c} \text{If—then } \dots \\ \hline \therefore \dots \end{array}$$

Possible to specify initial structures and rules such that *all* the validating forms whose vocabulary is limited to logic in the narrower sense can be generated from the initial structures by the rules. Due to Gödel (*not* his *celebrated* theorem).

There are infinitely many such validating forms within the vocabulary of logic in the narrower sense but every one is such that it *can*, if we are patient or lucky, be derived by the rules from the initial forms. Therefore the totality of the validating forms becomes definable as the forms which can be generated from such and such ones (enumerate them) by such and such rules. So *logical truth* (narrower sense) becomes definable as the property of *being obtainable by substitution in the blanks of a form which is obtainable by [such-and-such rules] from [such-and-such initial forms]*.

Here is the outline of the definition. To fill it in we have only to refer to any adequate and strictly formulated systematization of logic in the nar-

3. W. V. Quine, "Notes on Existence and Necessity," *Journal of Philosophy* 40, no. 5 (March 4, 1943): 113–127.

rower sense—e.g., chapters I–II of *ML*.⁴ The definition makes no use of “truth”; and it carries with it its rule of practical application.

But it gives us logical truth only in the narrower sense, excluding e.g., arithmetical truth. For this is often said to be reducible to logic, but it is reducible only to logic in the broader sense. The arithmetical vocabulary can be explained as abbreviation of logical vocabulary including ‘ $\in$ ’, but not excluding ‘ $\in$ ’.

It would be natural to expect that logical truth in the broader sense could be defined analogously, simply taking *more* initial forms, including some of the third kind (using ‘ $\in$ ’)—and perhaps adding some more rules.

But Gödel, now in his celebrated theorem (1931), proved otherwise. For, he proved there can be no consistent deductive system whatever in which all validating forms of broader logic—or ever all truths of elementary arithmetic—are deducible.

Arithmetical truth cannot be identified with demonstrability. There must be indemonstrable truths even of elementary arithmetic, no matter what consistent system we devise.

This result was a bombshell, for it had been an article of faith that any arithmetical truth was provable; that arithmetical truth, independent as it is of empirical method, *consisted* in demonstrability. But not so.

Arithmetic cannot be embraced in a proof-system. Therefore, *a fortiori*, logic generally, in the broader sense. Consequently we are frustrated in our desire for a definition of logical truth which will (a) avoid “truth”, and (b) convey with it an indication of a usable method for *establishing* logical truth. We are frustrated in this *if* we want logical truth in the broader sense.

And if we don’t, we are shelving the problem of accounting for the truths of mathematics, in particular arithmetic. Maybe, as hardy nominalists, we are willing to dismiss mathematics; and it is true—though I haven’t time here to argue the point—that mathematics, including arithmetic, involves assumptions uncongenial to the nominalist. But even such a dismissal of higher logic is an uncomfortable position, because certainly arithmetic is something to reckon with. (This is the second time I’ve tried that position and still no response. I think I’ll give it up. Arithmetic is something to reckon with.)

If we do want to define logical truth in the broader sense, we have, so far

4. W. V. Quine, *Mathematical Logic* (New York: Norton), xii + 344.

as I know, to use “truth”. E.g., in the way first suggested <see footnote 2>. Or another way, neater in some respects, is this:

A statement is logically true if obtainable by substitution in a form which is obtainable by [the specified rules] from [specified initial forms] and some true statements built up wholly of logical vocabulary.

But the proof of the equivalence involves difficulties beyond the scope of this talk.

This is enough about logical truth. Logical truth in the narrower sense can be defined in a way which is satisfactory by the most exacting standard, and does not presuppose the general idea of truth. Logical truth in the broader sense is definable with help of “truth”—and even this is some progress, for we are determined surely to sustain the general concept of truth, or a non-contradictory approximation to it, in some sense or other.

So let’s suppose logical truth taken care of now, in one sense or the other, and face the problem of defining *analytic* in terms of it. The question of taking one sense or the other is itself a far from trivial question, for probably the question whether arithmetic is analytic—hotly contested from Kant onward—will depend on whether we define analytic with help of logical truth in the narrower or broader sense. However, we can’t decide everything now.

Our definition of analytic <see footnote 2> in terms of logical truth depends on the idea of sameness of meaning. There has been a tendency among recent philosophers, especially those of the Vienna Circle, to short-cut this reference to sameness of meaning by appealing to *definition*. They say that

No bachelor is married

is, by definition of ‘bachelor’, merely an abbreviation of:

No man not married is married,

and therefore logically true.

But the definitions thus appealed to are a mirage. When do we first find them? Where is the standard dictionary of non-logical terms, which defines ‘bachelor’ as ‘man not married’, and defines each and every other word in one standard way, except for an official set of basic undefined terms?

Clearly there is none; and the *possible* dictionaries to this purpose differ enough among themselves so that many statements which would come out

analytic according to the definitions in one dictionary might not according to another. 'Bachelor' seems, perhaps, to admit of just one natural definition, so that

No bachelor is married

seems analytic by every dictionary; but then so much the worse for the example. How about:

Everything extended is colored?

Is this analytic or synthetic? What *is the* definition of extended? and of colored? And does it turn this statement into a logical truth or not?

Clearly we are lacking a criterion here. The criterion wanted is a criterion of sameness of meaning; and I think we do better to face this question and forget about the fiction of definition. The idea of definitional abbreviation is important in its proper place, but here it only confuses the issue.

Sameness of meaning is the real question, and if afterward we wanted to construct a dictionary of definitions, or constitution system, sameness of meaning would be our guide. Definitions would be so designed as not to disturb relations of sameness of meaning.

What I have said of definition applies equally, I think, to Carnap's more recent appeal to *semantic rules*. A statement is analytic if its truth follows from the semantical rules of the language. How, given even a complete account of linguistic behavior, do we discover what the semantical rules of a language are? and decide, in particular, whether "All colored objects are extended" is analytic? We have to have some criterion of this kind: if users behave thus and so, then this is a semantical rule, a rule depending on meanings; otherwise it is a rule depending on fact rather than meanings, so that the statements whose truth follows from it are synthetic rather than analytic.

The problem of a criterion of semantical rule is substantially the same as the problem of a criterion of sameness of meaning; and I think it is more natural to phrase it in the latter way.

An expression has the same meaning as another expression for a given individual at a given time. Person and time are variables which must be taken into consideration, in seeking our criterion in terms of linguistic behavior; though afterward we may, as a convenient approximation for purpose of generality, abstract from particular person and time and speak of language community.

So the problem is to set up a definition of sameness of meaning for a

given person at a given time. I have no such definition to offer; and I consider the problem a difficult one, yet one which demands solution—one whose solution is constantly presupposed in all our glib discourse about analytic statements, and in other discourses about meaning as well.

A first approximation is perhaps this:

two expressions have same meaning for x at t if substitution of the one for the other, in any statement believed by x at t , yields a statement believed by x at t .

But this proves unsatisfactory. For, sameness of meaning—in the sense needed for defining analytic—must be conceived broadly enough to connect expressions which are not *obviously* alike in meaning to x at t , but can be *shown* to be alike in meaning; e.g.,

‘4³²’ and ‘262,144’

Another complication is the existence of homonyms. In one sense, ‘sound’ has same meaning as ‘auditory stimulus’; in another sense, same meaning as ‘body of water open at two ends’. But we can’t treat ‘sound’ as two words; for, their difference is in meaning and not in form, and meaning is what we are trying to define. We have to say that ‘sound’ is synonymous with each of these two expressions, though they are not synonymous with each other. This, once we get down to business, proves to introduce very serious complications into the problem of defining sameness of meaning.

And even if we could define sameness of meaning by some elaboration and refinement of my first suggested definition in terms of belief, there would still be the problem of defining belief. X believes s at t . This means, no doubt, that x is *disposed* to act in a certain way at t , given proper stimuli; not that he does act. So the whole problem of *disposition terms* is raised—a problem considered by Carnap in “Testability and Meaning”⁵, but still, in my view, unsolved.

Nevertheless, clarity is a matter of degree. If we could get a definition of belief which made this term as clear as various disposition terms like ‘malleable’, ‘soluble’, etc., we’d be much better off than we are now. But belief isn’t in this state yet. And if we could get a definition of sameness of mean-

5. *Philosophy of Science* 3, no. 4 (October 1936): 419–471; vol. 4, no. 1 (January 1937): 1–40.

ing which was clear even relatively to the concept of *belief*, we'd be immeasurably better off in our understanding of sameness of meaning than we are now; but this is a problem which we haven't begun to solve. And the idea of analytic depends, in turn, on the idea of sameness of meaning.

Considering that sameness of meaning is so completely up in the air, we might, alternatively, try first to define analytic independently of this idea, and then get sameness of meaning, for other uses, by definition in terms of analytic. This last step is easy:

two expressions are the same in meaning if, when you put the one for another in a statement s to form a statement s' , the conditional 'If s then s' ' is analytic.

But through what channels, other than sameness of meaning, could we hope to get a criterion of analytic? Perhaps on the basis of relative reluctances in the face of contrary evidence, to discard a statement as false.

An unexpected observation may refute a whole theory, say of physics or biology, but we retain choice as to the particular constituent principle to reject. Analytic ones are the ones which we choose to reject last. E.g., we give up the special laws of Newtonian mechanics rather than the laws of arithmetic, when we find that a statement implied by these combined laws is refuted by Michelson-Morley experiment.

This conception suggests that analytic is a matter of degree. And, if so, then the relation of sameness of meaning defined in terms of analytic is likewise a matter of degree.

The ideas which I have offered in the course of this hour have been mainly negative: the obscurity of our conception of analytic, and the difficulty of doing anything about it. A good many adjacent problems have proved relevant to the discussion, and in these connections also the tenor has been, too often, that no solution is apparent. The ramifications of the central problem have been such, moreover, as to request a pretty formidable segment of the field of philosophy.

But I want to say in closing that my attitude is not one of defeatism, nor one of dismissing the problems as illusory. We have real problems here, meaningful problems worth working on. My feeling is that we philosophers, in connection at least with this complex of problems, haven't been doing very well; and that we should recognize that we haven't been doing very well, but not that we should give up trying.

— 3 —

Lectures on David Hume's Philosophy

1946

Edited by James G. Buickerood

Of the old timers, I feel most congenial to Hume.

—W. V. QUINE¹

Editorial introduction

Quine's name and philosophical views are not infrequently linked with Hume's. It cannot be said, however, that Quine himself discussed Hume's philosophy in any great detail in the works he published in his lifetime. Thus there has been a gap in our view of Quine's understanding and appreciation of an important predecessor of at least one dimension of his philosophical perspective. The following text of Quine's lectures on Hume fills that gap to a degree. This text consists of a revised version of the lecture transcriptions that first appeared in the first two volumes of *Eighteenth-Century Thought*.²

In 1946 Harvard remained on the academic schedule it had followed during the Second World War of three terms of equal duration per year with fall, spring, and summer terms. Before that year and before his leave from the university for naval service, Quine had been asked to teach a course in the history of philosophy, but until his return from the U.S. Navy, had been successful at fending off such requests. For the Summer 1946

1. W. V. Quine, in an interview with the semiannual serial, *Veery* 3 (1993): 21.

2. W. V. Quine, "1946 Lectures on Hume's Philosophy," ed. James G. Buickerood, *Eighteenth-Century Thought* 1 (2003): 172–254 and Buickerood, "Addenda to Quine on Hume," *Eighteenth-Century Thought* 2 (2004): 153–63.

term, he finally relented and chose to teach a course on Hume, Philosophy 14c, which was to meet at 10:00 A.M. on Mondays, Wednesdays, and Fridays. He began preparation during the winter of 1945–1946. Twenty-five students were enrolled in the course, of whom twelve were undergraduates and thirteen were graduate students.³ Quine was at that time, and remained, fundamentally unenthusiastic about teaching the course, as he recorded in *The Time of My Life*:

I tried to make a virtue of the necessity. The critical knowledge of Hume that I would need for my course would mesh with my own philosophical thinking, providing enrichment and perspective. The course, moreover, once given, could be readily given again. Rationalize as I might, however, preparation dragged. I dawdled. It was a struggle to keep ahead. By the end of the course my lecture notes were full and ready for a repeat performance in another year, but I could not bear to offer the course again. Determining what Hume thought and imparting it to students was less appealing than determining the truth and imparting that.

The conclusion of that summer term, concluded Quine, left “the ordeal of Hume forever behind me.”⁴

Still, at the conclusion of the 1946 lectures, Quine did give some thought to the approach he might take were he to teach Hume again in future. “If give course again,” reads an unnumbered card included with the lecture notes, “stress untenability of Hume’s doctrine ‘whatever is different is separable,’ or ‘whole is nothing but sum of its parts,’ used in his argument on cause; how this doctrine would invalidate his own theory of relations (no possible basis for contiguity and succession), and his theory e.g., of time (*Treatise* 1.2.3).”

Description of the manuscript

The bulk of the manuscript of these lectures now in the possession of the Quine estate was discovered by Douglas Boynton Quine soon after W. V. Quine’s death in December 2000. This primary portion consists of five

3. W. V. Quine, *The Time of My Life: An Autobiography* (Cambridge, MA: MIT Press, 1985), 190. Class meeting days and times are listed on the first numbered card in the primary portion of the manuscript. The enrolled students’ degree status was entered on Quine’s class roster.

4. *Ibid.*, 194.

hundred thirty-nine 3×5 index cards, five hundred eight of which Quine numbered in the sequence used to compose the text of the lectures. Thirty-one additional cards in this portion are unnumbered; these cards contain tables and supplementary materials described below.

In the summer of 2002, Douglas Quine and Roger F. Gibson, Jr. discovered an additional three cards and the typed class roster for the course in which the lectures were given in Quine's Emerson Hall office. Two of this portion's cards are in Quine's handwriting; the third is the printed text of the final examination given students in the course. At this time, Dr. Quine and Gibson also discovered an integral part of these lectures: W. V. Quine's annotated copy of Hume's *Treatise*, without which it would be very difficult, indeed impossible, to follow the detailed analyses comprising parts of these lectures.

This copy of the *Treatise* is undoubtedly the one Quine read and annotated in preparation for and during the course of his lectures.⁵ Throughout this text, but especially in book 1, Quine wrote, mostly in pencil (sometimes in red pencil), marginal comments that are reflected in the lectures; constructed a table laying out Hume's classification of impressions referred to in the lectures (and included below, 58); and assigned certain arguments, paragraphs, and passages of Hume's text numbers in distinct series. The latter two sorts of marginalia prove crucial to understanding Quine's lectures, and so are reproduced here in the exact locations Quine read and referred to them during his lectures. The numerals written in the manuscript corresponding to these numbered parts of Hume's text are rendered in this edition in bold type, reproducing the emphasis Quine's hand lent them on his note cards.

The two handwritten cards in the supplementary portion of the manuscript contain lists of texts and secondary sources Quine used or intended to use in the course of his lectures. The list of Hume's texts on one of these cards—*A Treatise of Human Nature; An Enquiry Concerning Human Understanding; Dialogues Concerning Natural Religion*—is repeated with

5. Quine's copy is a 1946 reprint of David Hume, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge (Oxford: Clarendon Press, 1888). This is the edition to which all subsequent *Treatise* page references refer. In addition to this text, Quine penciled a card identifying editions of Hume's works that he used to be the *Dialogues* and first *Enquiry* as found in Edwin A. Burt, ed. *The English Philosophers from Bacon to Mill* (New York: Random House, 1939). The *Dialogues*, as he noted, are also to be found in Charles W. Hendel, Jr., *Hume Selections* (New York: Charles Scribner's Sons, 1927).

more bibliographic specificity on a card at the opening of the series constituting the first portion of the manuscript, as I have noted. The present edition refers to those versions of these texts Quine evidently used in this course as identified here. The remainder of that card is taken up with Quine's reading recommendation for that term's reading period: Church's, Kemp Smith's, and Laird's books on Hume referred to in notes to the text below. The second handwritten card simply lists the titles of these three works in addition to H. H. Price's *Hume on Our Kn. of the Ext. World (sic)*, C. W. Hendel's *Studies*, and T. H. Green's edition of Hume's *Works*.⁶ The printed card in this portion contains the final examination for this course:

PHILOSOPHY 14c

HUME

Final. August, 1946.

Synopsise and discuss Hume's views regarding four of the following six topics. Indicate also the principal observations made by Church, Laird, or Smith, insofar as they relate to any of the chosen topics.

1. Abstract ideas.
2. Space and time.
3. Cause.
4. External objects.
5. Pride and love.
6. The artificial virtues.

Quine's handwriting is on the whole legible. There are places however where the penciling is too faint to make out. In some of these instances, I have been able to reconstruct what Quine wrote with a degree of confidence; in others—amounting to no more than about a dozen words—I have failed. Yet I believe that in no case of failure is the substance of the lectures appreciably diminished. Rather, so far as one can tell, the unreadable material is exclusively parenthetical, consisting, it appears, of one or another kind of self direction, such as a reminder to cover a point in fact covered in the text or a reminder to cover something in fact presented subsequently in the lectures.

The primary portion of the manuscript includes, in addition to the text

6. H. H. Price, *Hume's Theory of the External World* (Oxford: Clarendon Press, 1940); Charles W. Hendel, Jr., *Studies in the Philosophy of David Hume* (Princeton: Princeton University Press, 1925). The Green and Grose edition of Hume that Quine used is identified in n. 40 below.

proper, preparatory and supplementary notes pertinent to Quine's study of Hume and to his presentation of that material to his students.

One card at the beginning of this portion bears the title "W. V. Quine's 1946 Lectures on David Hume's Philosophy." The succeeding card lists the days and times of class meetings and the principal readings for the lectures as identified above, though with some more bibliographical specificity that will be reflected in editorial notes and interpolated quotations below. Twenty-three cards are densely covered with tables correlating pages of Hume's *Treatise* and references to that work in R. W. Church's *Hume's Theory of the Understanding* (1935), J. Laird's *Hume's Philosophy of Human Nature* (1932), and N. Kemp Smith's *Philosophy of David Hume* (1941), books to which Quine referred in the course of the term. Page references from the "*Treatise* to Church, Laird, Smith" cover three of these cards; page references in "Church to *Treatise*" cover three more. The tables detailing page references in "Laird to *Treatise*" take up ten cards; and seven cards are devoted to references in "Smith to *Treatise*."

The contents of the remaining unnumbered cards in this portion are:

- (1) A note on what Quine wished to focus should he repeat the course, quoted in full above (2).
- (2) "Order of battle, 5th–9th weeks.
Fifth week: Physical objects; identity.
Minds.
Mop up Book 1.
Sixth week: First lecture: last two sections of *Enquiry*."

Nothing further is entered here; Quine left this projected plan uncompleted.

- (3) After mentioning Selby-Bigge's editions of Hume's *Treatise* and "*Inquiry*," Quine wrote that he intended to "End up course with lectures explaining Kant's attempted answer to Hume and its shortcomings."

The remaining three cards contain abbreviated citations of, or allusions to, philosophical works pertinent to his interest in Hume that Quine consulted himself or recommended to students. The first three of these works are not mentioned expressly in the lectures. Identified and completed, these citations are:

- (4) J. N. Findlay, "Morality by Convention," *Mind* 54 (1944): 142–69.
- (5) "Hume, Causality: See" G. H. von Wright, *The Logical Problem of Induction* (Helsinki: Editio Societas Philosophica, distribuit Akateeminen Kirjakauppa, 1941). C. D. Broad, "Hr. von Wright on the Logic of Induction (II)," *Mind* 53 (1944): 97–119.
- (6) "Constitution system: Locke: vol. I, p. 279; vol. II, p. 28." That is, Locke, *Essay* 2.12.1, and 3.4.16, on the names of simple ideas and the constitution of complex by simple ideas.⁷

A final cache of lecture-related materials, here called the auxiliary portion, was found by Dr. Quine in February 2003 in a folder labeled "Current Assignments." As this material supplements rather than constitutes any of the lecture manuscripts, three apparently preparatory elements of this portion will be described in this introduction, while revisions and supplements to the lectures as delivered in 1946 will appear in editorial footnotes in suitable locations. This auxiliary portion consists of fourteen note cards and sheets of paper that can be usefully divided into seven sets:

- (1) A typed half-sheet containing a sketch of the texts and subjects to be covered in the course, Philosophy 14c.
- (2) A small note sheet, bearing an evaluation of some of the recommended secondary sources.
- (3) One note card with bibliographical information on the six secondary sources Quine recommended to students. A second card bears a reference to a secondary source criticism of Hume.
- (4) Three note cards contain the topics Quine assigned for the two papers he required of students attending the course.
- (5) Five unnumbered note cards containing an alternative to the primary portion of the manuscript treatment of Hume's *Treatise* 1.2.4–6.
- (6) A small sheet of notepaper on which Quine noted the citation of one piece of Hume correspondence.
- (7) A second copy of the final examination, transcribed above.

Why is this apparently disparate material combined in a folder separate from the primary portion of the lectures manuscript? Only two (nos. (4) and (7)) of these seven pieces can be construed as assignments. Yet, as

7. The edition of Locke that Quine used is identified in n. 18 below.

192

To I. II, parim
 (R. indirect definition) (Carnap) (Russell) (Quine)
 Key: contextual def'n. Cf. Russell, *phil*
 desir'ns.
 Appealing to operational criteria.
 Such is modern empiricism.
~~Some including~~ *apriori* ~~7+~~ *challenges*
 indirect "correction" of size comparisons. *
 Carnap, *Aufbau*. (But presupposes
 a logic which presupposes
 universals.)
 * Cf. also Bkly., *New Thy. of Vision*.

Figure 3.1a. The card in the primary portion of the manuscript on which Quine wrote the opening of his 8 July lecture on *Treatise* 1.2, the transcription of which appears on p. 83.

three other pieces (nos. (1), (2) and (3)) are nearly certainly preparatory notes, and as the two remaining pieces are most plausibly construed to contain material that revises (no. (5) figures 3.1a and 3.1b) and supplements (no. (6)) parts of the lectures as contained in the primary portion of the manuscript, I think it most likely that Quine preserved this material together in case he was required to repeat the course.

The first piece (above, no. (1)) is a typed note, bearing some penciled notations in Quine's hand. The sheet was prepared possibly as a draft syllabus for Philosophy 14c. The year, "1941," is typed at the top near center of this page, which is roughly one-third of a standard 8.5 x 11 sheet of paper. The physical character of the typescript⁸ and the name "Williams" penciled in the upper-right corner of the page suggest that Quine obtained it from Donald C. Williams, who had offered that course in that year. The evidence

8. The note was clearly not typed on Quine's own Remington typewriter, which notoriously left a floating 'w' on typed sheets. It is likely that the note originated with Donald Williams.

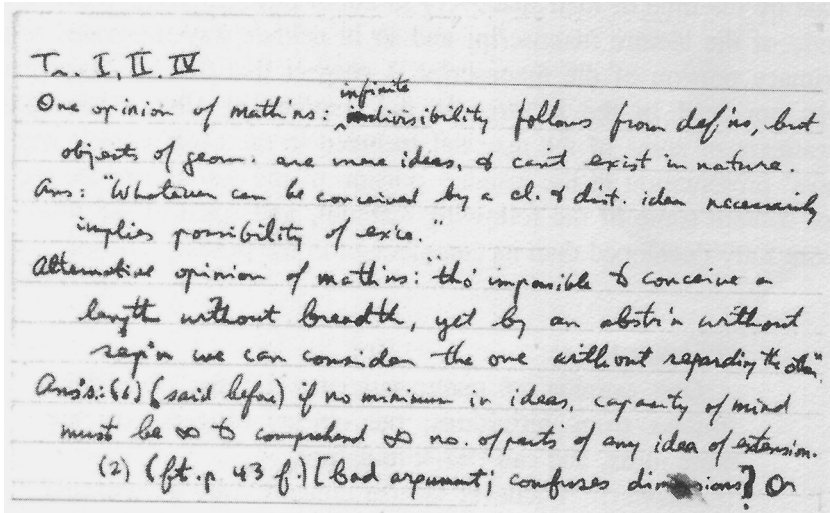


Figure 3.1b. The card in the auxiliary portion of the manuscript (no. 5) on which Quine wrote the opening of his revised lecture on *Treatise* 1.2.4.

therefore suggests that Quine used Williams's notes as a guide to contrive his own 1946 version of the course (Figure 3.2).⁹

Philosophy 14c. 1941 Williams

Treatise of Human Nature

Book I. Of the Understanding	Theory of Knowledge
Book II. Of the Passions	Human nature, morals and politics
Book III. Of Morals	

Enquiry Concerning Human Understanding

Section X. Of Miracles [with Book 1 of *Treatise*]

Dialogues Concerning Natural Religion

9. Donald Carey Williams (1899–1983), author of *Principles of Empirical Realism* (Springfield, IL: C. C. Thomas, 1966), was a member of the Harvard philosophy department for twenty-eight years and was department chairman from 1947 to 1954. On his death, Firth, Nozick, and Quine claimed that, during Williams's tenure there, the Harvard department "was strong in the history of philosophy, thanks largely to his courses on the British empiricists and the philosophers of the seventeenth-century [*sic*]." Roderick Firth, Robert Nozick, and W. V. Quine, "Donald Carey Williams 1899–1983," *Proceedings and Addresses of the American Philosophical Association* 57 (1983): 245–48 on 247.

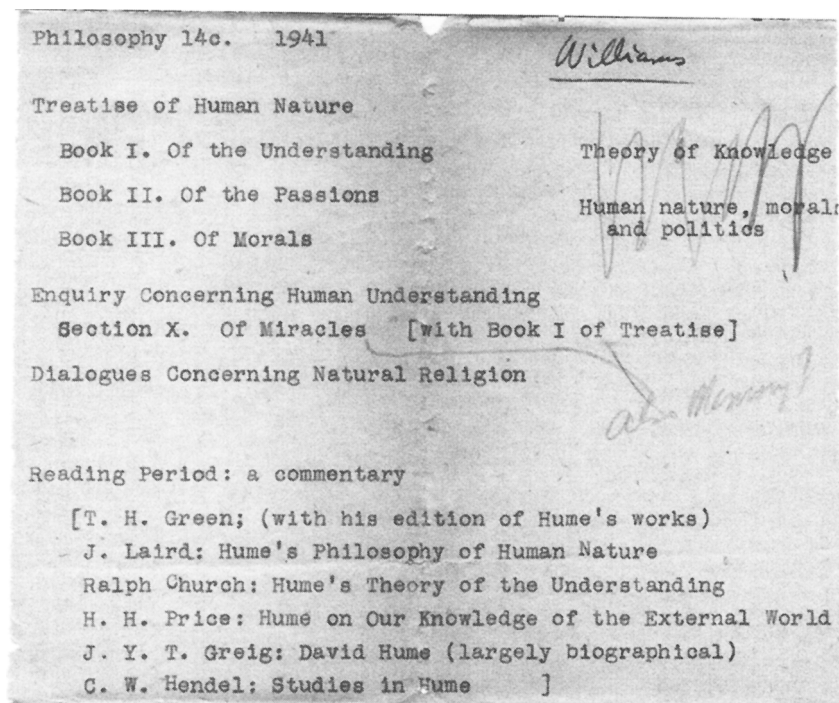


Figure 3.2. Donald C. Williams' syllabus sketch on which Quine penciled notes in preparation for his 1946 version of Philosophy 14c.

Reading Period: a commentary

- [T. H. Green; (with his edition of Hume's works)
- J. Laird: Hume's Philosophy of Human Nature
- Ralph Church: Hume's Theory of the Understanding
- H. H. Price: Hume on Our Knowledge of the External World
- J. Y. T. Greig: David Hume (largely biographical)
- C. W. Hendel: Studies in Hume]

These secondary and further primary sources are those that Quine recommended his students examine.

The following notes (as above, no. (2)) are handwritten on a small sheet of note paper, presumably drafted as Quine was deciding on the content and structure of the course:

Price 'Hume on our Knowledge of the External World' probably best.

Hendel good biographically.

R. W. Church

Probably vols. 1–2 *Treatise* and the *Dialogues*.

Maybe *Inquiry*. Ch^{ec} on Miracles and Memory.

The final element of the auxiliary portion to be discussed here is a set of two cards (above, no. (3)), one simply a handwritten note listing the six works cited above, and the other bearing but one line in Quine's hand: a reference to Ralph Church's review of the by-then standard criticism of the principle of ideas as copies of impressions in Hume's analysis of perception:

Criticism of Hume's "1st principle" (1.1.1).

Church, p. 27⁺¹¹

Editorial principles

This edition of Quine's lectures on Hume attempts to keep editorial intrusion to a minimum. In addition to the real merits of this principle in general, I have judged it particularly suitable in this case, inasmuch as the potential readership of this text falls into at least four overlapping, yet certainly distinct, classes: specialists in Quine's thought, (increasingly) students of the history of analytic philosophy, students of Hume, and philosophers and scholars focused mainly on the issues addressed here by both

10. This reconstruction is simply a best guess. Quine's hand is not entirely clear here. While the "Ch" is perfectly discernible, the final letter could possibly be "h" or "k." It is in any case clearly an abbreviated word followed by a period. The most pertinent speculation for its full form based on the reading of "h" is arguably "Church," indicating Ralph W. Church's *Hume's Theory of the Understanding* (London: George Allen and Unwin, 1935), but there is no significant discussion of miracles in that work at all. Given Quine's association of memory with miracles with respect to the first *Enquiry* in the above-described auxiliary text, it seems more likely that this scribbled note is a reminder to check on these topics in that work. Miracles are the subject of section 10; memory is referred to in numerous places in the *Enquiry*, though it receives no sustained analysis there.

11. Church, *Hume's Theory of the Understanding*, 19–39, esp. 26–28. The subscripted plus sign is Quine's characteristic indication of the page following that cited; this citation then means: Church, 27–28.

writers. The respective interests and backgrounds of members of these classes may very well differ from one another in ways that are likely to mitigate against the ready usefulness of any elaborately introduced and robustly annotated text.¹²

The following constitutes the entirety of the lecture notes written by Quine for the course he taught on Hume's philosophy in the summer of 1946 with some few exceptions and qualifications. Occasionally Quine evidently forgot that he had covered a point in a previous lecture and so repeated that point on a later date (e.g., on 61, below). Such repetitions and indications of them have been omitted. Neither can I see any useful purpose in including the twenty-three cards of cross references of pages in Hume's *Treatise* with pages in Church's, Laird's, and Kemp Smith's works on Hume. All other unnumbered supplementary cards of the primary manuscript portion have been omitted from the text proper, though the complete contents of each of them have been included or at least described in the introduction or footnotes to the text with sufficient indication of the source. Finally, the lectures, in places, contain notes set off by asterisks, some of which are clearly reminders to Quine himself to cover or to repeat some substantive point, take attendance, or announce some course-specific information such as student paper submission dates, or indications that he had done so. With some obvious exceptions, signaled by parenthetical asides integrated within the text absenting Quine's asterisks, these sorts of Quine's notes have been omitted, as have all crossings out.

This edition makes no attempt to disguise or otherwise mitigate the character of the following text as lectures, but I have decided some changes from the handwritten notes to be advantageous for publication. While Quine's handwriting is for the most part clear, its legibility does vary through the course of the manuscript, in places becoming rather difficult to make out. Doubtless, despite all efforts, some deciphering will prove erroneous. Again, though I have as a rule worked to keep editorial intrusion and overt textual interpretation at a minimum, a number of types of intervention seemed desirable in the interest of clarity and readability.

12. Michael Pakaluk's transcriptions of selected passages of these lectures is the only other publication that discusses or presents any of this material in any detail. The principles governing Pakaluk's transcriptions of those parts of these lectures he chose to publish differ considerably from the principles used in this edition, and his transcription includes nothing from this auxiliary portion of Quine's manuscript. See Pakaluk, "Quine's 1946 Lectures on Hume," *Journal of the History of Philosophy* 27 (1989): 445–59.

All editorial interpolations—additions, some footnotes, and some sorts of corrections—in the text and in the footnotes are included within angle brackets.

Original spelling has been retained, with very few exceptions. Straight-forward spelling errors, few as they are, have been silently corrected. Other spelling irregularities have been decided in favor of Quine's clear preference in his published work, where I have identified these.

I have not made sentence fragments and other irregular sentences conform to accepted grammatical standards in those cases where I judged their meaning sufficiently clear or where the lack of clarity is intrinsically interesting. Common, informal verbal contractions such as 'isn't,' 'can't,' and 'doesn't' have been left to stand as originally written.

Quine used or mentioned a few foreign words; these words are italicized. The meaning of Quine's very occasional colloquial German abbreviations has been silently rendered in English.

Unconventional contractions and verbal, numerical, and other symbolic abbreviations have been expanded into English. Verbal abbreviations are very numerous and not altogether consistent or systematic, and so are subject to interpretation. Among very many others, I have interpreted 'hp' as 'hypothesis,' 'H' as 'Hume,' 'a/c' as 'account,' 'dipl' as 'diplomatic,' 'thy' as 'theory,' 'phps' as 'perhaps,' 'diff't' as 'different,' 'q'n' as 'question,' 'thot' as 'thought,' 'psg' as 'passage,' 'vble' as 'variable,' 'prpns' as 'perceptions,' 'exx' as 'examples,' 'cspdg' as 'corresponding,' 'expce' as 'experience.'

Symbolic abbreviations and the shorthand use of symbols pose other interpretative challenges. Some symbolic expressions are strictly formal and have thus been rendered here as written. Other uses of mathematical and logical symbols pretty clearly constitute a species of shorthand Quine used, not intended to be interpreted strictly, and so are not reproduced here. Numerals in English sentences, as for instance, in 'more than 1 kind,' have been rendered here in English; in this case, as 'more than one kind.'

The equal sign ('=') has sometimes been interpreted as 'equals'; in other places as 'as,' or 'that is,' depending upon the immediate context of its occurrence and readability while retaining Quine's meaning. In other cases, such as its parenthetical appearance clearly indicating synonymy of the word following the sign with the word immediately antecedent to the parenthesis, the sign has simply been deleted, as the parenthetical construction clearly enough implies some degree of synonymy or equivalence of terms.

The plus sign ('+') following page numbers in a text quoted or referred to has been rendered as 'f.' or 'ff.' where it has indicated pages of a text referred to or quoted in the manuscript. In a few instances the plus sign is used to signify something of the order of 'and successors,' as in 'Kant ++'; such uses have been expanded in just this way.

Though each new paragraph is Quine's own, I have not retained all of his paragraph breaks, for a great many of them were clearly created merely as visual aids. In some parts of the manuscript, each sentence is indented as if it were a distinct paragraph. To reproduce this feature would work against the intelligibility and readability of the text.

Quotations from, and references to, identifiable sources (Quine often provided incomplete but sufficient indication of their source) have been identified, where necessary, but have in no case been altered from the manuscript version in the body of this text, even if inaccurate. Notes have been added with the correct citation in those cases where Quine's information is clearly erroneous, and with full bibliographical information in all cases possible.

Some quotations from Hume and from Locke are not written out in the manuscript but are clearly indicated there as having been read during the lectures. Quine tended to adopt such shortcuts more frequently as the term went on. These quotations have been added on the evidence provided by Quine's annotated copy of the *Treatise* and from a copy of the edition of Locke that it is clear Quine used. These additions are included, as per the above editorial stipulation, within angle brackets. It has rarely been possible to be certain as to precisely how much of a passage Quine quoted or intended to read in such cases; these sorts of notes to himself read something like: 'Read p. 608, foot,' indicating that the passage to be read lay on what I estimate to be the final quarter of the identified page. Yet it is clear enough from those (usually early in the course) passages that he did trouble to write out in his notes, that while the cited page number and, e.g., 'foot' correctly identifies the location of the beginning of the passage quoted, he often continued his quotation well within the succeeding page. Thus in an attempt to achieve completeness without pretense to utter fidelity to an event that cannot now be recaptured, I have perhaps quoted rather more of the cited source than Quine actually read in 1946.

Again on the basis of the principle of keeping editorial intrusion to a minimum, I have supplied notes merely with bibliographic and other immediately pertinent information to texts Quine mentioned in his lectures. If Quine referred to or quoted a text, the exact location of the passage and

the edition from which it is quoted is provided if possible or practicable. If Quine failed to identify a textual source for a position to which he alludes, I have attempted to supply this; in some cases I have failed. There are quite a number of names mentioned in the text—some more familiar to readers than others—for which I have supplied little information. To do any more would make this text unwieldy and inconsistent with what I take the spirit of the lectures to have been. Such omissions, in any event, will in many cases be readily enough rectified by readers with some acquaintance with early modern and classic analytic philosophy.

All book titles have been regularized in italic form. Quine usually accurately underlined them in his manuscript, but in some cases, such as the titles of Plato's and Aristotle's works, failed to do so. Minor punctuation irregularities in citations of texts, especially of Locke's *Essay* and Hume's *Treatise*, have been silently rendered consistent.

Quine's forms of citation of Locke's *Essay* and Hume's *Treatise* have been changed from the older conventions of, e.g., 'Bk. II, Ch. VII, §3,' and 'Bk. I, Pt. II, §2,' respectively, to the less cumbersome Arabic numeral system now common, reading '2.7.3,' and '1.2.2.'

All manuscript underlining for emphasis and titles has been represented with the use of italics. In cases where Quine carefully indicated his (rather than his subject text's) emphasis within a quotation by using a series of subscripted points rather than lines, the emphasis has been signaled by underlining within the quoted text (e.g., below, 62).

Quine included some asterisked footnotes in his manuscript. These are here numbered in the same sequence as editorial footnotes, but Quine's own notes are not contained within angle brackets.

Now and again Quine wrote what to the reader can seem enigmatic directions to himself to explain or illustrate a point or thesis, presumably extemporaneously, since no part of this manuscript contains such expository material. These directions are included within the body of the text where they originally appear. For the majority of these, I have attempted no reconstruction of such illustrations or explanations, even where the tack Quine is likely to have taken seems clear to me.¹³

13. Thanks to Douglas Boynton Quine for permission to edit these lectures and for his aid in deciphering W. V. Quine's handwriting and otherwise interpreting the text. I am grateful also to Roger F. Gibson, Jr. and Ernest Lepore. Finally, my grateful thanks go to Patricia Alexander Buickerood as always for her technical and other sorts of assistance with this project. Responsibility for errors is solely my own.

W. V. QUINE'S 1946 LECTURES ON
DAVID HUME'S PHILOSOPHY

David Hume was a Scotchman. Born Edinburgh 1711; died 1776. The house of Hume, also spelled Home, belonged to the Scotch nobility. Moderate means. David Hume studied at Edinburgh and then returned to the family country-seat in Berwickshire to study and meditate. At the age of twenty three he worked for a few months in a business office in Bristol; then gave it up and settled down in France to write his *Treatise*, which came out in 1739 (when Hume was twenty eight). By this time he had returned to his father's house and was continuing his philosophical writings; *Essays* came out in 1741. By this time he had begun to aspire to a chair at Edinburgh, but was disappointed. In 1746 and again 1748 he traveled in Europe as a diplomatic secretary. This was period also of *Enquiry Concerning Human Understanding*. Then returned to Scotland and continued his writing. Besides *Inquiry Concerning Principles of Morals* and *Dialogues Concerning Natural Religion*, his writing now took a political turn: *Political Discourses* (1751). And from 1753 on, *History of England*, finished 1761. But philosophy wasn't neglected; *Natural History of Religion*, "Of the Passions," "Of Tragedy," "Of the Standard of Taste" (1757). 1763–1769, posts in Paris and London as diplomatic secretary again. Then retired to Edinburgh on his pension and royalties.

So Hume's career as thinker and writer occupies three fields: philosophy, political economy, history. Hume was of some importance as an historian. His history was a considerable success. And even those who criticize it as tendentious are ready to admit that it was distinguished for its time and an important milestone in historiography. But Hume's greatness is in philosophy. And, within philosophy, his greatness is in theory of knowledge. Hume's philosophical writings treat of ethics and religion as well as theory of knowledge, but the theory of knowledge is the keystone of the structure.

Historically the theory of knowledge begins as a quest for certainty. Mathematics is the shining example of a science in which we can be certain of our conclusions. If this virtue were shared by the other sciences, notably physics and the rest of the natural sciences, there would have been little speculation over the theory of knowledge. Philosophers worry about theory of knowledge because the human understanding is in *trouble*; and its trouble is the uncertainty of knowledge about the world. Investigators in physics and other empirical sciences are aware that they are engaged in

framing hypotheses which can never be fully proved, can never achieve anything beyond a respectable degree of probability short of certainty. Physics traffics in opinions.

There is a degree of uniformity in the *opinions* of all concerned, in the matter of physics, which may well be envied by the philosophers who, each in his peculiar way, are out for certainty. It happens that we can often be more *certain* that something is highly *probable* than that something is certain. But this circumstance does not do away with the philosophical urge to find a bed-rock of certainty somewhere beneath the probabilities of natural science. And the motives for this quest are two-fold.

- (1) Uncertainty is an *unsatisfactory* end-point. So if we can't aspire to certainty in natural science itself, we want to penetrate back to underlying principles of natural science, broader and more fundamental, for which certainty *can* be claimed.
- (2) Further, it is argued that if there are not some such underlying certain principles, the so-called merely probable laws of science cease even to be probable; for, we can't say for *certain* that they are probable; they become merely probably probable; and it is not certain that they are probably probable, so they are merely probably probably probable; until finally every probability is modified by an infinite series of probabilities, which amounts to not being probable at all.

Briefly: there can be no probability without some certainty on which to base the probabilities.

I should say that I can't quite subscribe to this reasoning, for it overlooks the possibility of what in mathematics is called a convergent product. Thus, suppose the probability of a particular law of physics is $\frac{3}{4}$. I.e., there are three chances out of four that the law is true. (We never know probabilities of laws numerically this way, but let's go on for the sake of argument.) Now suppose our law is only probably probable, but that this second-order probability is $\frac{15}{16}$. (I.e., there are fifteen chances out of sixteen that the probability of the law was $\frac{3}{4}$.) And suppose, it is only probably probably probable, but that the third-order probability is $\frac{35}{36}$. And that the fourth is $\frac{63}{64}$, and the fifth is $\frac{99}{100}$; and, in general, the n th is $\frac{4n^2-1}{4n^2}$. Then the net probability is not vitiated by all these qualifications to complete indifference or toss-up ($\frac{1}{2}$), as might be expected. It reduces rather to

$$\lim_{n \rightarrow \infty} \pi \frac{4n^2 - 1}{4n^2} = \frac{2}{\pi} \text{ about } \frac{7}{11},$$

which isn't much worse than the original $\frac{3}{4}$.

But there isn't much comfort in this alternative, for it would be pretty hard to persuade ourselves that it held. Especially that it *certainly* held. So there remains sufficient motive for seeking certainty in the underpinnings of natural science: not merely in order to have something better than the probability of natural science, but in order even to make science safe for probability itself.

This quest for certainty goes back to Plato, and nobody knows how far beyond. The probability problem that we have just now been considering cannot, of course, be read into Plato. Nevertheless he regarded the science of nature as of an *inferior grade of knowledge*, and he was in search of basic principles, to which certainty might attach, *behind* the science of nature. Mathematics, then as now, was the prototype of certainty. Proofs in geometry and theory of number were accomplished by analyzing the concepts and finding basic relationships which, clearly and obviously and with full certainty, held among them. There were the *axioms*, self-evident truths—evident as soon as one clearly understands the concepts involved. And the other truths of geometry or theory of number followed from such axioms by successions of obvious steps of necessary inference. If we could only get a comparably clear understanding of other concepts, outside mathematics, we might hope for equally indubitable knowledge regarding them; mathematics merely happened to be that part of the job that was already pretty well in hand.

This analysis of concepts was *dialectic*, conducted typically through the Socratic procedure of cross-questioning. Viewed by Plato primarily as an analysis of *forms*; hence the word *Idea*, which originally meant *form* or *shape*, reminding us that geometry was the prototype of necessary reasoning. Plato was under no illusion that dialectic offered a royal road to factual knowledge about the world. Dialectic was an extension of mathematical certainty beyond the bounds of mathematics, but it was no substitute for the observation and guesswork of natural science. Physics remained a lower-grade accomplishment, incapable of being elevated to the level of genuine knowledge. But the ultimate realities *underlying* physical phenomena (viz., the *ideas*, for Plato) could be known with all the certainty of mathematics itself. Philosophy was a generalization of mathematics.

Hopefulness over the prospects of gaining fundamental certainties about reality had its ups and downs. Before Plato there were the Sophists, and after Aristotle there were the Skeptics; both groups represented a low in hopefulness, and both groups pointed to the inconsistencies of philosophers—inconsistencies between philosophies and inconsistencies within philosophies—to justify their defeatism. But hope springs eternal. Or *perennial* would be more accurate; it keeps coming back. Aristotle shared with Plato the hopeful view that the basic truths of objective reality could be known with certainty, through an intellectual apprehension of essences. And this was the tradition that dominated the Middle Ages. I'm not prepared to formulate this as a sharp issue; I think of it as a matter of degree: a very great deal about God and reality and the general nature of things was believed to be accessible to the human mind, with complete certainty, through the simple expedient of taking thought.

And thus we come down to the seventeenth century and Descartes, who, like the Sophists and the Skeptics of antiquity, was moved to attack the methodological complacency of the metaphysicians. Like the Sophists and Skeptics, Descartes was disturbed by the fact that the sweeping certainties of the metaphysicians were neither as convincing, taken separately, nor as compatible, taken together, as might reasonably be required. But Descartes was a mild man, and a man with a will to believe. He did not emulate the Sophists and Skeptics by abandoning himself and metaphysics to a nihilism of doubt and defeatism. His purpose was quite the opposite: to buttress metaphysics by performing a more conscientious and painstaking dialectic than ever had been known before. He hoped in this way to arrive at certainty in so genuine a sense that there could no longer be disagreements and inconsistencies in metaphysics, nor any future justification for the counsels of skepticism and despair. It is one of the great ironies of the history of thought¹⁴ that this new focusing of attention on method, for which Descartes's motive was the strengthening of metaphysical certainty, should have brought on—step by step through Locke and Berkeley to Hume—a golden age of metaphysical skepticism.

What I have said of Plato, that his ideal of certainty in philosophy was

14. Others: Berkeley's anti-hyloism for spiritual supremacy; Malebranche, Cordemoy, etc. against causality for glory of God. <See, e.g., George Berkeley (1685–1753), *A Treatise Concerning the Principles of Human Knowledge* (1710); Nicolas Malebranche (1638–1715), *Recherche de la vérité* (1674); and Géraud de Cordemoy (1626–1684), *Le Discernement de l'âme et du corps* (1666).>

inspired by mathematics, is yet more clearly true of Descartes. Descartes was a great mathematician, and the founder of analytical geometry. He was familiar with the way in which the most abstruse and surprising mathematical relationships can be broken down into a series of short steps each of which is self-evident on the basis of a mere understanding of the concepts concerned. In short, as Descartes says, *certain* knowledge resolves into *intuition* and *deduction*. Intuition, or contemplation of the basic concepts of a mathematical theorem, gives us knowledge of the self-evident truths or *axioms* which are needed; and then deduction leads us from these axioms to the theorem itself, step by step. Such is the nature, Descartes says, of *all* certain knowledge; but his inspiration is mathematics. So Descartes's method for reconstituting philosophy on a solid basis was to make a clean sweep of everything by provisional doubt, and then begin anew with self-evident axioms gained by intuition. "The things which we conceive very clearly and distinctly," he said, "are all true." And this virtue attached, he found, to his first axiom: "I think, therefore I am." "I see very clearly," he wrote, "that to think it is necessary to be."¹⁵ So here was a first axiom, furnished by intuition.

Descartes's next principle was the existence of God. Evidently he regarded this as a case of deduction rather than intuition, since he offered a *proof*—the so-called ontological proof; yet there is no attempt to indicate axioms from which the deduction proceeded, much less any rules of deduction. The so-called proof, however, was this: *I find* (perhaps by introspection; maybe this is an axiom furnished by intuition) *that I have an idea of a perfect being. But non-existence is an imperfection* (perhaps intuition again: apprehension of part of the essence of the concept of perfection: or, as we should now say, grasping part of the meaning of a term). *Therefore the perfect being of which I have an idea exists.* This proof is not original; goes back to Anselm.

Both of these arguments are bad, as you all know. The "cogito ergo sum" depends on a prejudicial phrasing, "I think," subject and predicate, imputing an attribute (thinking) to a substance, a mind. He assumed his "I am" in his "I think;" circular. Couldn't get it from mere assumption that thinking was going on. When he wrote "I see very clearly that to think it is necessary to be," he was quite right. To do anything it is necessary to be. In

15. <René Descartes, *Descartes: Selections*, ed. Ralph M. Eaton (New York: Charles Scribner's Sons, 1927), 30.>

fact, even not to do anything it is necessary to be. Everything *is*. Necessarily, everything *is*. *I*, if such there be, *am*—*cogito* or *non cogito*. If Descartes felt that it was immediately obvious to intuition that *I think*—imputing thinking to an object which is *I*—he might as well have felt it immediately obvious to intuition that *I am*—*without* any deduction, without any dependence on *I think*. If on the other hand he felt the thinking was more obvious than the being, all he was justified in accepting at the start was “Thinking is taking place.” From this he might fairly plausibly argue “Thought *is*,” but not “*I am*.”

The fault with the ontological argument is a misconception of existence. Non-existence isn't an imperfection, in the sense that non-existent beings are not perfect. Rather, there *are* no non-existent beings, perfect or otherwise. Everything exists, and so on one point Descartes is trivially right: any perfect being exists. Perhaps the real trouble is in the prejudicial phrase “I have an idea of a perfect being,” i.e., “There is a perfect being of which I have an idea.” Of course the latter implies existence, by assuming it; same would hold of a unicorn, *if* there were a unicorn of which I have an idea. But when I truly say “I have an idea of a unicorn,” I say it not in that sense but in the sense “I have a unicornuine idea,” and from this there follows no unicorn. The same holds for God. And the fact that perfection implies existence is no help—in the sense that every perfect being exists; indeed unicornuinity implies existence in exactly the same way.

It may very well be that, as Descartes says, “The things which we conceive very clearly and distinctly are true.” But, as he himself added, “there is some difficulty in ascertaining which are those that we distinctly conceive.” And in particular he overestimated the distinctness with which he conceived his “*cogito ergo sum*” and his ontological proof of God.

Descartes went on thus, using his method of affirming that which by his fallible standards was clear and distinct, until he had spun out a whole system of philosophy scarcely more rigorous and restrained than the mistrusted systems of his predecessors. God, once misproved to exist, served as a bulwark to Descartes's method of continuing to infer truth from clarity and distinctness; for God, being perfect and therefore good, would not deceive us—at least not in so ruthless a way as to allow us clear and distinct conception of falsehoods.

This is pretty lame, and I think that from the point of view of Descartes's actual motivation it puts the cart before the horse. God being the cart, and clarity and distinctness the horse. Because it is evident from Descartes's

Discourse on Method that the one principle of which he was basically convinced, prior to any theological inferences, was that clearness of ideas is a guarantee of truth. “The things which we conceive very clearly and distinctly are true.” And this was doubtless a generalization of Descartes’s experience with mathematics, and with the intuiting of mathematical axioms. I venture to say that Descartes became convinced of the validity of this precept inductively, through his mathematical experience, and afterward cast about for a reason *why* the things we conceive clearly should be true. Why should the self-evidence of mathematical axioms be a guarantee of their *truth*, rather than merely a compulsion to belief—possibly mistaken belief—on our part? And similarly for any other self-evident truth? And the explanation which Descartes concocted, after the fact, was that God is too good to make falsehoods appear to us as self-evident truths. Descartes’s theory was that we have innate ideas, implanted by God, and that we have only to reflect clearly upon them in order to see them in their true relations. (As we do, e.g., in reducing a theorem of geometry to its first beginnings.) And God wouldn’t have implanted these ideas in us otherwise than as they should be.

Leibniz, writing at the end of the same century (seventeenth), made a more devious and refined use of God in connection with certainties about the world. Since God is good, he must have made the world as good as it logically can be. If anything were better it would have to be at the expense of something else’s being worse, through *necessary* incompatibilities. And by the apprehension of clear and distinct ideas it is theoretically possible to determine such relationships of *necessary* incompatibility. It is likewise theoretically possible to determine in this way what is good, what is better than what else. So that, if the human mind were up to such complex reasoning, it could determine—purely by taking thought, and without recourse to observation and experiment—all the facts about the world. Reminiscent of Plato’s Idea of the Good, as the idea whose apprehension is the all-inclusive form of knowledge. Knowing what is good, what is best all things considered, is for Leibniz knowing what is the case.¹⁶ Actually infinitely complex, therefore impossible for us; so we fall back on methods of probability as a second best.

16. <See, e.g., G. W. Leibniz (1646–1716), “Discourse on Metaphysics” (1686) and *Monadology* (1714).>

Here we have a clearer awareness, than in Descartes, of the difference between truths of fact and truths of reason; not as a difference in reality (all is in a sense necessary, since the optimity of the world follows necessarily from the necessary goodness of God), but as a difference in method of knowing. *Ratio cognoscendi* vs. *Ratio essendi*. But Leibniz followed Descartes in his trust in the contemplation of clear and distinct ideas, innately implanted in us as a means of our penetrating far into reality by the pure light of reason. If he was more rigorous than Descartes, the difference is one of detail.

Descartes's focusing of attention on the methodology of knowing did not restrain him, nor various of his successors such as Leibniz, from flights of undisciplined metaphysical fancy. Yet it did have an important restraining effect elsewhere in later philosophy. As Eaton says, "It sets the fashion for philosophizing outward from the inner world of self-consciousness."¹⁷ This is Descartes's step toward Locke and Hume.

John Locke, toward the end of the same century (seventeenth), faced the methodological issue with complete directness by turning his attention squarely to the human understanding itself, its mode of operating, and its potentialities. The purpose of Locke's great work, *Essay Concerning Human Understanding* (1689), was "to inquire into the original, certainty, and extent of human knowledge, together with the grounds and degrees of belief, opinion, and assent." Book 1 of this essay is devoted to arguing that there are no innate ideas.¹⁸ Ideas implanted by God, and to be trusted for their veracity because of the goodness of God, had been¹⁹ a crutch for Descartes. This notion of innate ideas was indeed much older than Descartes, and a part of the long tradition. It appeared in Plato in an extravagant form as the doctrine of *reminiscence*, which appears for example in the *Meno* and *Phaedo*: the doctrine that so-called learning is only a matter of bringing out into the open what we have latently known all along. This is the *radical* doctrine of innate ideas: *all* ideas are innate, and all education is *literally*

17. <Eaton, "Introduction," in *Descartes*, xxx.>

18. <John Locke, *The Works of John Locke*, ed. J. A. St. John (London: Henry G. Bohn, 1843). This edition was reissued in two vols. in 1854, 1872, 1882, and 1892. Quine clearly used one of the two-volume issues, but which one is unknown. The pagination of each issue from 1854 through 1892 remained identical, so this passage (*Essay* 1.1.2) is to be found at 1: 128.>

19. <A>s we have seen <above, 55–56.>

education: *educing* what is already present within us. The doctrine of reminiscence must have seemed particularly apt as applied to mathematics: all manner of surprising things about circles and triangles could be brought out of an ignorant, untrained mind by the right sort of step-by-step questioning. You can extract the most abstruse theorem from an untrained mind by going back to self-evident fundamentals and asking leading questions, step by step, until the proof is complete.

This appeal of the example of mathematical method, which must have tempted Plato into his doctrine of reminiscence, surely operated equally to encourage Descartes in his belief in innate ideas. But Locke made a clean sweep of the whole theory of innate ideas. I believe more good than harm has come to subsequent philosophy from Locke's ignorance of mathematics; the hypnotic effect of contemplating the miracle of mathematical certainty was a danger to which Locke was immune. Locke viewed the mechanism of learning and understanding with a candid eye and saw that knowledge comes from outside, from experience; and that there was no natural place for innate ideas in the mental economy. And though Locke's throwing over the tradition of innate ideas was no doubt psychologically dependent on his not having the mathematical bent of Plato, Descartes, or his own contemporary Leibniz, the result was beneficial even for purposes of an eventual clearer understanding of the nature of mathematical knowledge itself.

Locke's repudiation of innate ideas was the beginning of *empiricism*: the doctrine that knowledge stems from experience. If there are no innate ideas, and innate knowledge of truths concerning them, everything we know must come to us through our senses. The mind begins as a blank tablet, upon which sense experience writes. In sweeping away innate ideas, Locke swept away cherished certainties and prepared the way for skepticism. It is not obvious why innateness should be so bound up with certainty. Couldn't we have innate *wrong* ideas as well as acquiring wrong ideas—if there are to be innate ideas? Would it be much more heinous of God to deceive us in the one way than in the other? The reason Locke's repudiation of innate ideas prepared the way for skepticism is simply this: he brought the knowing process out into the open, where common sense could be brought to bear on it. If knowledge is to be traced to sensations, let's trace it. We know our sensations pretty well, can enumerate them more or less, and thus have a working basis for detailed analysis of knowl-

edge. The mysteries of innate ideas afforded dark places where wishful thinking could flourish; and this Locke cleared away.

—To here June 19

Towards end of last meeting, suggested not evident why innateness of idea should be so bound up with certainty, and antithetical to skepticism. Different, indeed, later with Kant. In Kant, in the eighteenth century, innate ideas reappear in a form which *does* account for the certainty of innate knowledge regarding them. Kant's notion is that the structural limitations of the mind compel all experience to conform to certain structural norms in order that it be experienceable at all. But there was no comparable treatment of innate ideas by Locke's predecessors. But with Descartes and the others of Locke's predecessors, the connection of innate ideas with certainty seems to be essentially negative: merely that explicit examination of the knowing process was thwarted. But Locke, with his *tabula rasa*, laid the knowing process open to critical examination and thus prepared the way for skepticism. Yet Locke, like Descartes, was only a way-station in the movement toward skepticism. The end isn't reached until the next century, in Hume. It is characteristic of Locke to stop short of radical conclusions which his own observations bring within easy reach. There are several examples, as we shall see.

Locke's most central idea is the idea of *idea* itself. According to Locke our first ideas are direct sense-impressions. Further ideas are formed from these by combination—either through perceiving them in combination, or through combining them in the imagination. Locke showed a persistent tendency to mentalize, or subjectivize, the things which in the Aristotelian tradition had been spoken of as real essences. "The essences of the . . . species of things are nothing else but these abstract ideas" (3.3.12). "All the great business of genera and species, and their essences, amounts to no more but this: That men making abstract ideas, and settling them in their minds with names annexed to them, do thereby enable themselves to consider things, and discourse of them as it were in bundles, for the easier and readier improvement and communication of their knowledge; which would advance but slowly were their words and thoughts confined only to particulars" (3.4.end.).²⁰

An incidental consequence of Locke's attitude on essences was that in

20. <This citation is incorrect; it should be 3.3.20.>

biology he saw no reason to believe in the immutability of animal and vegetable species, such as had been supposed from Aristotle onward (3.6.14ff.). This is a question of fact, not to be decided *a priori*; but the doctrine of the subjectivity of essence was calculated to make the mind more receptive to the fact of instability of biological species. Even given the Aristotelian view that there was something more to species (biological and otherwise) than arbitrary or conventional grouping or degrees of resemblance, it would have been logically *possible* that heredity was not species-closed. But the thought was more tempting when species were pried loose, by Locke, from absolutism. So Locke paved the way for Darwin.

Locke's doctrine of subjectivity of essence, subversive of tradition though it was, is one example of Locke's stopping short of yet more radical conclusions which he has brought within easy reach. For, there are many passages which strongly suggest that instead of stopping with subjective *ideas*, he might just as well have thrown these over in turn and stopped with sheer *words*. We find "name" and "idea" running parallel in passage after passage, as if "name" might serve the purpose alone. "By the same way that they [people] come by the general name and idea of man, they easily advance to more general names and notions" (3.3.8). Again: "To be a man, or of the species man, and to have a right to the name man, is the same thing" (3.3.12). [In response to a question on "simplicity" or "epistemological priority," sketched my whole double-aspect theory of idealism and realism.²¹] Again: "Between the nominal essence and the name there is so near a connexion, that the name of any sort of thing cannot be attributed to any particular being but what has this essence, whereby it answers that abstract idea whereof the name is the sign" (3.3.16). Again: "The names of simple ideas, and those only, are incapable of being defined" (3.4.7). Here an awareness that other names *could* be defined—hence no need, we might suppose, of presupposing anything else (so-called ideas) to mirror them. Why not keep just the simple ideas (sense impressions), with their names; and then let *defined names* alone do the work of so-called complex ideas? Again: "Though therefore it be the mind that makes the collection, it is the name which is as it were the knot that ties them fast together. What a vast variety of different ideas does the word triumphus hold together, and deliver to us as one species! Had this name been never made,

21. <Possibly Quine reviewed the sorts of considerations he later published in, e.g., "On What There Is," *Review of Metaphysics* 2 (1948): 21–38.>

or quite lost, we might, no doubt, have had descriptions of what passed in that solemnity; but yet, I think, that which holds those different parts together, in the unity of one complex idea, is that very word annexed to it; without which the several parts of that would no more be thought to make one thing, than any other show, which having never been made but once, had never been united into one complex idea, under one denomination. How much, therefore, in mixed modes, the unity necessary to any essence depends on the mind, and how much the continuation and fixing of that unity depends on the name in common use annexed to it, I leave to be considered by those who look upon essences and species as real established things in nature" (3.5.10). And in summary: "The consideration then of ideas and words, as the great instruments of knowledge, makes no despicable part of their contemplation who would take a view of human knowledge in the whole extent of it: and perhaps, if they were distinctly weighed and duly considered, they would afford us another sort of logic and critic than what we have hitherto been acquainted with."²² Yet Locke never took the step of throwing the complex ideas overboard and letting names do their work for them. This was left for Tooke.

Tooke's doctrine of abbreviations: a single word is introduced in place of a complex of words; and there is no need to suppose a corresponding single idea introduced as a complex of ideas. For Tooke there *is* no *combining* of ideas. There are the simple ideas (corresponding to sense perceptions), there are the names for them, and there are the abbreviations of complexes of such names.²³ Tooke writes: "I consider the *whole* of Mr. Locke's essay as a philosophical account of . . . abbreviations in language" (p. 24).²⁴

22. <Locke, *Essay* 2.21.4.>

23. <Quine here inserted a card reading: "How many haven't had *any* history modern philosophy (Descartes and successors)? Read it (Weber, Windelband, Russell, Cushman, etc.), Descartes through Hume (including Leibniz, Locke, Berkeley) right away. Maybe it will illuminate your notes." The works referred to are: Alfred Weber, *A History of Philosophy*, trans. Frank Thilly (New York: Charles Scribner's Sons, 1896); Wilhelm Windelband, *A History of Philosophy; with Especial Reference to the Formation and Development of its Problems and Conceptions*, trans. James H. Tufts. 2nd edition (New York: Macmillan, 1901); Bertrand Russell, *A History of Western Philosophy* (New York: Simon and Schuster, 1945); and Herbert Cushman, *A Beginner's History of Philosophy*. 2 vols. (Boston: Houghton Mifflin, 1910–1911).>

24. <John Horne Tooke, "Ἐπεα πτερόεντα; or, *The Diversions of Purley*. 2 vols. (Philadelphia: William Duane, 1806–1807), 1. This is the edition Quine appears to be using—very possibly the copy given him by B. F. Skinner. See Quine, *Time*, 110. Elsewhere (e.g., in *The*

—June 21

Tooke suggests that if Locke had said “word” wherever he said “idea,” the essay would have been clearer and more correct. “Perhaps it was for mankind a lucky mistake (for it was a mistake) which Mr. Locke made when he called his book, *An Essay on Human Understanding*. For some part of the inestimable benefit of that book has, merely on account of its title, reached to many thousands more than, I fear, it would have done, had he called it (what it is merely) a *Grammatical Essay*, or a *Treatise on Words*, or on *Language*. The human *mind*, or the human *understanding*, appears to be a grand and noble theme; and all men, even the most insufficient, conceive that to be a proper object for their contemplation: whilst inquiries into the nature of *language* (through which alone they can obtain any knowledge beyond the beasts) are fallen into such extreme disrepute and contempt, that even those who ‘neither have the accent of Christian, pagan, or man,’ nor can speak so many words together with as much propriety as Balaam’s ass did, do yet imagine *words* to be infinitely beneath the concern of their exalted understanding” (24f.–n.).²⁵

Tooke’s repudiation of complex ideas should not be thought of as a theory that discourse calls for no mental activity; nor that adoption of complex definitions (“abbreviations”) involves no thought. I think the point is rather that the bringing in of intermediary unexplained *units*, “ideas,” is a poor way to clarify the process which connects stimulation with discourse. Obscures rather than explains; like the substantial forms of Toletus and the other mediævals, against which Descartes inveighed;²⁶ or like the *vertu dormitif* in Molière. Tooke has been followed (unknowingly?) by modern scientific linguists and by the behaviorists in psychology.

One more relevant passage from Locke: “This whole mystery of genera and species, which make such a noise in the schools, and are with justice so little regarded outside of them, is nothing else but abstract ideas, more or less comprehensive, with names annexed to them” (3.3.9). Might have been less noise still if these so-called abstract ideas, which make their share of

ories and Things [Cambridge, MA: Harvard University Press, 1981], 212), Quine quotes from what he identifies as a Boston, 1806 edition; I have located no such edition.>

25. <I.e., 24–25, n. (m).>

26. <Franciscus Toletus (1532–1596), Spanish Jesuit, professor of philosophy at Collegio Romano. See, e.g., his *Commentaria una cum quaestionibus in VIII libros De physica auscultatione* (Venice, 1600).>

noise in Locke's own discourses, were abandoned in turn in favor of the "names annexed to them." But this step remained for Tooke.

Vigorously though Locke objected to the doctrine of innate ideas, a remnant of that doctrine, or of the spirit underlying it, remained in his philosophy: Locke accepted unquestioningly the point of view that ideas come first and language after; and that language is a tool for expressing ideas which are already there to be expressed, and are there whether or not attaining expression. Hume, actually, is like Locke on this point; and so is Berkeley. Neither of them followed Tooke's course. Tooke was the pioneer in this particular form of nominalism; and Tooke wrote later—shortly after Hume's death.

But there were also other important places where Locke stopped short of more radical conclusions which he had brought within easy reach; and on these other points Berkeley or Hume did follow through. One notable case is Locke's theory of matter, and of primary and secondary qualities. Actually Locke distinguishes three sorts of qualities in bodies:

Primary qualities: "bulk, figure, number, situation, and motion or rest of their solid parts."

Secondary qualities: "the power that is in any body, by reason of its insensible primary qualities, to operate after a peculiar manner on any of our senses, and thereby produce in us the different ideas of several colors, sounds, smells, tastes, etc."

Powers ordinarily so-called: "the power that is in any body, by reason of the particular constitution of its primary qualities, to make such a change in the bulk, figure, texture, and motion of another body, as to make it operate on our senses differently from what it did before. Thus the sun has a power to make wax white, and fire to make lead fluid" (2.8.23).

Both second and third are powers, but second mistakenly not regarded so.

This point of view is close to that of Democritus and the other atomists. According to Democritus (420 B.C.), matter is made up of atoms which differ only in shape and size, not in substance. What we take to be differences in color, taste, etc., are just blurred ways of grasping differences in textures of atomic makeup. Though old as Democritus, this is also the point of view of modern physics (nineteenth century)—due in part surely to Locke himself. An ideal of physics: reduce everything to mass in motion. Explain all in terms of impact of billiard balls. Theory of chemical ele-

ments seemed to be a deviation from Democritus, for here were qualitative differences between atoms. But the ideal remained Democritan: absolute and irreducible difference between substances of atoms was felt unreasonable, and *explanatory* value would have inhered in a theory which made pseudo-atoms of these, composed of smaller *real* atoms in different combinations, so that qualitative differences again reduced to bulk, shape, motion.

Mendeléeff's periodic table hinted, in fact, that such an explanation might be possible.²⁷ In twentieth century, borne out conclusively. Differences of so-called atoms are differences of constitution from protons, electrons, etc. And we now know that elements can be transmuted: natural breakdowns from radium through uranium to lead; and artificially—even to creating a new element, Pl. We aren't back to Democritus, for there remains difference between a few kinds of sub- (or real) atoms (electrons, protons, neutrons, positrons). And current physics has departed in far more fundamental and bewildering ways: relativity of space and time; discontinuity;²⁸ indeterminacy; transmutability of energy and matter. But this is because we can't help it. We'd prefer the other kind of theory still, if it worked (and didn't have to be forced by excess complication). There is something fairly fundamental in our preference for explaining all properties in terms of what Locke calls primary; and it was to this preference that Locke gave voice. Yet all the qualities, primary and secondary, are conveyed to us by our senses. Why the distinction? Why the premium on qualities of the kind conveyable by the pressure component of the sense of touch? (As opposed to wet-dry, hot-cold, and non-touch.)

—To here June 23 <i.e., June 24>

Locke says why he thinks the secondary qualities aren't in the objects: "Let us consider the red and white colors in porphyry: hinder light from striking on it, and its colors vanish. . . . Can any one think any real alterations are made in the porphyry by the presence or absence of light?" (2.8.19) "Pound an almond, and the clear white color will be altered into a dirty one, and the sweet taste into an oily one. What real alteration can the beating of the pestle make in any body, but an alteration of the texture of it?" (2.8.20) *Hume* cites further reasons, in describing this philosophy

27. <Dimitri Mendeleev (1834–1907), Russian chemist, most prominent of a number of nearly cotemporaneous discoverers of periodic law.>

28. Minimal quanta of energy.

(1.4.4): "A man in a malady feels a disagreeable taste in meats, which before pleased him the most. . . . That seems bitter to one, which is sweet to another. . . . Colors reflected from the clouds change according to the distance of the clouds, and according to the angle they make with the eye and luminous body. Fire also communicates the sensation of pleasure at one distance, and that of pain at another."

The question isn't quite dispelled. Shape and size also change with our point of view. (Illustrate.) And motion. (Train.) We say the ellipse isn't in the penny, etc. O.K., say the changing color isn't in the porphyry, but some standard optimum-condition color *is* (analogous to optimum-condition circularity of penny). Point seems to be that we have *another* perhaps more *basic* sense which fixes those optimum-condition shapes: touch. Not so with color, etc. A reason why this sense is felt to be basic was advanced by late L. J. Henderson²⁹—fanciful but interesting: The sense which communicates the primary qualities is the most dependable sense, because the oldest: prenatal. In womb, the foetus senses no sound nor color, etc.; nor even temperature, wet-dry; nor, ordinarily, pain; just resistance (kicking). And it is this particular part of the sense of touch that conveys the *primary* qualities. Locke's view that color and various other qualities were subjective, and not inherent in the objects, was not new with him. As remarked, it goes back to Democritus; and it occurs also in seventeenth century in Hobbes and Descartes. But the distinguishing out of primary qualities, and exempting them from this subjectivity, was vital for Locke for a special reason: it was only thus that he was able to preserve his belief in the existence of material substance at all.

Prior to Locke, the traditional belief in material substance thrived. Substance was one thing, attribute another; and matter was a substance whose *existence* did not *depend* on the sensible qualities. Matter was a substance *having* the qualities, insofar as they were really objective qualities of matter, or *causing* them, insofar as they were subjective; and so long as this point of view persisted, a shift of one or another of the supposed qualities of matter from objective to subjective status in no way jeopardized belief in matter as such. But Locke changed this underlying situation by coming forward with a new, more skeptical doctrine of *substance*: "There is an . . . idea which would be of general use for mankind to have, as it is of general

29. <Lawrence J. Henderson (1878–1942), a physiologist and sociologist, and a founder and Senior Fellow of Harvard's Society of Fellows. See Quine, *Time*, 108–109.>

talk, as if they had it; and that is the idea of substance, which we neither have nor can have by sensation or reflection. . . . We have no such clear idea at all, and therefore signify nothing by the word substance, but only an uncertain supposition of we know not what, i.e., of something whereof we have no particular distinct positive idea, which we take to be the substratum or support of those ideas we know" (1.4.18).

Then on material substance, in particular, or *body*: "The primary ideas which we have peculiar to body, as contradistinguished to spirit, are the cohesion of solid and consequently separable parts, and a power of communicating motion by impulse. These, I think, are the original ideas proper and peculiar to body" (2.23.17). Locke is hesitant over the point, but seems to see that there is no clear alternative to recognizing that the idea of substance is no more than the idea of the aggregate of constituent properties; the substance is nothing over and above, or rather *under and below*, its properties. This is the only view consistent with Locke's empiricism—his insistence that all ideas (in particular that of substance or body) are ideas of sense or compounds of these. Body—so far at least as we have any idea of it—cannot go beyond the sensible qualities. And this is why it is so vital to Locke, if body (matter) is to be admitted at all, that there be *primary* qualities—qualities inherent in the matter and not merely in us. For the idea of matter is merely the idea of *these* qualities; and if *these* disappear—if all qualities become secondary—matter itself vanishes and there is just sensation, which is not sensation of anything at all.

And this is exactly what happens in the philosophy of George Berkeley, Bishop of Cloyne. Berkeley, who was two generations after Locke and a generation older than Hume, took the radical next step which Locke had declined to take: he reasoned that *all* properties were secondary, in that they were all sensory and there was no justification for giving one sense priority over another; and accordingly, agreeing in effect with Locke that matter was no more than the sum of its primary qualities, he concluded that there was no matter at all. Thus Berkeley concluded (*Principles of Human Knowledge*, section 8) that the words *sensible thing* and *idea* are synonymous; there is nothing outside, over and above our ideas themselves—over sense perceptions and the compounds formed in the mind of the ideas yielded by sense perception. Hume goes along with Berkeley in this, and goes him one better: same for spirit.

These considerations reveal a second way in which Locke failed to draw

the radical consequences of his reflections; failed to make a natural next step for which he had prepared the way.

The first instance was his failure to abandon complex ideas in favor of words; this he left for Tooke. The second instance was his failure to abandon substance; this step he left for Berkeley and Hume.

There is also a third important instance: *causality*. Locke writes: "In the notice that our senses take of the vicissitude of things, we cannot but observe that several particular, both qualities and substances, begin to exist; and that they receive this their existence from the due application and operation of some other being. From this observation we get our ideas of cause and effect. That which produces any simple or complex idea we denote by the general name, cause; and that which is produced, effect" (2.26.1).

Just what is this "producing" that we observe? All that sense impressions can convey is invariable concomitance or succession. [expand.] By Locke's own empiricist standards, there should be no additional idea of causality as "production." A case could be made out against it, in his own terms, as strong e.g., as Locke's own case against innate ideas, real essences, etc. Failure to repudiate causality, in the sense of necessary connection over and above observed concomitance, was a striking case of Locke's tendency to stop short of the more radical and distasteful conclusions. And Locke was stubborn on this one. In defense of causality he came out with arguments typical of the old metaphysical dialectic that Locke usually was so critical of. One of his arguments is far worse than typical; incredibly bad. "Whatever is produced without any cause, is produced by *nothing*; or in other words, has nothing for its cause. But nothing can never be a cause, no more than it can be something, or equal to two right angles." Quoted by Hume, with proper scorn, 1.3.3. Sheer play on words; reminiscent of Lewis Carroll;³⁰ hard to imagine Locke taking it seriously; a case, surely, of his frantically trying to evade the natural consequences of his point of view. It was left to Hume to face these consequences; they even passed Berkeley. But the consequent skeptical theory of causality, which denies all justification for any idea of causality beyond observed uniform concomitance devoid of any guarantee for the future, is the keynote of Hume's theory of knowledge (or non-knowledge).

30. <See below, 90–91.>

—To here June 26

(*Treatise* 1.1.1) What Locke called ideas are called by Hume *perceptions*, and divided into *impressions* and *ideas*. Impressions are our actual sense experiences as they come; ideas include memories of these, also objects of the imagination, also all other ideas in the ordinary present day sense of the term (whatever that is). In Locke's time as well as Hume's, I suppose "idea" had as part of its connotation the characteristic of being copied or otherwise derived, somehow, from something more real or fundamental: physical objects, for Locke, the causes of simple ideas in sensation. I suppose Hume's change of terminology reflects his added skepticism regarding those physical objects. He prefers to think of the sense experiences themselves as the fundamental things, and leave the question of external objects open for further debate. Therefore no longer appropriate to call the content of sensation *idea*, insofar as idea connotes being derived or copied from something *more* fundamental. Therefore the word *impression* instead. "Impressions are compleat in themselves" (p. 458). Unlike Locke: *two-fold* form: exact copies one of other; different only in intensity.

For Locke, same idea—now in sense, now in memory and imagination. Locke speaks of essences, Hume of events or types of events? Distinction between impressions and ideas: vivacity. But Kemp Smith, pp. 210f³¹ and 2½ <"in sleep, in a fever, in madness, or in any very violent emotions of soul, our ideas may approach to our impressions: As on the other hand it sometimes happens, that our impressions are so faint and low, that we cannot distinguish them from our ideas" (*Treatise*, 2)>. Or can we defend Hume by saying that it *is* a matter of degree with marginal cases and no further arbiter? Therefore not a case of confusion (in sense of error) so much as indeterminacy, settled by systematization?

Besides into impressions and ideas, perceptions are cross-classified into Simple and Complex (Figure 3.3). (Following Locke.) 3

<"There is another division of our perceptions, which it will be convenient to observe, and which extends itself both to our impressions and ideas. This division is into SIMPLE and COMPLEX. Simple perceptions or impressions and ideas are such as admit of no distinction nor separation. The complex are the contrary to these, and may be distinguished into parts. Tho' a particular colour, taste, and smell are qualities all united to-

31. <Norman Kemp Smith, *The Philosophy of David Hume: A Critical Study of its Origins and Central Doctrines* (London: Macmillan, 1941), 210ff.>

Impressions		
Ideas		
	Simple	Complex

Figure 3.3.

Impressions	  	
Ideas	  	
	Simple	Complex

Figure 3.4.

gether in this apple, 'tis easy to perceive they are not the same, but are at least distinguishable from each other" (*Treatise*, 2).>

How to construe "admit of no distinction nor separation"? Turns out in discussion of extension that (in case of color) must be minimally extended. Elsewhere 8½ scarlet seems simple. <"To give a child an idea of scarlet or orange, of sweet or bitter, I present the objects, or in other words, convey to him these impressions" (*Treatise*, 5).> And explicitly on p. 637: "Blue and green are different simple ideas."

Among the simples, one-to-one correspondence between impressions and ideas (Figure 3.4). 4

<"After the most accurate examination, of which I am capable, I venture to affirm, that the rule here holds without any exception, and that every simple idea has a simple impression, which resembles it; and every simple impression a correspondent idea" (*Treatise*, 3).> But subject to exception later.

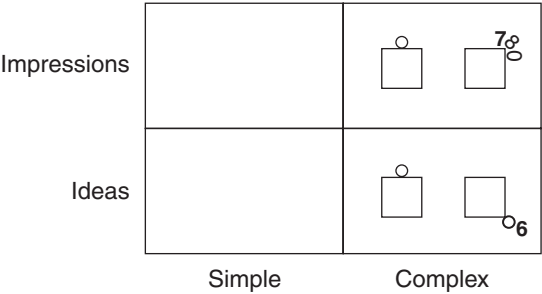


Figure 3.5.

Not so with complex (Figure 3.5). 5

<“I must make use of the distinction of perceptions into *simple and complex*, to limit this general decision, *that all our ideas and impressions are resembling*. I observe, that many of our complex ideas never had impressions, that corresponded to them, and that many of our complex impressions never are exactly copied in ideas” (*Treatise*, 3).>³²

Exceptions to correspondence between simples: 10—serious; from Descartes, *Rules*³³—11 (Figure 3.6).

<“Suppose therefore a person to have enjoyed his sight for thirty years, and to have become perfectly well acquainted with colours of all kinds, excepting one particular shade of blue, for instance, which it never has been his fortune to meet with. Let all the different shades of that colour, except that single one, be plac’d before him, descending gradually from the deepest to the lightest; ’tis plain, that he will perceive a blank, where that shade is wanting, and will be sensible, that there is a greater distance in that place betwixt the contiguous colours, than in any other. Now I ask, whether ’tis possible for him, from his own imagination, to supply this deficiency, and raise up to himself the idea of that particular shade, tho’ it had never been conveyed to him by his senses? I believe there are few but will be of opinion that he can; and this may serve as a proof, that the simple ideas are not al-

32. <The bold numerals 6 and 7 in this illustration signify the following passages in *Treatise*, 3, as indicated in Quine’s annotated copy. 6: “I can imagine to myself such a city as the *New Jerusalem*, whose pavement is gold and walls are rubies, tho’ I never saw any such.” 7: “I have seen *Paris*; but shall I affirm I can form such an idea of that city, as will perfectly represent all its streets and houses in their real and just proportions?”>

33. <Descartes, *Rules for the Direction of the Mind*, rule 14.>

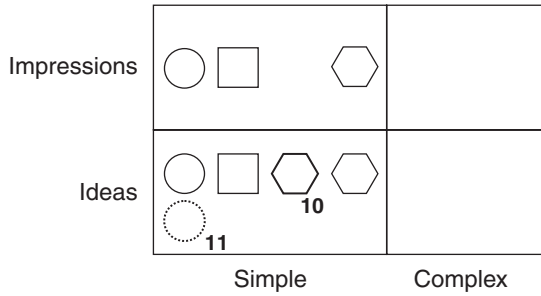


Figure 3.6.

ways derived from the correspondent impressions; tho' the instance is so particular and singular, that 'tis scarce worth our observing, and does not merit that for it alone we should alter our general maxim" (*Treatise*, 6).> 11: <"But besides this exception, it may not be amiss to remark on this head, that the principle of the priority of impressions to ideas must be understood with another limitation, viz., that as our ideas are images of our impressions, so we can form secondary ideas, which are images of the primary; as appears from this very reasoning concerning them. This is not, properly speaking, an exception to the rule so much as an explanation of it. Ideas produce the images of themselves in new ideas; but as the first ideas are supposed to be derived from impressions, it still remains true, that all our simple ideas proceed either mediately or immediately from their correspondent impressions" (*Treatise*, 6f.).>

[A student observes in re 10: since the interpolated simple idea isn't faint as compared with any corresponding impression, it should *be* an impression!]

—To here June 28

Impressions cause the ideas. 8 <"That all our simple ideas in their first appearance are deriv'd from simple impressions, which are correspondent to them, and which they exactly represent" (*Treatise*, 4).> Like Locke: "all ideas come from the senses" (but different in terminology "impression"). As Hume himself says, equivalent to denying the existence of innate ideas. 12 <"the present question concerning the precedency of our impressions or ideas, is the same with what has made so much noise in other terms, when it has been disputed whether there be any *innate ideas*, or whether all ideas

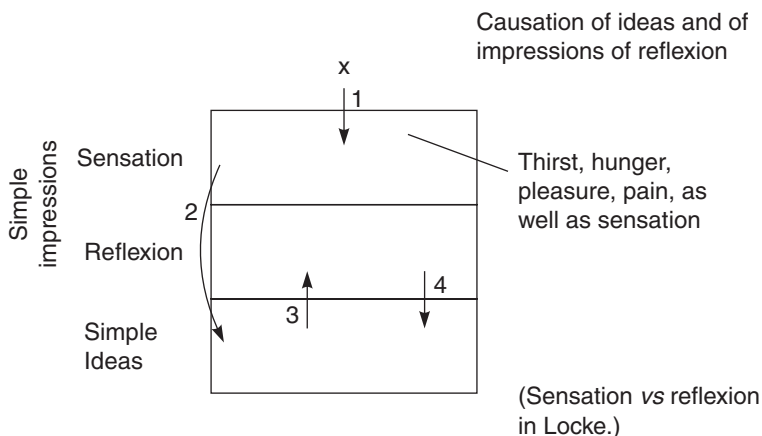


Figure 3.7.

be derived from sensation and reflexion” (*Treatise*, 7).> Like Locke even to the pineapple: 9 <“We cannot form to ourselves a just idea of the taste of a pine-apple, without having actually tasted it” (*Treatise*, 5).> But unlike Locke: cause of impressions unknown. (Except insofar as impression of “reflection” rather than of “sensation” which is the next distinction.)

(*Treatise* 1.1.2) Causation of ideas and of impressions of reflexion (Figure 3.7).³⁴

Church (*Hume’s Theory of the Understanding*, 21): “Since impressions of reflection are subsequent in time to those of sensation, and therefore ‘secondary’ and derivative, why should they not be called ideas?” For, Hume seems to have limited “idea” previously to the derivative or copy. The answer, as Church says, is in Hume, p. 458: Because impressions of reflection are “original facts and realities, complete in themselves, and implying no reference to other passions, volitions, and actions.”³⁵ In what sense “origi-

34. <The bold numerals 1 through 4 in this illustration signify the following assertions in *Treatise*, 7f., as numbered in Quine’s copy of the book: 1: “An impression first strikes upon the senses, and makes us perceive heat or cold, thirst or hunger, pleasure or pain of some kind or other.” 2: “Of this impression there is a copy taken by the mind, which remains after the impression ceases; and this we call an idea.” 3: “This idea of pleasure or pain, when it returns upon the soul, produces new impressions of desire and aversion, hope and fear, which may properly be called impressions of reflexion, because derived from it.” 4: “These again are copied by the memory and imagination, and become ideas.”>

35. <Ralph W. Church, *Hume’s Theory of the Understanding* (London: George Allen and Unwin, 1935), 21.>

Ideas	Impressions:	vivid	Constrained	Simples: Exactly alike except for vivacity
	Memory:	medium		
	Inspiration:	faint	Free	

Figure 3.8.

nal facts and realities, complete in themselves”? Irreducible characteristic? Perhaps a qualitative difference: The compelling quality of emotions, so different from memory and imagination, so forceful and involuntary, like sensations. Really perhaps: an internal sensation (James-Lange: “hope” – “fear”).³⁶ But also they are not representative (supports Smith’s point, read <above> that vivacity isn’t really it³⁷)—whereas ideas (insofar as *simple!*) are. (Except when got by interpolation.)

(*Treatise* 1.1.3) New division of ideas into Memory and Imagination (Figure 3.8).

But Hume is vague: memory an idea? p. 8: “somewhat intermediate betwixt an impression and an idea.” And p. 84 he even speaks of “impressions of memory.” But basically yes; and should be, by criterion of non-originality (*sufficient* condition). Though imagination (“perfect ideas”—p. 8) have distinction of freedom of combinability. Simple ideas can be recombined in imagination in all ways. “Wherever the imagination perceives a difference among ideas, it can easily produce a separation” (p. 10).

(*Treatise* 1.1.4) But this “freedom” of imagination (separability and *recombining*) is not capricious, but guided by gentle forces:

Resemblance,
Contiguity,
Cause and effect.

Cause-and-effect (and the others) can be mediate (e.g., fourth cousin), but then becomes correspondingly weak in its influence on the imagination. 1

36. <The James-Lange theory of emotion entails that emotion consists of the patient’s perception of his or her bodily changes as they occur. See William James (1842–1910), “What is an Emotion?,” *Mind* 9 (1884): 188–205; idem, *The Principles of Psychology*. 2 vols. (New York: Henry Holt, 1890); 2: 442–85; and Carl G. Lange (1834–1900), *Über Gemüths-bewegungen* (Leipzig: T. Thomas, 1887).>

37. <See above, 68, on Smith, *Philosophy of David Hume*, 210ff.>

<“That we may understand the full extent of these relations, we must consider, that two objects are connected together in the imagination, not only when the one is immediately resembling, contiguous to, or the cause of the other, but also when there is interposed betwixt them a third object, which bears to both of them any of these relations. This may be carried on to a great length; tho’ at the same time we may observe, that each remove considerably weakens the relation. Cousins in the fourth degree are connected by *causation*, if I may be allowed to use that term; but not so closely as brothers, much less as child and parent. In general we may observe, that all the relations of blood depend upon cause and effect, and are esteemed near or remote, according to the number of connecting causes interpos’d betwixt the persons” (*Treatise*, 11f.).>

Aristotle was the pioneer in association of ideas; and resemblance, contiguity, and *contrast* were the principles (*On Memory and Reminiscence*³⁸). No clear reason for dropping *contrast*; absorbed into resemblance? Causality is an addition—but an ephemeral one, as we shall see. Word “association” came with Locke; pejorative. Causes of association “are mostly unknown, and must be resolved into *original* qualities of human nature, which I pretend not to explain” (p. 13). Like Newton’s gravity (Laird says Hume says³⁹). 2 <“Here is a kind of ATTRACTION, which in the mental world will be found to have as extraordinary effects as in the natural, and to shew itself in as many and as various forms” (*Treatise*, 12f.).> And “There is a certain regular mechanism, which is susceptible of as accurate a disquisition, as the laws of motion, optics, hydrostatics, or any part of natural philosophy” (Hume, end of *Dissertation on the Passions*⁴⁰). Complex ideas issue from above forces of association, and are classifiable into: Relations, Modes, Substances.

(*Treatise* 1.1.5 (on relations) and 7) Let’s consider relations first from a logical point of view, before examining Hume. And attributes, before considering Hume on abstract ideas. Begin with *predicate*. Simplest kind is an intransitive verb. *One-place predicate*. More generally, whatever you can say about an object is a one-place predicate. Best represented as a matrix in one variable. Now an *attribute* is an abstract entity associated with a one-

38. <451b19–20.>

39. <John Laird, *Hume’s Philosophy of Human Nature* (London: Methuen, 1932), 37.>

40. <Hume, “A Dissertation on the Passions,” in *The Philosophical Works*, ed. Thomas Hill Green and Thomas Hodge Grose. New ed. 4 vols. (London: Longmans, Green, 1882–1886), 4: 166.>

place predicate. Class? Could be; depends on point of view; classes differ from attributes only in "principle of extensionality." Classes (or attributes) come, indeed, to outrun predicates, according to classical mathematics; so not *definable* as "abstract entities associated with one-place predicates," besides vagueness thereof. But this is how we get at them from language (or, as nominalist supposes, how we project them from language). "Universals." These, if we think of them as ideas, are the clearest case of Hume's "abstract ideas." Or, if we don't, the ideas *of* them are the abstract ideas.

Now besides one-place predicates, we may consider two-place. Simplest kind is transitive verb. More generally, whatever you can say (truly or falsely) about two objects. Best represented as a matrix in two variables. Just as attributes are abstract entities associated with one-place predicates, so *relations* are abstract entities associated with two-place predicates. Extensionally—just as attributes may be thought of as classes—so relations as classes of ordered pairs; or, double-barreled classes. Just as abstract as classes, but Hume doesn't treat them under abstract ideas, nor like them. Let's first consider Hume's treatment of abstract ideas. (section 7).

(*Treatise* 1.1.7 (order changed)⁴¹) Hume is a nominalist. He does not believe in universals. Section 7 is about the ideas, and hence perhaps not about the universals themselves, and indeed Hume therefore has little to say clearly on the point of nominalism properly so-called. His thesis in section 7 is rather that there are no abstract ideas except in the sense of concrete ideas used abstractly (with help of language). "No impression can become present to the mind, without being determined in its degrees both of quantity and quality" (p. 19), and ideas are copies of these impressions; an abstract idea is "a contradiction in terms" (p. 19). Perhaps the one place where he commits himself on universals as distinct from their ideas is on foot p. 19, where he simply *affirms* nominalism as "generally received in philosophy." "'Tis a principle generally received in philosophy, that everything in nature is individual, and that 'tis utterly absurd to suppose a triangle really existent, which has no precise proportion of sides and angles." So he takes nominalism for granted, and uses it to prove his conclusion about no abstract ideas, which he infers thus (foot p. 19f.): "If this therefore be absurd in *fact and reality*, it must also be absurd *in idea*; since nothing of which we can form a clear and distinct idea is absurd and impossible." Lat-

41. <Presumably the note refers to the sequence of arguments 1–3 identified in the succeeding paragraph.>

ter principle is a vestige of rationalism used often by Hume: “Nothing of which we can form a clear and distinct idea is absurd and impossible.” “Anything that we can clearly conceive is possible.” So Hume uses this principle to infer that there are no abstract ideas from ontological nominalism: Argument 3.

Not his only argument for the claim that there are no abstract ideas. Another already mentioned:⁴² Argument 2.

And his last argument of all is foot p. 18ff.: abstraction doesn’t imply a separation, because can’t separate the precise length of a line from the line (what kind of thing remains?). And what is not distinguishable is not different. (Essentially same as Argument 2: can’t have line without its length.) Argument 2 is that the abstract idea itself would be indeterminate, therefore impossible; Argument 1 is that the *remainder* after abstracting away the abstract idea would be indeterminate, therefore impossible. The doctrine is Berkeley’s. And like Berkeley, Hume says that what purport to be abstract ideas are particular ones which have “become general in their representation.”

How it works:

- (1) Association by resemblance: present perception induces memory of past similars.
- (2) Association by contiguity: general *name* learned in application to various of those similars.
- (3) Through this compound route, name evokes complex idea of several appropriate similars, as needed. (“Readiness of the imagination to supply others as needed” thus explained; this is the element of generality.)

Note that “similarity in a special respect” is needed here. Suppose first idea raised by word “apple” is a particular green apple seen an hour earlier; *suitable* resemblant is a red apple, but not a green leaf. Our past associations by resemblance must be sorted, so that only those are in “readiness” which resemble the given one in the same *respects* as others do in connection with which the word had been learned by contiguity. <“when a globe of white marble is presented, we receive only the impression of a white colour dispos’d in a certain form, nor are we able to separate and distinguish the colour from the form. But observing afterwards a globe of black marble

42. <Above, previous paragraph, referring to the passage quoted from *Treatise*, 19.>

and a cube of white, and comparing them with our former object, we find two separate resemblances, in what formerly seem'd, and really is, perfectly inseparable. . . . When we wou'd consider only the figure of the globe of white marble, we form in reality an idea both of the figure and colour, but tacitly carry our eye to its resemblance with the globe of black marble; And in the same manner, when we wou'd consider its colour only, we turn our view to its resemblance with the cube of white marble" (*Treatise*, 25).> So a case not merely of resemblance but of overlapping resemblance-*systems*. Hume's analysis of this is wanting. But no criticism of his point of view. (Possible approach: word-habit is readiness to bring forth just such particular ideas as resemble *all* the contiguity-named ideas *more closely* than some of them resemble each other).

—*To here July 1*

In short, a general term has as its "meaning," or anyway behavioral content, the particular "readiness" or "power" or habit which we have of producing right concrete instances in idea. It has no abstraction as additional meaning. Hume defensibly declines to "explain" the existence of such habits. Way the mind works; observation says so. But he gives analogues (pp. 22ff.).

Hume's attitude on abstract ideas throws new light on his notion of simple ideas and impressions, and brings contradictions. Colors would be abstract ideas. Can't have green without a shape. Yet we saw (p. 637, cited above, <69>) "Blue and green are different simple ideas." Maybe a simple idea should *really* be fixed in color and minimal (in view of his remarks on geometry) in size.

Determinacy of ideas in all details is basis of Hume and Berkeley's nominalism. Has since been objected (by W. E. Johnson, Laird, et al.⁴³) that Berkeley and Hume's basic assumption of determinacy of impressions and ideas is wrong. But Moncrieff (here)⁴⁴ points out that Hume would say in cases of indeterminacy that we aren't having the idea at all; *merely* using the word—which indeed we do. As to impressions themselves, question of determinacy depends on what we mean by an impression. Construe impression as something which we can note more or less carefully—as if we were immediately to forget and falsify a detail or delete a minimum-sized

43. <W. E. Johnson, *Logic. Part I* (Cambridge: Cambridge University Press, 1921), xxix; Laird, *Hume's Philosophy of Human Nature*, 58.>

44. <I.e., Bruce Moncrieff, a graduate student attending this course.>

(i.e., simple) part? Then Hume is sustained. Construe as whole locus of vacillation—then Laird etc. are sustained.

(*Treatise* 1.1.5) Hume treats relations quite differently.⁴⁵ What difference does he feel? Relations not so clearly understood in those days. If abstract idea (attribute idea) gives way to ideas of instances, then idea of relation should give way perhaps to ideas of ordered pairs; but too sophisticated. Problem was double: problem of things being connected with others not themselves, unlike attribute; *and* the problem of the abstractness of the connection. It is former that seems to occupy Hume to exclusion of latter. Basically, Hume's notion of relation is *association* of ideas (by resemblance, contiguity, and causation)—an irreducible connection, yet doesn't quite claim to account for "philosophical" relations. Thus,

Relation:

Common language: connection of ideas for association as above.

Philosophical: any circumstance in which we care to compare.

[Logical extreme nowadays: any class of ordered pairs.]

Seven sources of relation in philosophical sense:

1. Resemblance*: This relation needed for all comparison.
2. Identity: (persistence through time) "common to every being whose existence has any duration."
3. Space and time: a species comprising many (distant, above, after, etc.).
4. Quantity or number: a species comprising many.
5. Degree of a quality: a species comprising many.
6. Contrariety.
7. Cause and effect.

Only example of contrariety is existence *vs.* non-existence. 1 < "no two ideas are in themselves contrary, except those of existence and non-existence" (*Treatise*, 15).> And later (p. 66) Hume says there *is* no idea of existence, really, except the idea of the object in question itself. So not much left of contrariety! So he needn't have worried about "exception" (paragraph 6, p. 15). Cause and effect to be examined later.

All these relations obviously, as described (pp. 14f.), turn on contiguity and/or resemblance, which for Hume are the basis. Except "distant" under paragraph 3, Space and Time. Maybe it is for this kind of thing that Hume

45. Doesn't reject (like abstract idea), but admits (Relations, Modes, Substances: section 4).

vaguely thought contrariety needed. Cf. his general neglect of negation. He is logically unclear.⁴⁶ Question arises about resemblance: how can simple ideas resemble without *something* in common, and therefore not simple? Hume went into this as an after-thought: p. 637.

Just what is Hume's thesis about relations? That all relations (philosophical sense, therefore corresponding to all matrices in two variables) are of one of the seven kinds? And that all these reduce to resemblance and contiguity (and cause; eliminable);

plus logical negation (supply);

plus other logical connectives (?);

plus non-relational particulars (simple impressions and their copies and compounds)?

I.e., that all two-place predicates are translatable into these terms, insofar as significant? *No*. Would be tough to establish this. He doesn't claim it; philosophical relations come in as a grudging, irreducible concession. "Hume's theory of Relations."⁴⁷ Important in any case that there *are* primitive relations (resemblance and contiguity), irreducible to the ontology of impressions and ideas. Do we have ideas *of* them? And are they abstract, or at least subject to Hume's criticism of abstract ideas? I leave this to you to ponder.

Papers

Announce: 'Ideas of relations in Hume.'

Read list of paper topics, and commentaries.⁴⁸

46. See last paragraph of section <5> on p. 15: seems to allow negation as a matter of course.

47. <Possibly an allusion to Ralph W. Church, "Hume's Theory of Philosophical Relations," *Philosophical Review* 50 (1941): 353–67.>

48. <Two unnumbered cards in the auxiliary manuscript contain the details of the assignment Quine must have dictated here:

Due July 19.

1st paper: ≤1000 words exclusive of quotes.

All topics are raised by <*Treatise*> 1.1 (pp. 1–25), and touched on in class. But you are expected to include any relevances from elsewhere in book 1. Also consult commentaries, by index.

Choice of 5 topics:

1. Ideas of relations in Hume.

—To here July 3

(*Treatise* 1.1.6) Difference between mode and substance is difference between river and the water that passes through it. Other example of mode: a dance. Not to be confused with abstract idea; another *kind* of distinction. Thus “river,” “dance,” etc., are abstract *too*; but *each* river or dance is a mode. Similarly for substance. “We have no idea of substance, distinct from that of a collection of particular qualities” (p. 116, end first paragraph). [One of the natural conclusions which Locke stopped short of. Repudiation of the substratum.] [Even Locke’s example of gold, malleability, *aqua regia*.] For, only the particular qualities can come through the senses. Similarly for modes. But are they really collections of qualities? Can he mean this? For qualities are gone with abstract ideas; and anyway, is collection thereof a higher abstraction still? Maybe he says more nearly what he means middle p. 16, beginning paragraph two: “Idea of substance as well as that of mode is nothing but a collection of simple ideas, united by the imagination.” Now we know: simple ideas, not qualities; and collection is merely combination in a complex idea, which *is* the imagination’s business. Idea of a piece of gold (a substance) is just our total complex impression of it, or rather its copy in an idea. Similarly for our idea of any particular mode, e.g., a river or a particular procession.

(*Treatise* 1.2.3) Hume’s theory of space (extension) is in the long paragraph on p. 34. It is the idea of visual points (colored minima) occurring together. Therefore idea of visually complex impressions in general (general in the sense of abstracted from them, *selon* <i.e., according to> theory of abstract ideas; see last sentence of paragraph). Visually simple impressions can’t give it, for they have only color and feel; too small for shape, which is reserved to complex impressions. It is the idea of any complex of

-
2. Hume’s theory of relations.
 3. The simplicity of impressions in Hume. (A shade of blue seems simple; but elsewhere abstract; and elsewhere minimum size seems essential.)
 4. Hume’s distinction between impressions and ideas. (Interpolated ideas: why not called impressions? Impressions of reflection: why not called ideas? Also, Hume admits that his vivacity criterion fails of sharpness at points.)
 5. Individuality of perceptions in Hume. (Can a perception recur, or are they two like perceptions? Can a perception endure, or is this a series of like perceptions and how long is each?)

The fifth topic is crossed out, though still legible. After crossing it out, Quine had a change of mind and printed “STET” vertically in the left-hand margin alongside that paragraph.>

simples together as for contiguity-association. Finally (concrete content) it is the idea of any extended object (used representatively, with words, in the manner of abstract ideas). An abstract idea boils down to a *word*, with a "readiness" to bring forth, in connection with it, any of many relevant concrete ideas. And in the present instance it is the word "extension" and a readiness to bring forth any idea of an extended object (which equals a congeries of contiguous minimum color-points or touch-points).

The treatment of time (p. 35) is analogous except that here it is succession instead of contiguity. But there is a difference. Space-points may be qualitatively alike (purple), yet several are better than one and suffice to give extension. Whereas time-points consecutive and qualitatively alike are no good; there must be *change*. Feels, probably, there is a real difference: numerical difference of space-points is capable of being sensed, when thus contiguous, even if there is nothing else. But numerical difference of moments is not. When we *do* sense duration of an unchanging presentation it is because of other changes in background—e.g., internal senses; and then it is the *latter* that gives idea of time.

Extension, or space, comes from contiguity; time from succession. But these are relations. Are they then ideas of relations? No. Hume is pretty explicit in the case of extension: idea not of the contiguity itself but of the complex (in any given concrete case) which *exhibits* the contiguity within it. Presumably so also with time. But there we must recognize complex impressions which are successions, not just configurations. The hearing of a song is as good a *single* complex impression as the seeing of a painting. This is O.K., but worth recognizing clearly. Now a new difference between extension and time, for Hume, has to do with the bounds of the complex impressions which are the concrete embodiments of the respective abstract ideas of extension and time. In the case of space, the concrete complex impressions may, I suppose, be anything up to content of whole visual field at any one time; no more. (Though it is not very clear just how far to carry Hume's subjectiveness. Sometimes, e.g., he even seems to allow the *same* simple idea to be tangible and visual—as if construction of external objects were under way.) But what limit for time?—In the extreme, time would become a single concrete life-long complex impression. For Locke, like Hume, extension and duration (or time) were from contiguity and succession. But, unlike Hume, they were ideas of reflexion; reflecting on arrangement of sensations, or on train of sensations. Hume has no place for such reflexion—which is a source of general difficulty for him—because of

his doctrine that all ideas (practically) must copy impressions. He does have impressions of reflexion, but these are limited to passions, emotions, desires and aversions. “None of which,” (as he says foot p. 33) “I believe, will ever be asserted to be the model, from which the idea of space is derived.” So he is driven to his theory of ideas of space and time as abstract ideas having complex impressions as their concretions.

(*Treatise* 1.2.1) Let’s look back now to Hume’s negative argument for *finitude*. In section 1, argues our ideas of space and time not infinitely divisible. Might distinguish two senses of divisibility of an idea (e.g., of an extended particular):

- (1) division into simples of which it is a complex idea,
- (2) an ultimate relation between one *simple* “big” idea and various *simple* “small” ones.

In either case, infinite divisibility of ideas (or of corresponding impressions) not to be hoped for, because it would presuppose infinitely small extensions as impressions (whether constituent or related), which is in any case absurd. But, which of (1) and (2) is Hume? Apparently (1); see the phrase “*simple and indivisible*,” p. 27, line 7. Conflicts with notion (in places previously cited) that a color is simple. But this conflicted *also* with abstractness of color, which was likewise documented.⁴⁹ Still, as we see, requiring simples to be minimal in extent is *not* essential to his present conclusion that ideas are not infinitely divisible; this would hold under either sense.

But then a final remark undermines the whole: “This however is certain, that we can form ideas, which shall be no greater than the smallest atom of the animal spirits of an insect a thousand times less than a mite” (p. 28). How? Not copies of simple impressions! And not compounds having other *smaller* ones as *parts* (if this is Hume’s sense of part). By some indirect derivation? *How*, then? And, if so, couldn’t we go on without end in like fashion? Or *should* he have said: no idea, but way of using words? There is a great unsolved problem of the status of indirect scientific constructs like electrons etc. And his mite-particle is one such. Maybe they should be tied into the system without presupposing ideas in any *image*-sense (for which Hume has a predilection), but by some derivative constructions regarding use of words and not insisting on ideas at all (spirit of Tooke; might have

49. <Above, 68–69 and 77, respectively.>

been carried further and applied to all complex ideas, as Tooke wanted). This is no solution, but only a hint of a program. Remains to see what the verbal constructions are like, what ones are acceptable to an empiricist in Hume's spirit.

—To here July 5

(*Treatise* 1.2, *passim*)

(Re indirect scientific constants. Electron; mite-particle; geometric point.)

Key: contextual definition. Cf. Russell, descriptions.⁵⁰

Appealing to operational criteria.

Such is modern empiricism.

Some inkling pp. 47f., talking of indirect "correction" of size-comparisons.⁵¹

Carnap, *Aufbau*.⁵² (But presupposes a logic which presupposes universals.)

If such a construction program were put into effect, question of infinitely divisible space would become: are there smaller and smaller entities, without end, even in this realm of contextually defined pseudo-objects or constructions? Might still say no (modern physics); criteria of most remote sorts peter out finally. Then Hume's condemnation of space as unempirical remains valid. Could still *explain* classical geometry as a fiction convenient in reasoning but having no genuine content. But this is an *acceptance* of Hume's allegation, coupled with a pragmatic *use* of geometry, as a short-cut, in full knowledge of its strict absurdity. Jeffreys has a chapter on physical geometry (mensuration) in *Scientific Inference*.⁵³ Same really needed for arithmetic. Hume doesn't face the question: what is our

50. <Bertrand Russell (1872–1970), "On Denoting," *Mind* 14 (1905): 479–93; idem, "Philosophy of Logical Atomism," *Monist* 28 (1918): 495–527; *Introduction to Mathematical Philosophy* (London: George Allen and Unwin, 1919); Alfred North Whitehead and Bertrand Russell, *Principia Mathematica*. 3 vols. (Cambridge: Cambridge University Press, 1910–1913), 1: 66ff.>

51. Cf. also Berkeley, *New Theory of Vision* <sections 52–78>.

52. <Rudolf Carnap (1891–1970), *Der logische Aufbau der Welt* (Berlin: Weltkreis-Verlag, 1928).>

53. <Harold Jeffreys, *Scientific Inference* (Cambridge: Cambridge University Press, 1931), 107–30.>

idea of a trillion? Really have a trillion parts each of which is a copy of a simple impression?

Contextual: ‘There are three things x such that Fx ’ for

$$\begin{aligned} & \exists x \exists y \exists z [Fx \cdot Fy \cdot Fz \cdot x \neq y \cdot x \neq z \cdot y \neq z \cdot \\ & \sim \exists w [Fw \cdot w \neq x \cdot w \neq y \cdot w \neq z]]'. \end{aligned}$$

And similarly for all numbers. The ideal of empiricist construction might be formulated thus:

Assume: statement connectives; quantification; identity; indefinitely many *empirical* predicates, one or many places. (Problem: which are empirical?) Express all other *entire* sentences in this language, or reject.

—To here July 8⁵⁴

What I developed last time can be summarized as a contrast between three senses of empiricist “construction:”

54. <Five unnumbered cards from the auxiliary portion of the manuscript contain an alternative to parts of Quine’s 8 July (figures 3.1a and 3.1b) and 10 July lectures as transcribed here. The most plausible interpretation of the auxiliary material included in this note would appear to be that this alternative was intended as a revision of that portion of the given lectures, for it touches on subjects that Quine identified in the note included in the primary portion quoted above (37) as subjects to which he particularly wished to draw attention were he to teach the course again. Given the fact that these auxiliary portion cards are unnumbered and (at least by the time of their discovery in 2003) physically separated from the body of the lecture manuscript and so in neither way integrated with the primary portion of the manuscript, I suspect that this material was not read in the lectures in the summer of 1946. Quine’s initial treatment of some of the material included in the cards constituting these days’ presentation in the lectures does not contain some of the following material, and some of the following is more fully developed than its complement in the primary portion:

(*Treatise* 1.2.4)

One opinion of mathematicians: infinite divisibility follows from definitions, but objects of geometry are mere ideas, and can’t exist in nature.

Answer: “Whatever can be conceiv’d by a clear and distinct idea necessarily implies the possibility of existence.” <p. 43>

Alternative opinion of mathematicians: “tho’ it be impossible to conceive a length without any breadth, yet by an abstraction without a separation, we can consider the one without regarding the other.” <*Ibid.*>

Answers: (1) (said before) if no minimum in ideas, capacity of mind must be infinite to comprehend infinite number of parts of any idea of extension.

(2) (foot p. 43f.) [bad argument; confuses dimensions? Or proves wrong thing—that e.g., “surfaces are indivisible in depth”—to which mathematician agrees!]

- (1) Naïve (Hume): construction of ideas from constituent ideas (and constituent *spatially*—making cramps).
- (2) Sophisticated: construction of terms from constituent words (which may indeed have ideas connected, but this thought becomes less essential and important now).
- (3) Hyper-sophisticated: construction of *contexts* from constituent words. This carries us much farther than (2), yet is certainly no less admissible. An important discovery. Russell.⁵⁵

This contrast applies not only to the construction of mathematical ideas (number, geometrical elements, etc.), but also physical objects, electrons, everything. But the immediate question that led us to these considerations was that of the infinite divisibility of space.

Hume's arguments to the contrary turned on his naïve conception of construction: an idea of an infinitely divisible object in space would have

Meaning of '=', '>' between lines or surfaces? <p. 45> Comparison of number of points? But incomparable if infinite. Or by congruity? But this means one-to-one agreement of parts; therefore reduces to comparison of number of points.

How actually? First, rough immediate comparison of impressions. Second, mediate [cf. Berkeley, *Theory of Vision*] by reflection. Ideas of comparison are hybrid of both. No precise idea. Infinitesimal is unwarranted extrapolation.

Ideas of straight line, curves, etc. similarly two-fold and imprecise. "<Our> appeal is still to the weak and fallible judgment, which we make from the appearance of the objects, and correct by a compass or common measure." <p. 51> [Operationism. See Jeffreys on Mensuration.]

Mathematical certainty is an illusion, and thus, dubious.

(*Treatise* 1.2.5)

Idea of space or extension is idea of visible or tangible points distributed in a certain order. Therefore no idea of vacuum, unless vacuum be identified with the mere "capacity of becoming a visible and tangible distance" <p. 64> [therefore a reflective idea, as Locke would say?] [But what *is* capacity, potentiality?] [Digress on counter-factuals, etc.]

Two ideas of distance: direct (interval filled by body) and inferred (dimness etc. analogous to first type).

(*Treatise* 1.2.6)

To reflect on anything as existent is to reflect on it. (Usual argument: what other impression?) [To reflect on anything as externally existent is to make an abstraction common to all impressions of outer sense? Doesn't quite say.]>

55. <See n. 50 above.>

to be made up of infinitely many component ideas, and this is impossible. The issue would have to be reconsidered under the sophisticated version of construction: can we define geometrical language *in context* in such a way that

- (a) all full statements containing them boil down to empirically acceptable vocabulary (empirical predicates and logical connectives, etc.),
- (b) the propositions of infinite divisibility become true?

Actually there is an equal problem, not recognized by Hume, in the infinite divisibility of numbers themselves—and even in the infinite generability of whole numbers.⁵⁶ In *all* these problems, the answer—even for the sophisticated notion of construction—is very likely *no*.

This set of problems involves difficulties a proper consideration of which would take us too far from Hume. In part the problem depends on how *much* in the way of auxiliary logical signs we accept. If we accept the whole machinery of modern logic (not just the signs enumerated last time), we can indeed get infinite extent and divisibility of number system. And by devious methods involving appeal to Descartes's analytical geometry we can perhaps hook in the infinitely divisible conception of geometry. *But* this logic entails assumption of abstract objects, contrary to Hume's nominalism—and contrary, I think, to whole Humean spirit (whereas the sophisticated form of construction as such is not). Toeing the mark—using only the mild logical auxiliaries hitherto listed, which do not violate nominalism—I expect the answer is *no* all along the line, so that the negative conclusion given by Hume's own naïve conception of construction comes to be sustained.

I've spoken earlier of how Descartes's thought, and indeed Plato's, was influenced by the impressive certainty of mathematical reasoning.⁵⁷ Hume has seen fit not merely to question this certainty, but to repudiate a conclusion of the mathematicians themselves: infinite divisibility of space. On p. 71, Hume remarks that geometry is inexact and lacking in demonstrative certainty because "it depends on appearances." More particularly, because it depends on shapes, which are complex ideas whose properties reduce to spatial relations *between* simpler ones. And there are external relations, concerning which we cannot gain the *certainty* which we can

56. Former reduces to latter, by taking fractions as pairs.

57. <Above, 52–54.>

have over *internal* relations. This contrast between external and internal relations begins on p. 69. *Resemblance* is an *internal* relation, in the sense that "it is invariable so long as the resembling ideas remain the same." Similarly for *contrariety*, *proportion in quantity or number*, and *degrees in any quality*. Not so for *identity* (as duration), *relations of time and place*, or *causation*.

Seems arbitrary. Survival of old distinction between essence (property) and accident? Why aren't the space-relations of three spots as *intrinsic* to the complex idea of the three spots, as the proportion $\frac{3}{2}$ between it and the idea of two of the spots?

Hume isn't good at logical abstractions. Much better in his conclusions than in the logical underpinnings which he sometimes gives them. He has found trouble in geometry, in the nonsense of infinitely divisible space; he has found no trouble in arithmetic; therefore he supposes a theory of internal vs. external relations which regards external ones as objects of only fallible knowledge, and rests geometry on them, while regarding internal ones as objects of certain knowledge, and resting arithmetic on them (proportions of quantity and number). Actually, as already remarked, maybe he should have felt trouble also in arithmetic—over infinite divisibility, and infinite extensibility.

Of the exactness and certainty of arithmetic he says (p. 71): "There remain . . . algebra and arithmetic as the only sciences, in which we can carry on a chain of reasoning to any degree of intricacy, and yet preserve a perfect exactness and certainty. We are possessors of a precise standard, by which we can judge of the equality and proportion of numbers; and according as they correspond or not to that standard, we determine their relations, without any possibility of error." The basic method is pairing off units, for equality. In more complex cases we "proceed in a more *artificial* manner" (p. 70)

Compare this with Descartes's "intuition and deduction." *Internal* relations, such as one-to-one correspondence of given units, can be *intuited* with complete certainty. The more artificial manner is step-by-step deduction, where each step is in turn intuited certainly. Not so with geometry. But Hume was a little uneasy about his cavalier treatment of geometry. Top p. 72 he tries to concede that geometry works out pretty dependably in *practice*, but he does not make it clear why it should.

By the time of the *Enquiry*, his repentance was complete: he granted geometry, along with arithmetic, absolute certainty. (Section 4, part 1, beginning): "All the objects of human reason or enquiry may naturally be di-

vided into two kinds, to wit, *relations of ideas* and *matters of fact*. Of the first kind are the sciences of geometry, algebra, and arithmetic; and in short, every affirmation which is either intuitively or demonstratively certain.”⁵⁸ He has got away from his old division of relations into internal and external, with its prejudice to geometry. The new division is reminiscent of it, but the phrasing suggests more clearly the following point of view:

- (a) what we can know by considering and comparing the content of ideas, without which they wouldn’t *be* those ideas, vs.
- (b) what we may suppose or observe about the way impressions do or will succeed one another.

A copy of Leibniz’s truths of reason vs. truths of fact; but brings us close to modern (Kant and successors) *analytic vs. synthetic*.⁵⁹

To get this distinction in modern form, useful to switch again (à la Tooke) to talking about words and meaning—relations instead of ideas. And then we get theory of mathematical truth as with modern logical positivists et al.: true because this is the way the words are properly used; otherwise their meaning would be altered. Same for logical truth itself, and easier to see there: ‘ $p \supset p$ ’ because this is the way we use ‘if’. [Illustrate by telling about Lévy-Bruhl.⁶⁰] But difficulty: drawing line. Which principles e.g., of physics, etc.? Some, reasonably. And then there is still, as before, the question (though abandoned in *Enquiry*) of infinite divisibility: *can* the words of such a geometry make sense for an empiricist? Doubtful, as already remarked, even under a sophisticated (but nominalistic) construction.

—July 10

I explained view that logical truths are true by virtue of meanings of (i.e., way we resolve to use) signs; therefore certain. Same doctrine holds for mathematics. In fact, mathematics reduces to logic; but this is a strong

58. <The text of the *Enquiry Concerning Human Understanding* that Quine appears most often to have cited is included in Burt, ed., *The English Philosophers*, 585–689. This passage, with some variation in spelling, is on 598.>

59. <See. e.g., G. W. Leibniz, *Monadology* (1714), sections 33ff.; and Immanuel Kant (1724–1804), *Critique of Pure Reason*, “Introduction,” section 4 (A6/B10–A10/B14).>

60. <Allusion unidentified with confidence, but cf. Quine, *Word and Object* (Cambridge, MA: MIT Press, 1960), 59n.; and idem, *Ways of Paradox and Other Essays*. Rev. ed. (Cambridge, MA: Harvard University Press, 1976), 109.>

logic incompatible with nominalism. Still, even failing such reduction, certainty of mathematics can be attributed to same source: true by virtue of how we use signs. This, as seen, is close to Hume as of *Enquiry*. But there is still a question *what* mathematics is certain? With infinitely divisible space or not? Because it might be (and probably is, given nominalism) that *no* meaningful use of signs could embody *such* principles.

Hume's advance in *Enquiry* is *good*, insofar as recognizing that geometry *and* algebra can be demonstrative because dependent on sheer meanings. But he should continue (as empiricist) to be critical of possibilities of meaningfulness. Perhaps he was in a hopeless quandary because lacking sophisticated idea of construction of ideas; sensed that anything in the way of a critical attitude toward meaningfulness carried him too far, given the poverty of the naïve method of construction; so that at the period of the *Enquiry* he abandoned this side pretty completely, at least for mathematics, and adopted an attitude of submissive acceptance. But the problem is still alive, and worth reconsidering now from the point of view of an enlightened empiricism—empiricistic and nominalistic as before, but armed with the sophisticated conception of construction. But we shall not.

(*Treatise* 1.3.3) Then we come to the most celebrated and iconoclastic of Hume's doctrines: the doctrine of causality. The doctrine, in its negative part, amounts to this: there is no possible logical basis for belief in causal law and scientific prediction. There is no more than a logically unjustifiable habit of mind, or faith in so-called scientific method. Our naïve view of cause: necessary connections, in nature, between events. Things have to be thus and so when others are thus and so. And indeed every event has antecedents of this kind rendering it inevitable. Philosophers long before Hume found trouble justifying what we so naïvely assume on this subject.⁶¹

61. <A small sheet of notepaper contains a reference to Hume's correspondence with John Stewart (d. 1766) concerning presumably the *Treatise* 1.3.3 opening of Hume's argument to the effect that there is no rational ground for belief in causal law and the possibility of justifiable prediction in the sciences. There is no explicit mention of this text in the primary portion of the lectures manuscript. Quine penciled this note possibly in preparation for this course. But it is more likely that he noted this letter after he had delivered his lectures, saving the reference against the possibility of offering the course again in future. The focus of Quine's interest in the letter is Hume's attempt to rectify Stewart's misunderstanding of especially the *Treatise* discussion of causal law, and Quine had not expressly antici-

Certainly the arguments in favor of causality which Hume attacks were indefensible. Hobbes's (p. 80) was best of a bad lot: appeal to "Principle of Sufficient Reason," which figured so prominently also in Leibniz: no reason why something should happen in one moment rather than another if causality not granted. But Principle of Sufficient Reason is, in effect, assumption of causality itself. *Petitio principii*. Samuel Clarke's (p. 80): (1) cause must precede effect, (2) anything uncaused would have to be its own cause, therefore (3) anything uncaused would have to precede itself. First assumption doubtful, though Hume doesn't object to it.⁶² (Actually it may be asked how an immediate cause *can* precede effect; *actio in distans*; and therefore how *any* cause can.) But Hume doesn't need to, for he can rightly attack second premiss: *petitio principii*. Locke's argument (p. 81) for cau-

pated and rebutted this not-uncommon misreading in the primary portion of the lectures manuscript. (Hume's name, in capitals, is printed vertically in the upper left-hand corner of this sheet.)

HUME

Stewart's letter to: "I never uttered such a proposition as that anything happens without a cause. What I said was. . . ." (In Greig.) Descartes already said there were no necessary connections *within* nature. (Only via God.)

This apparent quotation is not from any letter from Stewart to Hume, but rather captures the sense of a passage in *Hume's* letter to John Stewart, M.D. and Professor of Natural Philosophy at Edinburgh, written in February 1754. Hume wrote to Stewart on the occasion of the latter's misrepresentation of the *Treatise* discussion of causation in the essay, "Some Remarks on the Laws of Motion," which Hume was at that time editing for publication. What appears to be a quotation is not in fact a quotation of this letter but a paraphrase of the opening of the pertinent passage. See Letter no. 91, in *The Letters of David Hume*, ed. J. Y. T. Greig. 2 vols. (Oxford: Clarendon Press, 1932), 1: 185–88 on 187. The portion of Hume's letter in which Quine was interested here reads: ". . . I never asserted so absurd a Proposition as *that any thing might arise without a Cause*: I only maintain'd, that our Certainty of the Falshood of that Proposition proceeded neither from Intuition nor Demonstration; but from another Source. *That Caesar existed, that there is such an Island as Sicily*; for these Propositions, I affirm, we have no demonstrative nor intuitive Proof. Would you infer that I deny their Truth, or even their Certainty? There are many different kinds of Certainty; and some of them as satisfactory to the Mind, tho perhaps not so regular, as the demonstrative kind.">

62. <See, e.g., Thomas Hobbes (1588–1679), *Liberty and Necessity*, in *The English Works of Thomas Hobbes of Malmesbury*, ed. Sir William Molesworth. 11 vols. (London: J. Bohn, 1839–1845), 4: 276; and idem, *Elements of Philosophy*, in *English Works*, 7: 85. See also *The Leibniz-Clarke Correspondence*, ed. H. G. Alexander (New York: Philosophical Library, 1956), throughout.>

sality—whatever is produced without any cause is produced by *nothing*, has nothing for its cause—is a sophistry (and indeed elementary linguistic confusion) reminiscent of Lewis Carroll, *Through Looking Glass*, pp. 139–140.⁶³

There *were* indeed philosophers before Hume who repudiated causality—and with as much ulterior purpose as those who defended it. Thus Malebranche and other followers of Descartes repudiated causality—so far as matter upon matter—for the glory of God. (Hume cites, pp. 158ff.). But Hume's argument is without ulterior purpose; and, not stopping with matter, it repudiates cause—in the naïve sense of necessary connections between matters of fact—altogether.

First thing for Hume to examine is this alleged necessity itself. He argues that there is no necessity but constant conjunction (p. 88), because there is no further connection that the senses could give us. No such “necessity” can be *real* necessity in the sense of necessary truths of reason (mathematics), because of conceivability of break of connection (p. 89). The appeal here to conceivability of the opposite is a vestige of the rationalist tradition. A poor criterion, because hard to decide what we are imagining and what we are merely verbalizing without imagining. *Through Looking Glass*, pp. 100–101.⁶⁴ Modern survival of the principle: *contradictions are meaningless* (i.e., can't be clearly conceived). Wittgenstein et al.⁶⁵ See especially p. 95: “What is absurd is unintelligible.” Difficulties: (1) no test of meaningfulness; (2) indirect proof. But Hume's argument remains that the so-called necessity in supposed causal connections is no more than constant conjunction; and this is important.

Now constant conjunction is observed only in present and memory; yet it leads us to expect *same in future*; *this* is apparently bound up somehow with our feeling of a necessary connection. To account for this, Hume inquires into nature of expectation, therefore *belief*. P. 94: Belief consists in

63. <Lewis Carroll, *Through the Looking-Glass, and What Alice Found There* (London and New York: Macmillan and Co., 1895). This is pretty clearly the edition Quine used. This passage, where Alice claims to see “nothing” on the road to the White King's astonishment and envy, is in chapter 7, “The Lion and the Unicorn.” Cf. above, 67.>

64. <Carroll, *Through the Looking-Glass*, chapter 5: “Wool and Water,” where the White Queen claims to believe sometimes as many as six impossible things before breakfast in response to Alice's assertion that one cannot believe impossible things.>

65. <E.g., Ludwig Wittgenstein (1889–1951), *Tractatus Logico-Philosophicus*, trans. D. F. Pears and B. F. McGuinness (London: Routledge and Kegan Paul, 1961), 4.461.>

entertaining an idea *as realized in existence*. But existence is not a predicate; an idea of a horse or unicorn is no different from the idea of same as existing. (Or of God; cf. my remarks on Anselm's ontological proof of God, *à propos* Descartes.⁶⁶) So difference between believing and not must lie in manner of entertaining the idea, not in the idea itself.

Hume's solution is same as in his distinction between memory, imagination, and impression: *vivacity*. So, definition of belief (p. 96): *a lively idea related to or associated with a present impression*. Particularly suited to belief in a constant conjunction continuing: present reappearance of A, lively idea of B as expected. More *general* version of same theory, of belief as vivacity simply, occurred way back on p. 86, before the theory was argued: "the *belief* or *assent*, which always attends the memory and senses, is nothing but the vivacity of those perceptions they present; and . . . this alone distinguishes them from the imagination." Certainly fits, if the distinction between memory and imagination itself is right; for, the big difference between the two is that we *believe* memory. And there are minor empirical corroborations which Hume also notes, notably (p. 117): "I have often heard in conversation, after talking of a person, that is in any way celebrated, that one who has no acquaintance with him, will say, *I have never seen such-a-one, but almost fancy I have; so often have I heard talk of him.*" In particular, then, the belief accompanying so-called necessary connection—rather, accompanying the experience of past constant conjunction—is the extra *vividness* with which a new encounter of A leads us to imagine the conjoined B *by association*. Association by constant conjunction is based *always* on resemblance no less than contiguity. For, it is based on "repetition"—but there is no perfect repetition; only resemblance among succeeding quasi-instances of the quasi-repeated: All expectation of one object because of another is by custom (p. 102).

—To here July 12

So the "necessity of connection" is just the constant conjunction; and the expectation engendered is accounted for by association. Whence then our *idea* itself of necessary connection in this sense? For Hume it is an idea copied from an *impression of reflection*—internal sensation of compulsion to associate (foot p. 165). Observed *constant conjunction* invokes this impression, since it strongly compels us to associate (by resemblance plus association)—see p. 156. The "necessity" we feel is subjective—necessity on our part to associate or expect; and it is this that we sense internally when

66. <Above, 55–56.>

we sense the impression of reflection which is the prototype of the idea of necessary connection. And so expectation by so-called scientific method is a species of sensation (p. 103).

An interesting confirmation of this theory of Hume's comes from contemporary psychiatry: in the neurotic phenomenon of *indecision*. A person whose troubles or mental discomforts are too much to be conveniently borne usually protects himself by repressing his feelings *in general*, so that there results an apathy and a dull depression, taking the place of the grief he is fleeing from. Emotions generally—impressions of reflexion, in Hume's phrase—are less *vivid*, converging in a dull neutral. And what happens to belief? Pros and cons become about equal and there is indecision—as there should be by Hume's theory, if the sense of necessary connection is an impression of reflexion!

Now *cause*, provisionally taken as a basic type of association-relation along with resemblance and contiguity, comes now for Hume to *reduce* to these latter; nothing additional. Definition appears top p. 170: <“We may define a CAUSE to be ‘An object precedent and contiguous to another, and where all the objects resembling the former are plac'd in like relations of precedency and contiguity to those objects, that resemble the latter.’”> Thus:

(type A)	(type B)	
o	—	horizontal = contiguity (or succession)
	—	
o	—	
	=	
o	—	
	—	
o	—	Then, granted the <i>whole columns</i> horizontal <i>also</i> = cause-effect.
o	—	
o	—	
	=	
o	—	
	—	
o	—	
	—	
o	—	(Note there may be extras of type B.)
	—	

No doubt “resemblance” in Hume's definition is too vague, for it could allow resemblance in unlike respects. Suitable refinement would probably be like what I suggested in connection with Hume's theory of abstract ideas.

Thus: If $a_1, a_2, \dots, a_n$ are respectively contiguous to (or similarly succeeded by) $b_1, b_2, \dots, b_n$; and everything that resembles all of $a_1, \dots, a_n$ more than some of $a_1, \dots, a_n$ resemble each other is included among $a_1, a_2, \dots, a_n$; and similarly for $b_1, b_2, \dots, b_n$. Trouble here is that we cover only full correlation; no allowance for there being extra b 's with other causes.

Maybe "respect of resemblance" is fundamental and must be developed more fully, in connection with an improved theory of abstract ideas. And then certain respects (as relevant to science) to exclusion of others. This is a live problem still today. Also might allow for concomitant variations. A misgiving that might from a *too* empiricistic view arise in connection with *any* such definition of cause is its reference to absent (past and *future*) pairs of resemblants. Hume notes this and so proposes an alternative (p. 170, middle): <"CAUSE is an object precedent and contiguous to another, and so united with it, that the idea of the one determines the mind to form the idea of the other, and the impression of the one to form a more lively idea of the other."> Odd—determining cause by introspection. Almost "cause if I think it's a cause." The previous approach suits our usual attitudes better: never *know* it's a cause (past, present, future), but *conjecture* from samples.

Problems of cause and prediction can be separated:

- (1) Cause—as a necessary connection, in some irreducible sense, vs. mere concomitance (constant conjunction).

Hume has elected the latter. But

- (2) Prediction might still be hoped for: observations of past *constant conjunctions* giving us *grounds* for rightly expecting continuation of same (without thought of necessary connection in any further, irreducible sense).

Hume has explained our prediction psychologically: "Why we expect," in terms of association. But question remains: is it in *fact*, apart from our psychological vagaries, more likely than otherwise that the conjunction will continue? Is our psychological propensity *just*? This is the problem of prediction, or *induction*—a central problem of philosophy. Hume's answer is no.

—To here July 15

Uniformity of nature p. 89: <"If reason determin'd us, it wou'd proceed upon the principle, *that instances, of which we have had no experience, must*

resemble those, of which we have had experience, and that the course of nature continues always uniformly the same.">—(the traditional principle by which to try to justify induction, prediction) is sheer postulate. P. 90: <"probability is founded on the presumption of a resemblance betwixt those objects, of which we have had experience, and those, of which we have had none; and therefore 'tis impossible this presumption can arise from probability."> Reappears p. 104. In a sense it has a ground, viz., observed always true; but this would be circular as argument for induction. In effect: induction proved true by induction. (p. 91 and again p. 105.) It works, yes; or has; but that is irrelevant, unless we *assume* causality in order to infer from this past success that the success will continue. Hume's negative doctrine is inevitable, I think, in any thoroughgoing empiricism; and it does not depend on the more extreme or questionable features of his particular underlying system of elements and psychology.

This is not to say that we cannot develop refined techniques of induction. Modern statistical theory. Essence of explanation of why we can do this is on p. 105: we use induction to establish induction. And we thus inductively arrive at a science of statistical method. Hume uses this principle to counter the objection entertained p. 107, viz., that <if "belief is nothing but a more forcible and vivid conception of an idea; it shou'd follow, that that action of the mind may not only be deriv'd from the relation of cause and effect, but also from those of contiguity and resemblance."> Actually, as Hume points out by various examples, we often *do* believe or expect because of such "accidental" associations. E.g., p. 113, *credulity* as over-sensitivity to resemblance; p. 114, *impiety* as over-dependence on resemblance. Such was Locke's *derogatory* use of "association." And note "extravagance," p. 120. But better thinkers *don't*; because better thinkers are the ones who have done a more thorough job of *applying induction to induction*—i.e., empirically arriving at scientific method. Thus, for Hume the *primitive* form of induction is *simple induction*: sheer accumulation of instances (top p. 131). Simple conditioning, as observed and measured in experiments on rats (Skinner)⁶⁷. Last section of the part, pp. 176–78, makes the animal point. Accounted for very well by Hume's mechanism of vividness through association by contiguity and resemblance. This much is all at the unreflective stage (p. 133; we might say "*unreasoned*"?).

67. <B. F. Skinner (1904–1990) published a series of papers involving experimentation on rats from 1930, some of which formed the basis of his *The Behavior of Organisms; an Experimental Analysis* (New York: D. Appleton-Century Co., 1938). Skinner was a Junior Fellow with Quine at Harvard in the 1930s.>

But *contrariety* drives us farther (p. 133; p. 150). Contrariety makes us give up one hypothesis, and we shift to a second best, which we yet don't believe strongly enough (not vivid) because too few instances. So we try to get *more* for it, by experiment (manufacture of impressions) (p. 138). Sometimes (as recognized p. 104, foot) a *single* experiment may decide us—but this is because of indirect implication *via rules of scientific method*, together with other more specific *scientific beliefs*, already established *and knit together (for greater vividness) in a system of "physical objects"* (p. 108).

These general rules of scientific method are *themselves* established by induction, or derived from general principles (e.g., uniformity of nature) so established. They include all the refinements of statistical method. Familiar ones, less refined than some, are *Mill's methods*—the now classical ways of discovering causal connections (though Mill was ninety-five years younger than Hume):

- (a) Method of agreement: if given event accompanies a, b, c, d, . . . and a, e, f, g, . . . , then *a* is the cause.—See Hume, paragraph 5, p. 174!
- (b) Method of difference: if given event accompanies a, b, c, d, . . . and not b, c, d, . . . , then *a* is the cause.—See Hume, paragraph 6, p. 174!
- (c) Method of concomitant variations: Same as paragraph 7, p. 174!⁶⁸

Hume doesn't say why these are right; *impossible to defend*. But he says why *believed*. In each case [observe], association by contiguity and resemblance. And the rules are believed *by induction* because they work, or because derived from principles (uniformity of nature, etc.) which induction establishes—ultimately by sheer undeliberated conditioning. In this way we may, by scientific method, be led to so-called beliefs which *don't* correspond to greater vividness (p. 141) in any immediately perceived sense of impression-of-reflexion "necessity." Perhaps an extension of "belief," definitionally extended through rules of scientific method (which ultimately rest on beliefs in the direct sense).

—To here July 17

Hume's point of view on scientific method is startlingly summed up p. 149: <"all reasonings are nothing but the effects of custom; and custom has no influence, but by invivening the imagination, and giving us a strong conception of any object."> And the best summary of the central argu-

68. <John Stuart Mill (1806–1873), *A System of Logic, Ratiocinative and Inductive*, 3.8.1–6.>

ment is in bottom paragraph p. 166. <“Before we are reconcil’d to this doctrine, how often must we repeat to ourselves, *that* the simple view of any two objects or actions, however related, can never give us any idea of power, or of a connexion betwixt them: *that* this idea arises from the repetition of their union: *that* the repetition neither discovers nor causes any thing in the objects, but has an influence only on the mind, by that customary transition it produces: *that* this customary transition is, therefore, the same with the power and necessity; which are consequently qualities of perceptions, not of objects, and are internally felt by the soul, and not perceived externally in bodies?”> So probability is wholly subjective. The only things we can *know*, apart from immediate perceptions, would seem to be the demonstrative sciences of arithmetic and algebra (plus, as seen in *Enquiry*, geometry). But now in a curious argument, pp. 180–82, he repudiates even these things! For, we can err in calculation; we check, but this can err; we check the check, and so on in infinity; probabilities on probabilities, reducing to nothing.

This is precisely the argument which, as I pointed out in the first lecture, fails to allow for possibility of convergent product.⁶⁹ And this is just the case where we’d expect a convergent product, the superimposed probability being larger each time. Still, Hume has a point—if *none* of the probabilities are (except subjectively) any good *at all*, which is his position. A check of figures is at once an empirical judgment which can go wrong; we cannot get beyond *causal*, non-arithmetical support here. At best, trusting memory (which is itself of the nature of causal inference, though Hume doesn’t cover this point very well in his analysis of causality). We are driven to “animal faith” all along the line. Hume doesn’t say we *shouldn’t* trust our scientific methods. Merely that this is ultimate. Vain to ask for a more ultimate justification of anything than that *these methods themselves* support the thing in question. “Meaningless” to ask, as modern logical positivists phrase it.

After cause, the next great skeptical thesis in Hume concerns physical objects. But in rejecting our assurance of physical objects, unlike the case of cause, he was preceded by Berkeley. I’ve already touched on this—how Berkeley took the step for which Locke with his hesitating empiricism had prepared the way.⁷⁰ What Locke said of the subjectivity of the secondary qualities, Berkeley said of the so-called primary qualities as well; so that

69. <Above, 51–52.>

70. <Above, 66–67.>

there was no longer any empirical evidence for physical *objects* at all, except insofar as conceived as bundles of sense impressions (or constructs in some more devious sense—say by modern contextual definition). I remarked that various of Descartes’s followers had in part anticipated Hume in denying causality, viz., among material objects, and that they had done so with an ulterior theological purpose.⁷¹ The same was true of Berkeley’s denial of all matter: it secured God’s position as cause of our impressions, since impressions could no longer be attributed to the efficacy of matter. This seems a bit transparent, since any *problem* in accounting for impressions which is occasioned by abandoning matter is insofar an argument *against* abandoning matter, and evidence *for* it. However this may be, Hume’s skepticism with regard to material objects, like his skepticism with regard to cause, was altogether disinterested (see p. 409). Not skepticism to bolster belief, but skepticism as such.

Hume’s negative arguments are essentially Berkeley’s, and need not detain us. The essential point is that there is nothing but experience to go on, and experience gives us only the appearances, not an additional reality behind them. Here, as in causality, Hume doesn’t proclaim the *opposite* belief. In causality he is in *favor* of believing in predictions, etc., but it is animal faith, and vain to ask for ultimate evidence. Similarly as to bodies: <“Thus the sceptic still continues to reason and believe, even tho’ he asserts, that he cannot defend his reason by reason; and by the same rule he must assent to the principle concerning the existence of body, tho’ he cannot pretend by any arguments of philosophy to maintain its veracity. Nature has not left this to his choice, and has doubtless esteem’d it an affair of too great importance to be trusted to our uncertain reasonings and speculations. We may well ask, *What causes induce us to believe in the existence of body?* but ’tis in vain to ask, *Whether there be body or not?* That is a point, which we must take for granted in all our reasonings” (p. 187).> So the treatment of bodies, like that of cause, divides into two parts: (a) the negative part, arguing that there is no ultimate evidence for necessary connection (or physical objects), and (b) the positive part, explaining the derivation of our ideas of necessary connections (or of physical objects). So let’s turn to latter.

Purpose of hypothesis of external, continued existences is mentioned foot p. 198: simplifies our system of impressions, gives more coherence.

71. <Above, 91.>

Reduces multiplicity of entities. Particularly (as anticipated p. 108) it simplifies the causal network. Here is a *purpose*, then, but Hume must still say, in terms of his psychology, what the *mechanics* of arriving at ideas of *external continued existences* are. He says the “continued” is prior to the “external” (p. 199). His reasoning, as appears from succeeding pages, is this:

1. Successive resembling impressions are welded into a so-called *enduring* single object. Even others not successive are welded to it under certain conditions.
2. This *identification* among temporally different impressions encounters difficulties when thought of as identity of the impressions themselves, so we “externalize”—i.e., we say the enduring object is something other than the impressions.

On 1: Why are *successive* wholly resembling impressions to be viewed differently *at all*—hence giving an *enduring* object instead of one momentary one? Because synchronous with changing *backgrounds*, foot p. 200f.: “time, in strict sense, implies succession, and when we apply its idea to any unchangeable object, ’tis only by a *fiction* of the imagination, by which the unchangeable object is suppos’d to participate of the changes of the co-existent objects.” Seems a little roundabout on Hume’s part—saying it is a fiction of participation in change, instead of letting time in the first place rest on any changes of the co-existent objects themselves, without fiction. But a detail; let pass. Further on 1: the welding of such impressions into an enduring object leads Hume to some curious remarks on *identity*: for identity enters as relating these welded impressions, insofar as each is thought of *as* the fused object itself. Hume asks (p. 200) whence the idea of identity.

There ensues a confusion, similar to the confusions infesting identity right down to Wittgenstein. (a) “In that proposition, “*an object is the same with itself*, if the idea express’d by the word, *object*, were no ways distinguish’d from that meant by *itself*; we really shou’d mean nothing.” But that’s as it *should* be; we are not supposed to mean anything here, except a logical truth; it *should* be trivial. This doesn’t mean identity itself is trivial, though Hume seems to think so—see p. 201: “We cannot, in any propriety of speech, say, that an object is the same with itself, unless we mean, that the object existent at one time is the same with itself existent at another.” *Of course we can*. We can even say: “The first complex impression I ever

had containing blue was also the second I ever had containing yellow.” Same fallacy appears more strikingly in Wittgenstein: [explain].⁷²

Hume, in the former passage (p. 200) in which he complains of triviality of “an object is the same with itself” where identity is properly construed, apparently felt this wasn’t quite a convincing objection, for he added another: that furthermore the proposition would not “contain a subject and a predicate, which however are implied in this affirmation.” Three rebuttals:

- (i) *petitio principii*; why implied?
- (ii) can’t it have a subject and predicate which however happen to be the same?
- (iii) (the *real* one) a statement of identity ordinarily has two *different* terms (call them subject and predicate if you will), not to be confused with the designated objects (which are one).

—To here July 19

The true purpose of the identity concept, not appreciated by Hume and others, is as a means to linearizing statements which would otherwise run something like this (Figure 3.9).

We hark back, in the linear rendering, by saying “itself,” “the former,” “the same,” it, etc. And often it is necessary to point out only afterward, as in afterthought or because there wasn’t room for full provision in the running statement, that such-and-such further terms belong back at the same point.

Moreover, *descriptions* give rise to identities which are not even products of sheer meanings. E.g., Morning Star and Evening Star; but not necessarily just physical objects; also (cited before) “the first complex impression I have had containing blue that was also the second I ever had containing yellow.” Hume’s naïve sense of construction, forcing him to seek ideas as straight composition of copies of simple impressions, makes him less willing to accept this real notion of identity—perhaps—than if he knew about contextual definitions.

But not whole story, for it was one of the logical signs which *I assumed* in speaking of contextual definitions, and Hume would still be entitled to ask *whence*. *Not* contextually defined. He might ask same of logical negation (which he seems to assume uncritically) and of quantification (which he didn’t fully know about). But all these are legitimate questions. Actually

72. <Possibly Wittgenstein, *Tractatus*, 5.5303; cf. Quine, *Word and Object*, 117.>

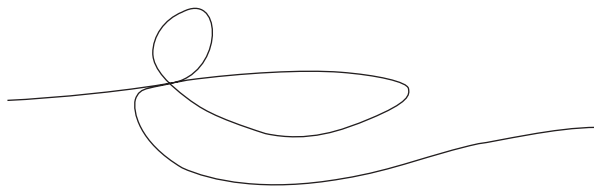


Figure 3.9.

identity *can* be defined, if we really have a catalogue of *all* our basic empirical predicates. [Explain.] This is probably the right way to handle the whole thing, if the catalogue can be supposed given. Question remains of explaining negation, quantification, conjunction; maybe we *have* to take these without further comment.

Anyway, let's see now what Hume does about accounting for idea of identity. His course is already determined by the line on p. 201 previously quoted: "We cannot, in any propriety of speech, say, that an object is the same with itself, unless we mean, that the object existent at one time is the same with itself existent at another." Not quite as though Hume makes identity consciously *mean* "duration," i.e., recurrence with full similarity (especially successively), but rather that this is the *source*, and that identity thus derived is confusedly taken as sameness out-and-out, i.e., identity properly so-called. Mechanism is this:

- (1) Suppose our *total* impressions, background and visceral and all, stay the same (or exactly repetitive) for a time. This *is* identity *sensu stricto*, for no time has passed!—by meaning of "time." So here identity is *properly* applicable.
- (2) Suppose next the background alone changes, giving us a measure of time. If a particular impression still stays unchanging (or exactly repetitive), the case is enough like (1) so that we *confuse* it with (1) and *impute* identity.

—To here July 22

- (3) Next natural shift—recurrence after interruption—still enough like situation (2) so that we repeat the practice also here, other things being equal. But this is the weakest stage of the three; we decline to identify here when in conflict with systematic inferences based more on (2) and causal reasoning—concluding then mere resemblance rather than identity.

The source of the *idea* of identity is evidently an impression of reflexion: the sense of the *uninterrupted progress of thought*. See p. 256: “it must be the uninterrupted progress of the thought, which constitutes the identity.” And, as Hume eloquently describes on pp. 255–261, this breaks down into association by resemblance, contiguity, and causation—as indeed it should. This fits our conception of physical objects well: unite the contiguous as parts of the object; the resemblant as stages of the objects; and cut off the contiguous beyond where resemblance (i.e., duration in time) reinforces it. And depend on causation (as seen) to decide interrupted cases. Hume points out how continuity operates strongly. It should, involving resemblance and contiguity. Strong enough to identify stages of river as *same* river despite even the known (inferred) change in substance.

In choosing this example perhaps he had Heracleitus in mind—who urged the fallacy of identifying, just as Hume does, and used the river as his example: can’t step into same twice. This effecting of strictly improper but natural identifications among our sense data, says Hume, is the extent of the vulgar view; no dichotomy yet between objects and impressions. Latter is made philosophically, when paradoxes arise; e.g., subjectivity of impressions (pushing one eye, etc.) without changes in “object” according to causal network. Interruption vs. continuance. And thus we have our physical objects. Thus attributing *interruption* to perceptions, *continuance* to objects (p. 215).

Modern empiricism could, compatibly with Hume’s *basic* orientation, depart from his view of physical objects in a radical way: Don’t attribute “enduring object” to *fallacious* identification of impressions; instead, get identity as *I* have explained; then externalize “enduring” object from the start as the object of this identity. Then what Hume analyzes as source of idea of *identity* (viz., uninterrupted progress of thought) is to be viewed rather as source of uniting as *manifestations* of one object. Identity of *this* object (or any) with self *of course* holds, and is explained in our new way. All of Hume’s observations about the rôle of contiguity, resemblance, and causation, and especially continuity, come now to figure in the explanation of how we form *things*, not identity. “Principle of individuation” becomes not a criterion for identity but a criterion for thinghood. Identification ceases to be a fallacy. Heracleitus becomes wrong: we *can* step into the same *river* twice, but in so doing we are not stepping into the same material substance; *river* is one kind of thing (a mode), *substance* another; *impression* a third. River and substance are two ways of combining

events (physical point-events or impression-events) into things. *Both* valid; things can overlap. The explicit accounting for the several external objects (over and above the above general characterization of our ways of limiting them) should proceed by some kind of logical construction *from* (not confusion *with*) impressions.

Russell undertook to sketch such a scheme in *Our Knowledge of the External World*. Begin with so-called *perspectives*, viz., whole fields of momentary experience, between which we can recognize part-similarities for correlation. Imagine many different perspectives (corresponding to different minds or “points of view”) simultaneously. Class of all the parts of all these corresponding to one given constituent impression of one of them is a *momentary-state-of-a-physical-object*. Can correlate such states in turn by taking neighboring sets of perspectives in time, and get *thing* in its temporal entirety—with adjustments based on causal laws (and complicating the definition) where there are questions of “right” identification after interruption.⁷³ One fault: intersubjectivity of starting point. Carnap tries to redo it avoiding this; also being more rigorous all along the line.⁷⁴

—To here July 24

But another fault remains still with Carnap: assumption of a logical apparatus that provides classes, classes of classes, etc.—anti-nominalistic. The desirable would be to get along without such machinery, by subtle use of contextual definition—not available to Hume (and not accomplished yet). Contextual definition would explain all new predicates, of kinds applicable to so-called physical objects, in application to variables or contents referring to so-called physical objects. But these whole contexts would be explained ultimately, as wholes, as abbreviations of long expressions made from our primitive vocabulary. (Probably impossible.) Identity, then—as a small detail—would be then extended also to physical objects, by a parallel to my basic definition of identity as applied to perceptions. The “vulgar” view would not enter as a stage; but Hume was unconvincing here anyway. Remains a general problem on the score of psychology *vs.* epistemology. Maybe Hume is nearer the truth genetically; but ours would be an account of how it *might as well be*—an account of how firmly we *can* reconstitute the world. If the proposed modern construction is impossible (as it may

73. <Russell, *Our Knowledge of the External World as a Field for Scientific Method in Philosophy* (London: George Allen and Unwin, 1926), 94ff.>

74. <Carnap, *Aufbau*.>

very well be) then from an empirical view there is *only* the genetic account: psychology of meaningless verbal behavior.

We now leave physical objects and turn to minds. Hume's attack on the belief in mental substance is like that on necessary connection and on physical substance: *no corresponding impression* (p. 233). It's his old argument on substance, as of early pages: most we can have is abstract idea embodying (and concretized by) the several observed states of the alleged substance. But the alleged mind in particular *self*, is peculiar in that it is not a substance in which the perceived qualities are supposed to inhere, but rather in which the perceptions themselves occur. Hume's usual form of argument is still available, though (p. 251): "If any impression gives rise to the idea of self, that impression must continue invariably the same, thro' the whole course of our lives; since self is suppos'd to exist after that manner [substance]. But there is no impression constant and invariable." So, just as there is properly no justifiable physical object but the sum-total of the impressions "of the object," so, Hume holds, there is properly no mind but the sum-total of *all* perceptions "by the mind." P. 253: "They are the successive perceptions only, that constitute the mind." There is no receptacle, no theater. The mind is the life, the sum total of experiences. In this, as in cause, Hume out-Berkeleyed Berkeley. Hume applied to the mind the same destructive considerations that Berkeley had applied to body.

As in the uniting principle of physical objects (or as Hume says, the basis of identity), so with the self, causality is the most important. More important even than memory, for we forget experiences, but still consider that they were *ours*. Continuity (reducing to resemblance and contiguity) and causality (reducing to same) are the main bases. Here, as with physical objects, we should speak perhaps of constitution rather than of a seemingly spurious "identification" of the distinct. But in either case we have a subversion of popular beliefs, insofar as the latter seem to involve something unchanging and "simple"—which is lacking either way.

Magnitude of the step can best be brought out by this sort of supposition: You have your choice: you and someone else are simultaneously going to be changed into two quite different people, different places, different appearances, different memories. One of the new incarnations is going to be ugly, hated, miserable; other is going to have everything good. Do you care which? By present doctrine you shouldn't!—for it means nothing to say it is you rather than he. No continuity or causation (nor even memory, which, however, Hume says, isn't the main thing anyway top p. 262). Popu-

lar feeling is that there is something *further* which could be said to be you or not. Hume says: find a justification in experience for *such* a conception.

Book 1 of *Treatise* is commonly considered most important part of Hume. *Enquiry* covers largely the same ground. Smoother, and in some respects more correct; because written years later; but less detailed and on the whole less provocative. Provides, however, a good *review* of book 1 of *Treatise*, as well as new matter here and there. And a good piece of literature. Its one hundred fifty pages divide into twelve sections, of which the first seven go over the arguments of the *Treatise* with regard to impressions and ideas, association, belief, causality. Apart from condensation and omission, the main novelty of these sections is the point already touched on in class regarding mathematical certainty:⁷⁵ beginning section 4, he relents toward geometry, as compared with his attitude in *Treatise*, and accords it the same good will that he previously had reserved for arithmetic and algebra. It is here that he introduces his contrast between *relations of ideas* and *matters of fact*—according certainty to mathematics on the ground that it concerns merely the relations of ideas and is therefore under our control—a point of view toward mathematics which, essentially, had been shared by Locke.

—To here July 26

I had recurred to Hume's remarks in the *Enquiry* on mathematical knowledge. How he relents toward geometry. The significant shift in the *Enquiry*, as regards philosophy of mathematics, is to the conception of *relations of ideas vs. matters of fact*. Former are under our control and afford certainty. It's not altogether clear how this notion fits with the rest of Hume's doctrine. Can't contrast ideas with matters of fact in the sense of mind vs. external world, for no external world is assumed. Only possible contrast is between complex ideas (voluntary combinations in imagination) and impressions (as they come, involuntarily). But this isn't a clear contrast between mathematical and empirical science, because voluntary compounding occurs in both; and ultimate dependence on simple impressions is required equally in both. The answer is suggested early in section 7: difference merely of precision. If we could develop other, non-mathematical ideas with more precision, perhaps much of the same certainty could be hoped for in other fields as in mathematics. And there is no reason why mathematics should have a monopoly of certainty. Mathematics has done

75. <Above, 86–88.>

better than other branches just because of (a) accident of having attained precision, or (b) easier topics for precision. This *is* mathematics; any additions to the realm of comparable precision would be *called* mathematics too.

This view harks back to Leibniz's program of a *universal characteristic*: analyze all our concepts precisely in terms of clear and distinct ideas; then, when a problem arises, *calculemus*.⁷⁶ This is probably right. It doesn't say we can know everything; what relates to the succession of impressions remains unknowable until experience reveals it. Leibniz himself recognized this distinction, in his distinction between truths of reason and truths of fact. What can be known with certainty (given precision enough in our ideas) is that which proceeds from their mere manner of composition—or, from the more elastic point of view of Tooke and of modern “logical construction:” that which proceeds logically from mere definition of terms. This version in terms of definition isn't accepted by Hume himself. He distinguishes, on next to last page of *Enquiry*, between statements true by definition (“*where there is no property, there can be no injustice*”—because injustice is defined as violation of property) and the certainties of mathematics (“*the square of the hypotenuse is equal to the squares of the other two sides*”).⁷⁷ But given more intricate conceptions of definition than Hume knew, perhaps those others (mathematics) reduce also to “analytic”—following from definition of terms. Surely so if for Hume they follow from relations of our ideas—if the definitions exhaust the ideas, embodying all connections not matters of fact. Even logical truth, not worried about by Hume, is on same footing. Only we finally get to a point where appeal strictly to *definition* seems hardly feasible—but still our certainties at that level may be in same spirit, viz., analytic in the sense of following from the ways we see fit to *use* our terms (badly called “implicit definition”).

A few other scattered points in sections 1 through 7 worthy of notice: One remark, end second page of section 2 (immediately after middle of section 2), capable of giving Hume's opponents a sense of indignation, and rightly: <“when we analyze our thoughts or ideas, however compounded or sublime, we always find that they resolve themselves into such simple ideas as were copied from a precedent feeling or sentiment. Even those

76. <E.g., G. W. Leibniz, “On the General Characteristic,” in idem, *Philosophical Papers and Letters*, ed. Leroy M. Loemker. 2nd ed. (Dordrecht: D. Reidel, 1969), 221–28.>

77. <Burt, ed., *The English Philosophers*, 688. All references to Hume's *Enquiry* are to this text, unless otherwise noted.>

ideas, which, at first view, seem the most wide of this origin, are found, upon a nearer scrutiny, to be derived from it.”> (p. 594). Perfectly easy to find plenty of ideas without corresponding impressions: resemblance etc., cause, existence, substance, probably identity, etc. Hume's normal approach is the other way around—*objecting* to an idea, rejecting or revising it, *because* it has no corresponding set of impressions. E.g., less than two pages farther on (end of section): <“When we entertain . . . any suspicion that a philosophical term is employed without any meaning or idea (as is but too frequent), we need but inquire, *from what impression is that supposed idea derived?* And if it be impossible to assign any, this will serve to confirm our suspicion.”> (pp. 595–96). This is the *inquisition* of which Reid spoke, as quoted by one of the commentators.⁷⁸ This pair of passages, two pages apart, a beautiful case of circularity. Actually of course the first passage is the one at fault; Hume's doctrine of ideas copying impressions has no *such* basis, but is his ultimate standard of meaningfulness: his empiricism.

Mid next section (3), surprising remark: “I do not find that any philosopher has attempted to enumerate or class all the principles of association.” Seems to overlook Aristotle, who speaks of this subject in terms of principles of *recollection*, in *On Memory and Reminiscence*; as I mentioned before, he is in two-thirds agreement with Hume: resemblance, contiguity, *contrast* (instead of causality).⁷⁹

On last page of section 4 [p. 617] an interesting analogy is drawn by Hume between his doctrine of induction and “pre-established harmony.” The allusion is to Leibniz. Leibniz, like Hume, was in trouble over relations. Though Hume was less ready to recognize his trouble. And Leibniz's trouble was for a different reason: not because he wanted his ultimate realities to be simple impressions, but because he wanted them to be simple souls. But the result was the same: no mechanism for connecting them. Leibniz faced it out (unlike Hume with his simple impressions). The problem was: how can souls act on one another, or perceive one another (in particular perceive matter, which after all is a congeries of small souls)? Answer: they can't! Hence no causation or interaction among souls for

78. <Laird, *Hume's Philosophy*, 25–26. Laird refers to Reid's comment in *An Inquiry into the Human Mind on the Principles of Common Sense*, section 8. See Reid (1710–1796), *Inquiry*, in *The Works of Thomas Reid, D. D.*, ed. Sir William Hamilton. 7th ed. 2 vols. (Edinburgh: MacLachlan and Stewart; London: Longman, Green, et al., 1872), 1:144.>

79. <Above, 74.>

Leibniz; but they are so created as to go on parallel, like watches synchronized. So this idea of preëstablished harmony is one of the famous ways of dealing with problem of cause or interaction.⁸⁰ Another is *occasionalism*, discussed by Hume near end of part 1 of section 7. Doctrine of Malebranche and other followers of Descartes. This is the doctrine mentioned earlier, whereby God is sole cause, and the so-called causes in nature are *occasions* for God to intervene. Thus different from preëstablished harmony.⁸¹

Since cause is so central to Hume, well to know of these celebrated alternatives. But in particular Hume's cited analogy between preëstablished harmony and his own doctrine of induction puts latter in an interesting light: as if:

—To here July 29

- (1) Nature *in fact* goes on in its normal causal ways, but we can't *know* it in any direct sense. (Does Hume believe it does? *Sure*; it's human nature to *believe* it!)
- (2) Meanwhile man goes on *believing* in these causal connections, without any access for proper knowledge, because of a *parallel* mechanism in the *mind*, so-called custom! God has preëstablished this harmony between each mind and nature (and therefore, by transitivity, also between the minds!) Quite unexpectedly Leibnizian.

Apart from the points I have touched on, the *Enquiry* in sections 1–7 and also 9 and 12 is mainly a repetition, in rephrased and condensed form, of book 1 of *Treatise*.

Section 8 repeats in effect pp. 399–412 of *Treatise* and adds more. It constitutes a fine little twenty-page essay on freedom of the will. This is a great traditional problem of philosophy. If the events of the world are determined by causal laws, how is the will free to change things? And, if not, how can there be praise and blame, ethics, reward, religion? Hume begins by arguing in support of causal laws, including human actions along with other phenomena. This may seem strange in view of Hume's skepticism over causality, but it needn't. Hume believes in the regularities of nature

80. <E.g., G. W. Leibniz, "Discourse on Metaphysics," and "Correspondence with Arnauld," in idem, *Philosophical Texts*, trans. Richard Francks and R. S. Woolhouse (Oxford: Oxford University Press, 1998).>

81. <Above, n. 14, and 98.>

and counsels us all to, but says they can't be established in any more ultimate terms. It is our nature to believe in them, and he goes on from there. No call to draw line in this respect between objects and persons. And he cites many instances to show that it is common sense to believe in causality among actions, or uniformity of nature including human nature. The unyielding quality of jailer and executioner is as causal as the unyielding quality of the prison walls and the axe; and as predictable. But there is still freedom of the will, in the sense of "*a power of acting or not acting, according to the determinations of the will.*" This fits with determinism; merely that the will is a link in the causal chain. But is determined by earlier links—as indeed is presumed popularly by predictions of action, by psychology of motives, by threat and incentive. Hume goes into the details of such causes of will in *Treatise*, pp. 413–18, and especially book 3. Will consider briefly later.⁸²

But, continuing now with section 8 of *Enquiry*: as Hume points out in an interesting footnote, there is a special reason why people are tempted to think the causal chain itself may lapse in human actions: frequent sense of indifference, indecision, *not* being motivated or caused. Yet when we *do* act, we are caused—even if trivially. Here we are feeling as caused things, not as observers of external caused things. Latter method always sustains determinacy, even when directed to ourselves; and former doesn't count—for a caused stone doesn't feel at all. But there is also *another* reason for wanting to think of the will as not only free (in the sense of an effective part of causal chain) but undetermined. This other reason is justification of praise and blame: for how can we praise or blame constrained volitions? Why doesn't all blame go back to God as first cause (if our actions *are* caused) and not to us? This raises the problem of evil: Why isn't God guilty? Hume mentions the optimistic solution: Polyanna, Leibniz, Émile Coué, and Mrs. Eddy. Mrs. Eddy's solution is that there is no evil. Leibniz's is that there is evil, but no more than logically necessary compatibly with greater goods; best of all *possible* worlds. The latter exonerates God, but doesn't explain still why *man* is to be blamed. (Mrs. Eddy's removes the whole problem, but is implausible.)⁸³

82. <Below, 124.>

83. <See, e.g., G. W. Leibniz, *Theodicy* (1706). Émile Coué (1857–1926) was a French psychotherapist whose views were fashionable in the 1920s. His pertinent works include *La Maîtrise de soi-même par l'autosuggestion consciente*. Nouv. ed. (Nancy: L'auteur; Paris:

Hume's answer is that praise and blame are instinctive and ultimate like induction. Mechanism of it—or of our moral standards—is subject of *Treatise*, book 3. Even if we know intellectually that a crime is part of a good whole (he goes on to say in *Enquiry*) we instinctively react against it—as we react with pain against the pulling of a tooth for our general good. Hume leaves it at this. One wonders incidentally how serious he is about his hypothesis that everything is right with regard to the whole.

This was a ticklish religious point which he didn't need to face in this particular connection. But we begin to see what he felt about it on turning to section 11, where he quotes his imaginary friend personifying Epicurus. For his friend argues, and very well, against theological arguments concerning the nature of this world of experience—arguments seeking to picture it as better than it is. Hume's supposititious friend begins by considering the *teleological argument* for existence of God—the argument from design. There is so much system and pattern in the world, surely it couldn't have come about by blind chance. This is the most popular argument down to present day. Another argument, considered earlier, was the *ontological*: existence follows from the definition of God. I have already criticized this; and we saw earlier that Hume rejected it.⁸⁴ And recurs in *Dialogues*.⁸⁵

A third well-known one is the *cosmological*—that of first cause. Everything must have a cause, therefore the totality of events in the world must as a whole have a first cause. It is defined as God. This is really a part of the teleological argument, in that the latter also takes God as cause, but argues further that this cause, God, must be intelligent and purposeful in view of the evidence of design in the world. Actually the cosmological argument itself has a conspicuous flaw. By the same reasoning that the totality of the world must have a cause (viz., that anything has a cause), we could ask in turn for the cause of God. If it is a satisfactory answer to say that God is

Oliven, 1923); in English translation, *Self Mastery Through Conscious Autosuggestion* (New York: American Library Service, 1922). Mary Baker Eddy (1821–1910), founder of the Christian Science movement, articulated her doctrine that there is no evil in *Science and Health with Key to the Scriptures* (Boston: Christian Scientist Publishing Co., 1875).>

84. <Above, 54–56 and 91–92, respectively.>

85. <Possibly Quine had in mind Hume, *Dialogues Concerning Natural Religion*, part 9, in Burt, ed., *The English Philosophers*, 734f.>

self-caused, or eternal and therefore uncaused, why not say this of the world itself in the first place?

The teleological argument, or argument from design, is free from this defect. It is—as Hume recognizes—a purely empirical argument, like any argument in natural science. We see that, in general, where there are marks of intricate organization there is a designing mind behind the thing. (Or do we? Surely most cases of intricate organization, from snow crystals to living animals, are not man made; to suppose design is to beg the question.) But suppose we do; this is, if justifiable at all, an empirical generalization. And existence of God is inferred from it as an empirical conjecture.

In the section, 11, under consideration, Hume stresses this empirical character, but declines to say whether he feels it is a *bad* empirical argument, *bad* science. Discussion of this topic is not needed for the point he is going to make regarding optimism. He reserves appraisal of the teleological argument to *Dialogues Concerning Natural Religion*, sheltering himself, there as in section 11, behind a fictitious anonymity. In section 11, for sake of argument, the teleological argument is accepted. And it is agreed that the Creator would have indeed to be ingenious and able—really inconceivably so—to produce so remarkable a world. But then how do people argue for optimism—that it is the best of all possible worlds, or that there are better things to come, rewards for the just, etc.? On the ground that so good a Deity, and so powerful a one, would not find it in him to leave all the imperfections and injustices which we seem to find in the world; he must have made adjustments, of a kind that it is not given us to see, but only to infer from God's attributes. And how did we infer God's attributes? From the nature of the observed world which he designed. How, then, are we warranted in arguing that God's attributes are really such as to lead him to make the world *better* than we observe it?—Here is the main point of section 11.

It is remarkable enough that God was able to make as ingenious a world as this one as we know it; and perhaps it is remarkable that God was benevolent enough to put as much good in the world as we can actually observe in it—though this point is more tenuous. But what ground for inferring that God is even more skillful and even more benevolent than needed for those already incredible achievements—and hence concluding that the world is really even better than it looks? The net content of such an argu-

ment is that, because the world looks as good as it does, it must be even better than it looks; and such reasoning is quite lacking in empirical basis, direct or indirect.⁸⁶

—*To here July 31*

The general tenor of section 11 is that it is a lost cause to try to establish religious tenets by reasoned argument, *a priori* or *inductive*. Section 10 on miracles takes the same line. It is the nature of a miracle to be—other things being equal—incredible; and we have *real* evidence for believing it only if the evidence in favor is so strong that it would be just as much of a miracle for that evidence to exist *without* the alleged miracle as with it. This doesn't in fact happen; testimony can be wrong unmiraculously—especially remote testimony, remote from us and/or from the fact reported.

All of this may be viewed as argument against religion. Actually it is argument against both *deductive* and *naturalistic arguments* for religion. It is argument that the only way to religion is a third, independent one: *faith*. And there are plenty of theologians who are in full agreement with Hume on this point. Hume says of divinity and theology, at very end of *Enquiry*: “its best and most solid foundation is *faith* and divine revelation.”⁸⁷

It is especially noteworthy in Hume's case that he is not being any rougher on religion than on science!—for science also rests ultimately on *faith* in inductive principles incapable of non-inductive substantiation. Faith which, in the latter case, has a mechanism consisting in *custom*; but this is no justification; anything has *some* mechanism. Pinning things on faith doesn't mean *rejection* by Hume. He is all in favor of induction, and

86. <Though there is no clear evidence of this in the primary portion of the manuscript, at about this time Quine must have assigned the second paper to his students, for the auxiliary portion includes a card indicating this topic and its due date of 16 August:

Due Aug. 16.

2d paper:

1000-word digest of the three points of view in *Dialogues Concerning Natural Religion*

(Philo, Cleanthes, Demea).

Note e.g., that there is much disingenuousness in Philo; debating technique; skip his first professed views—his mock piety—and get down to his real views.

Just a schedule of the *main* tenets of each man; and, so far as possible, brief support of each.>

87. <In Burt, ed., *The English Philosophers*, 689.>

admits to it; it is even his method in the empirical psychology that constitutes much of the *Treatise*.

So it doesn't mean rejection of religion. Does *he* accept it on faith just as he does induction? Could. Of course not quite free to say he doesn't. But he doesn't close his eyes to the fact that there *is* conflict between science and religion, anyway a natural religion that talks about miracles or otherwise makes counter-scientific statements about the world. Such was the religion of his day, and the fundamentalism of today, and much even of more moderated religions. There is such conflict, and, insofar, it is a conflict between faiths; not between faith and reason, or between faith and science, but between faith and faith; for scientific method itself goes back to faith of its own kind, in the form of acquiescence in the mechanism of custom.

So it is a small miracle, as Hume says (end section 10), when we allow religious faith to supervene over scientific faith in cases of conflict. Which is *right* it is, presumably, impossible to say except *from a scientific point of view* or *from a religious point of view*. We may prefer the one, insist on it, despise those who deviate from it in favor of the other; but there is no higher court. Such is Hume's skepticism.

As applied to science it is the doctrine of *positivism*: it is vain to seek certainty behind science (the problem I spoke of at the beginning of the course), vain to seek a foundation for science beyond science itself. Summed up at end of *Enquiry*: <“While we cannot give a satisfactory reason, why we believe, after a thousand experiments, that a stone will fall, or fire burn; can we ever satisfy ourselves concerning any determination, which we may form, with regard to the origin of worlds, and the situation of nature, from, and to eternity?”> [*Enquiry* pp. 687–88; cf. p. 689.]

Book 2 of *Treatise*: When Hume divided the impressions into those of sensation and those of reflection in book 1, he had little to say of latter; reserved to book 2. Now he elaborates his classification, insofar as concerns impressions of reflection, as follows (Figure 3.10):⁸⁸

Takes up the psychology of origins of impressions of reflection. Omits that of impressions of sensation because not mental enough for him; physical and physiological, insofar as there is anything to be said of origins.

Part 1 of book 2 devoted to one pair of impressions of reflection: passions of pride and humility. Opposites, but paired: same *object*: self. Since

88. <This tabular classification is written on Quine's copy of Hume's *Treatise*, 275; i.e., at the opening of Book 2.>

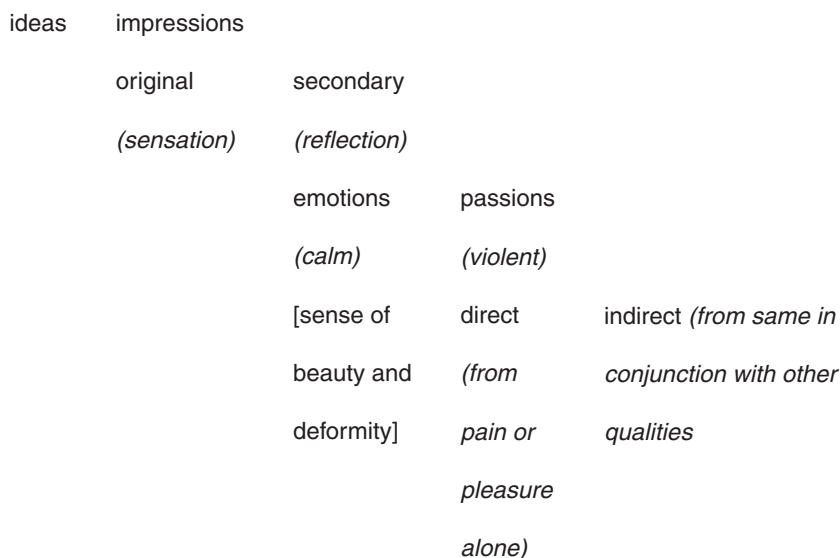


Figure 3.10.

pride differs from humility, *causes* differ; therefore cause is not *object*. (Also: if cause were self, pride and humility would be perpetual.) Relation to *object* is not of effect to cause, but a kind of association: ‘turning our view to.’ Foot p. 278: “the productive principle. . . . <e>xcites the passion, connected with it; and that passion, when excited, turns our view to another idea, which is that of self. . . . the one produces . . . [the passion], and the other is produc’d by it. The first idea . . . represents the *cause*, the second the *object* of the passion.” The pointing to self is “not only by a natural but also by an original property” (p. 280); i.e., ultimate, no further resolution; we merely observe that it always happens—like association of ideas.

It *isn’t* a *case* of old association of ideas; is it a sign of omission from earlier account thereof? This is not *association in imagination*, but evocation of an idea by an impression; but no excuse, because same is true of the regular association. Commonly: *impression* evokes an *idea* in *imagination* by resemblance, contiguity, etc. Here we evidently have, in pride → self, *another* ultimate principle of association, along with contiguity and resemblance? Similarly for humility. And similarly for many others, e.g., (Hume says) hunger → idea of food. Something different, all this, from what Hume calls association of ideas. More forceful and less controllable, per-

haps. ("Association" was a "gentle force.") It is an additional, irreducible factor in Hume's dynamics of perceptions, *besides* association.

Now to *causes*, as against *objects*, of pride and humility.

Causes: valuable and/or undesirable *qualities*.

Subjects: that in which those causative qualities inhere (own mind, body, possessions, etc.).

So object is *self*, i.e., total train of perceptions; causes are qualities, i.e., abstract ideas. No such things really, but we can easily see what Hume means: review his account of cause:

(instances of quality <i>a</i>)	(instances of pride)
a_1	p_1
a_2	p_2
$\vdots$	$\vdots$
a_n	p_n

By his literal terminology, each *instance* of the quality *a* causes pride, and they are *similar* to one another. The *respect of similarity* is quality *a*, if we admit qualities for the moment; and Hume *loosely* says now the *quality* is the cause. Not really even as loose as it looks, for Hume, because to speak of an abstract idea is *really* (he holds) to speak of any one of its instances. But it does mean that the causes coalesce with the subjects! Add to this a further complication in that the subjects aren't perceptions but extended objects (or self)? But for Hume these *are* same, and it is rather the so-called identity of an extended object with itself that he repudiates.

But, phrasing causes of pride and humility in terms of *qualities*, as Hume does, next note that not *all* qualities cause pride and humility. Only those which cause *pleasure* and/or *pain*, which are *sensations* (p. 285). And not all these, always, but only when inhering in a subject 'part of ourselves or nearly related' (Figure 3.11).

Now what of this relation between subject and object of pride? Sometimes *very* near: sheer identity. As where cause is virtue (vice for humility parallel), or beauty of own person (deformity for humility parallel). Or external: object is property, or a blood-relation, or a product of own workmanship, or native land, or a man whom the object physically resembles. Closer the relation, stronger or more likely the pride. Note that, from point of view of a rigorous formulation, no sharp line between case where object is identical with subject and case where not. For should we say the cause (quality) is elegance and the subject possessions, or the cause (quality) is

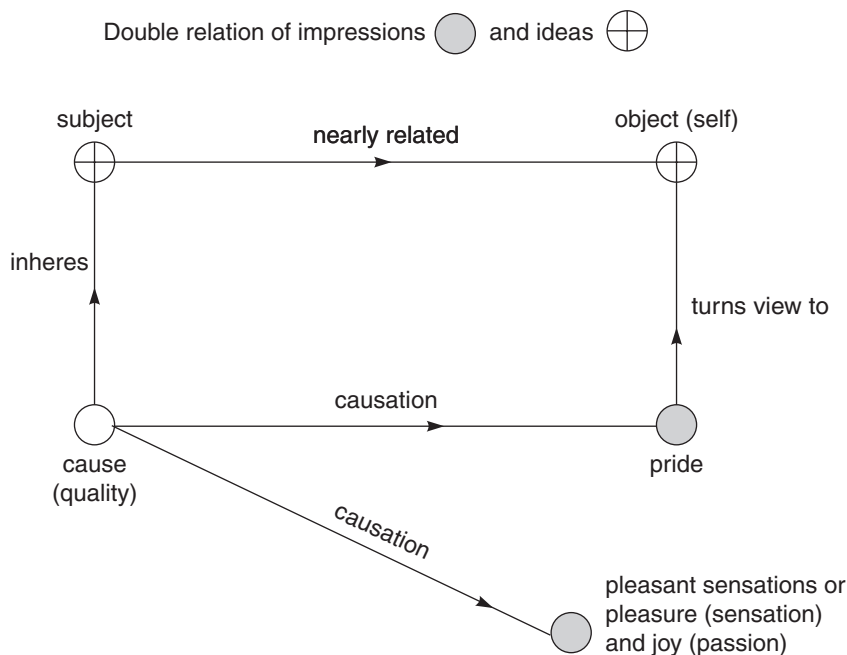


Figure 3.11.

wealth and the subject self? Or again: cause (quality) is refinement and beauty of city, etc., and subject is Boston, or subject is self (as object) and quality is that of being Bostonian? Any quality of an object related to *s* (self) is another quality of *s*, unless we have a standard of intrinsic vs. extrinsic—which Hume tacitly assumes. Common sense to assume it, but how formulate? Oh well.⁸⁹

How would it be in terms of sheer impressions and ideas, rather than qualities and enduring physical objects? [Re-doing above diagram.] Let's give special license to "self," otherwise can't do much (Figure 3.12).

89. <On a separate unnumbered card Quine included a précis of a twenty-minute presentation on Hume's doctrine of double relations given by a graduate student in the course, Robert Yost: "In theory of passions, Hume's 'double relation' seems to have two meanings: (1) p. 284, where he summarizes his basic principles, doubleness is: (a) passions related by similar affective tone, (b) causes of them related by inherence in same subject; (2) elsewhere, e.g., p. 289, where e.g., pride is made to involve doubleness, it is in (a) as before. (b) inherence of cause in subject closely related to self (to which passion 'points'). Not clear how pride can partake of (1). Strength of pride is attributed to doubleness—hence presumably (2); but (2) is not in his basic principles. Or does (1) equal (2) finally?">

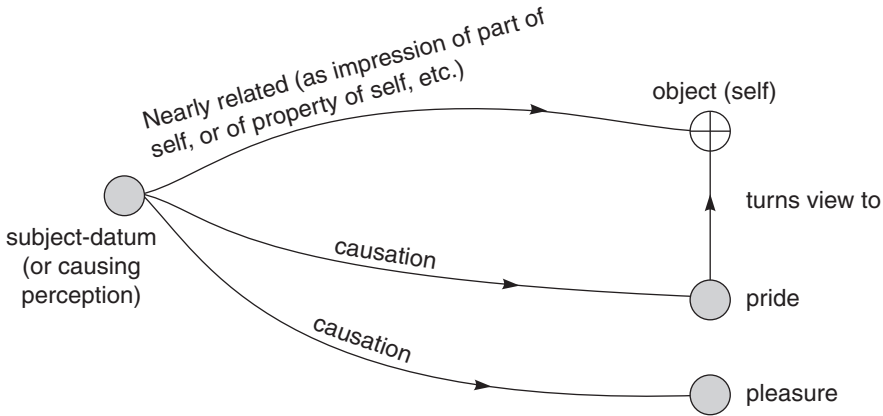


Figure 3.12.

How *close* this relation between subject and object of pride? Section 6:

- I. Closer than for joy. (Because joy is easier, having only a subject and not also an object to enliven.)
- II. Oligopoly on the subject. (Otherwise, perhaps, the *self* isn't limned so clearly; weaken association of ideas.) Hume seems to put insufficient emphasis on superiority—*invidious* comparisons—which is so much of the essence of pride. For his system it is rather incidental—merely a means to *accentuating* the *self-association* by method of differences or contrast from other persons.

—To here August 2

Last time, began considering respects in which relation between subject and object of pride (or humility) must be close. Saw two:

- I. Closer than for joy.
- II. Such that the subject (or other things having same quality, i.e., cause) bears the relation to few if any persons other than the self.

Hume mentions three more:

- III. Other people must enjoy the subject.
- IV. Connection must be firm and lasting.
- V. Connection gains by fitting general laws—i.e., by falling under an already conceived type which usually fits I–IV. (Rank, etc.)

Both IV and V are familiar factors in strengthening an association. But how does III work? Due to *sympathy*. Other person's enjoyment of object

enhances our own, for we *sympathize* and thus imbibe part of his pleasure to add to our own; thus one of the components of the pride-complex (viz., causing pleasure) is increased, increasing whole.

What *is* sympathy? How does it work? Manifestations of feelings in others are, for us, associated with the feelings themselves, because of past contiguity between such feelings and similar outward manifestations observed in our own bodies. So the feelings are vividly associated, therefore present in imagination (standard association theory of Hume). But then a special added event: *they are turned into impressions and felt* (p. 317). Here is a turning of the tables: impressions copying ideas. (But not in the first place, of course.) A mysterious transmutation. Hume had indeed provided for causing of impressions of reflection by ideas; but didn't lead us to expect it to be a copying. Just an ultimate phenomenon that Hume observes; no question of explanation needed. Seems to happen primarily or exclusively with passions; thus (p. 353) "Every lively idea is agreeable [odd], but especially that of a passion, because such an idea becomes a kind of a passion, and gives a more sensible agitation to the mind, than any other image or conception." Fits pretty well with rest of Hume's psychology. Difference between idea and impression is only vivacity. If idea of a passion becomes more lively, e.g., through association (by past uniform contiguity, or cause) with outward manifestations in another person, perhaps the increased liveliness of the idea will be enough to push it over into the impression category—just a matter of degree.

Why only passions, and not also sensations? Two possible answers why *not*, plus a third answer: that it *can* happen also to sensations.

- 1) Perhaps the manifestations of passions in others are a much better association than anything we can find for sensations. People don't manifest latter. Nearest they can come is describe in words. If anyone were artist enough with words maybe he could bring it off—and give us sensation of taste of a boiled potato by sheer description. But writers aren't that good.
- 2) (better) *Resistance* to letting an idea of a sensation get vivid to the point of being an impression, because of our systematization of impressions of sense into a physical world. There is no comparable systematization of passions, so no such resistance. And this brings us to—
- 3) Sometimes, if people don't set as much store by systematization as we do, an idea of sensation *does* make the grade and become an

impression! So-called hallucination, and insanity—disorganization of the systematization of impressions. Happens also in dreams. Sleep is a relaxation of our pursuit of system; a normal periodical insanity.

Hume should have allowed for copying of ideas into impressions *generally*, and at the beginning. Departing thus from doctrine of impressions of sensation always being caused outside; but not dangerously—because he can still say that the *first* event in any set of qualitatively like impression events *is* caused from outside; though another of the set *may* be caused from inside by copying the copy. This would rub out the main line dividing impressions of sensation from impressions of reflection, but so much the better. And it would make a virtue of what Hume, in book 1, felt to be a slight imperfection—the question of hallucinations as seeming to blur his conception of impressions vs. ideas.

So *sympathy* is the contagion of another's passion—through our associating so vividly to his manifestations as to bring the idea of the passion up to impression strength. But we can also get the effect without help of another's *manifestations*; but only seeing a causal *situation*. Example on p. 385, foot: man asleep, in danger of being trampled. No reason why this association by causality shouldn't also vivify idea to the point of impression of sorrow, if seeing another's manifestations can do it. Sympathy is needed, we have seen, in explaining pride and humility; viz., to explain why these are increased when other people enjoy (and/or suffer from) the *subject*.

Hume also makes much further use of sympathy. He uses it to explain the gregariousness of the human animal, p. 363: <“In all creatures, that prey not upon others, and are not agitated with violent passions, there appears a remarkable desire of company, which associates them together, without any advantages they can ever propose to reap from their union. This is still more conspicuous in man, as being the creature of the universe, who has the most ardent desire of society, and is fitted for it by the most advantages. . . . A perfect solitude is, perhaps, the greatest punishment we can suffer. Every pleasure languishes when enjoy'd a-part from company, and every pain becomes more cruel and intolerable.”>

Something wrong here: why pain more cruel alone? Why not less, since unaccentuated by sympathy with pain in others (supposing a generally unpleasant situation, not just a pain in the one person)? Recalls also the old statement p. 353, where he was also discussing gregariousness: “Every lively

idea is agreeable.” There is something in this; grief is more agreeable than depression, apathy. There is an agreeable quality to tragedy, because of vividness, despite disagreeableness of the content of the ideas. Maybe liveliness is itself, other things being equal, pleasant rather than neutral, and maybe this is ultimate; such is Hume’s notion.

Hume uses sympathy also as basis for some queer ideas on aesthetics: Back on pp. 298ff. his basic views on beauty appear: the instances of beauty are those visual qualities which cause pleasure. (Or: those perceptions, having visual qualities such that all perceptions having these visual qualities cause pleasure.) Similarly for deformity, in relation to pain.

—To here August 5

This much seems all right. But *how* does beauty give pleasure? E.g., functional beauty: comfort, convenience. But not *this* pleasure to *me*, if not mine—unless by *sympathy* (p. 364). Similarly personal beauty, largely health and vigor. Again by sympathy; indeed a case of functionality. Also there are other standards of beauty than functionality. Typically, *balance*. This he explains (p. 364) again by sympathy: conveys ideas of fall, harm, pain. What we would call *empathy*, not sympathy, not sharing feelings of owner of the beautiful or ugly object, but imagining feelings of *being* the object. (Likewise for personal beauty as health-functionality.) But this distinction needless here; Hume’s account covers all such cases under sympathy. Present case is type of man in fall; idea of pain is vivified to impression not by seeing manifestations of the feeling, but by seeing causes calculated to bring it about (trampling, toppling). But this is far-fetched. If we can vivify the idea of painful fall to an impression of pain by looking at a lopsided design, we surely ought to be able to get an *impression* of the taste of a boiled potato from a very middling verbal description. And what then of the beauty of balance—giving impression of pleasure of *not* toppling? But this is an impression that we don’t even get from *real* non-toppling. Hume’s sympathy is getting badly overworked.

Main topic of part 2 (of book 2) is love and hatred. Hume’s theory is that these are like pride and humility, but that the *object* is other than self. All the rest stands: subject, cause (quality), causing of pleasure, relation of subject and object (Figure 3.13).

This link between pleasure (or pleasant sensations) and pride should also be drawn into the old diagram. It is what makes for the *double relation* that Hume so often refers to (e.g., p. 333): “of ideas [subject] to the object

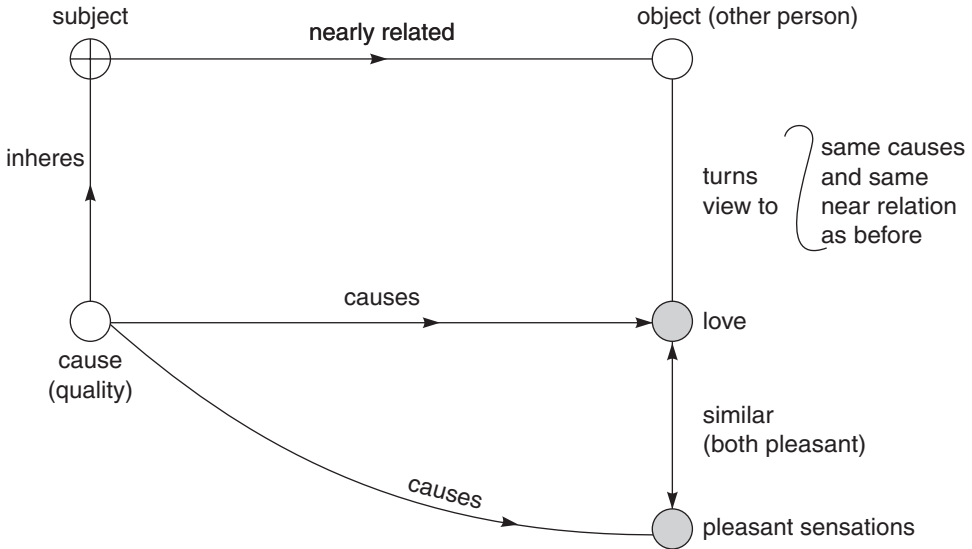


Figure 3.13.

of the passion, and sensations [pleasant] to the passion itself.” So, Hume says, pride, humility, love, hatred form a square (Figure 3.14).

Then there is *this* connection, also, between love and pride: love of a near relation (admiration, etc.) induces also pride, since the ‘turning view to object’ goes on easily to self. And the ‘near relation’ of subject to object is a fairly near relation to self. Hume says easier in this direction than opposite, because near relation etc. reminds us of (or ‘turns view to’) self more easily than vice versa. This is because mind passes more easily from remote to near than vice versa; likewise it passes more easily from less important to more important; perhaps because latter is intrinsically more vivid and therefore takes less associative power to make it vivid enough for awareness. “Whatever has the greatest influence is most taken notice of; and whatever is most taken notice of, presents itself most readily to the imagination” (p. 342).

Since love (and hatred) can vary as to object, a new consideration arises not operative in pride (and humility): love can transfer to a different object closely related to the first, since the needed *near relation* between subject and object will often hold almost as nearly to the new, related object (Figure 3.15).

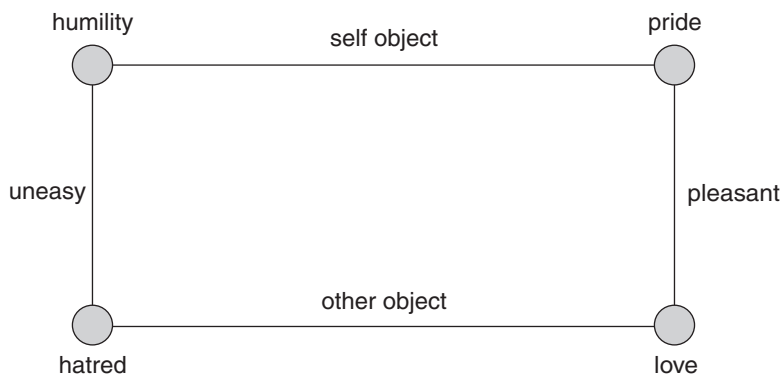


Figure 3.14.

And, since attention goes more easily to the near or important than to the far and unimportant, this transfer (*directing the view* onward to new object) should be easier from less to great than vice versa. Yet, says Hume, the reverse is the case: more easily love son on account of loving father; servant on account of master; etc., than vice versa. Reason is lame: p. 344: love of an important person is *ipso facto* the greater, so there is more of it to spread to lesser connections than vice versa. Why wouldn't (contrary to Hume) pride, for the same reason, diffuse itself into love?

Love and hatred carry with them another pair of passions: *benevolence* and *anger*. These are "originally" conjoined to love and hatred; no further explanation; ultimate. He includes various further descriptive details of love and hatred, mostly quite plausibly: That the reason we love more for deliberate benefits than for accidental ones, because evidence of a *lasting* disposition in the person to be beneficial. Item IV of the five features of the "narrow relation" as described for pride (and carried over for love). That we love relatives or old friends more than strangers because of durability; same IV again. And other points, of less philosophical than psychological interest.

Sympathy comes in again, as basis for our love of the rich and powerful; gives reasons to believe this more central than expectation of benefits. And sympathy operates directly, of course, to explain passion of *pity*. Indeed, same thing? No, a passion added to the sympathetically contrasted impression of sorrow. For, the same sympathetically imbibed impression can be accompanied by a passion opposite to pity, viz., *malice*. Like love and hate,

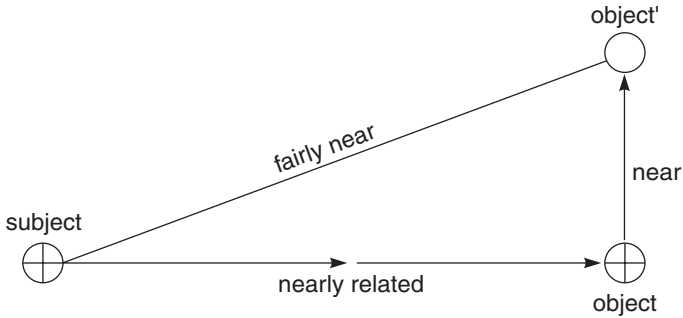


Figure 3.15.

pity and malice have subject and object; same structure; only the associated sensations are unpleasant. But in the case of malice, while the associated sensations are unpleasant, the quality of the passion itself is *pleasant*. Why this? And *envy* is in a like and opposite state: sensations pleasant, passion unpleasant. Essentially, explains them in terms of pride and humility, by a new principle of *relativity*: magnitude or strength or vividness of an impression is increased by contrast.

Hume's example: warm and cool hand, put in same water; water feels cool and warm respectively. Another example he might have cited: after-images, opposite to original. How does this explain malice? Our own pleasures seem *more* pleasant in contrast to the bad lot of the object of malice. Thence the pleasure investing malice. Increase in pride. And envy? Our own pleasures seem less pleasant (or pains greater) in contrast to the happy lot of the object of envy. Thence the pain investing envy. Increase in humility. Sometimes we get malice and envy, sometimes pity or sympathetic joy. Both opposite forces are at work; what decides? Their relative strength in the given case. Sympathy can be increased by vividness—sight of the suffering, extremeness of it, etc.; and the opposite—the factor of contrast with our state—can be increased by closeness of parallelism, facilitating the comparison and contrast.

—To here August 7

Review division,⁹⁰ for there is yet a further one: direct-indirect. Foregoing—pride, humility, love, hatred, pity, malice, envy, etc., were indirect: depended on relations over and above pain and pleasure-giving. Direct:

90. <Above, 114.>

desire, aversion; grief, joy; hope, fear. And the *will*: “*the internal impression we feel . . . when we knowingly give rise to any new motion of our body, or new perception of our mind*” (p. 399). *Internal impression*. Sensation or reflection? Latter, presumably; emotion or passion (calm or violent)? He hedges: “and tho’, properly speaking, it be not comprehended among the passions” (p. 399). This is subjunctive; not a flat denial. And on p. 438 he comes out with the affirmative: volition is a passion, and indeed a direct passion. And certainly volition is will, especially in view of foregoing definition of will.

Next he argues the freedom (or efficacy) of will *and* the fact that the will is caused—as already seen in *Enquiry*.⁹¹ And goes on to consider in detail *how* the will is caused. On this subject his important thesis is that man is *not* a rational animal in the sense of being guided primarily by reason; on the contrary, (1) “reason alone can never be a motive to any action of the will,” and (2) “it can never oppose passion in the direction of the will” (p. 413). Reason never enters except occasional or an engineering basis—finding means. “Reason is, and ought only to be the slave of the passions” (p. 415). Passion, he argues, is “original,” copying nothing in nature; therefore irrelevant to reason, and incapable of contradicting it—except in the case of a misconception of fact causing a passion. Only in this, and in engineering, is the reason relevant. We *think* reason guides actions more than it really does, because we confuse calm desires with reason. So study of causes of will becomes first a study of causes of direct passions generally. Hume begins with the violent and what turns a calm passion (or emotion?) into a violent one.

Passions readily unite; the stronger gaining added strength from the force of the weaker. “<A>ny attendant emotion is easily converted into the predominant.” Opposition and uncertainty increase passions. Ultimate observation? No explanation. “Puts spirits in agitation.” *Custom* has an *ultimate* tendency, other things being equal, to make us (a) do things more easily, (b) incline to things (cf. nostalgia). Contrary to uncertainty, spirits subside. *Surprise*, contrariwise, by force of uncertainty, augments passions generally, including unpleasant ones. Also surprise is *cæteris paribus* pleasant. (Because increased vividness is pleasant? And surprise is a sudden [ultimate] attraction of attention?) Another force that strengthens passions is *imagination*. For, passion depends on the contemplation of the subject in

91. <As on 108–109 above.>

	GOOD	EVIL
Direct passion from certain or probable	JOY	SORROW
Direct passion from uncertain	HOPE	FEAR
Direct passion from good or evil apart from <i>expecting</i> (respectively)	DESIRE	AVERSION

Figure 3.16.

idea (impressions of reflection, caused by ideas of subject), so naturally vivified by increase of this component. Hume doesn't avail himself of this obvious explanation, but asks merely if the effect of imagination on passions proceeds from the principle "*any attendant emotion is easily converted into the predominant*" (p. 424). In particular then contiguity in space and time, on the part of the subject, are a force for strengthening passions. For, these strengthen imagination. Ultimately, laws of association. How about "absence makes the heart grow fonder?" Hume would have a pat explanation here too: uncertainty. But looks like contradiction. Maybe explanation is: uncertainty on anything tends to heighten *all* present passions on whatever subject; but increased clarity of knowledge or imagination of the subject heightens particular passion toward that subject. This would be consistent, but there is some question how fully it fits experience.

Then (pp. 438ff.) to main lines again: *all passions are founded on pain and pleasure*. Their subjects are always goods and evils (construed, as Hume does, hedonistically). Mind has *original* tendency to seek pleasure and avoid pain (Figure 3.16).

These latter directly actuate the WILL.

Book 2 ends with a rather appendicial little section on curiosity, or love of truth. Says the search for truth is essentially the same, in its motivational structure, with *hunting*. The *chase*. This point of view is reflected interestingly in Portuguese: the word for research is *pesquisa*. Not hunting, exactly, but what it does suggest is *fishing*: *pesca*. Same in Spanish. (My Portuguese etymologist proposes an unrelated etymology, but I should like to suppose he is wrong.) Why like hunting? Pleasure consists in the action, attention, difficulty, uncertainty—yet at the same time there must be an *idea* of utility, though it be only token. Why pleasant? Uncertainty makes for more forcefulness of feelings generally; discovery makes for increased vividness of an idea; latter (said Hume earlier) is *cæteris paribus* pleasant; and pleas-

antness augmented by general agitating effect of previous or concurrent uncertainty. And maybe (Hume says little of all this) utility is needed in order that the uncertainty be serious and thus effective as agitating spirits and enlivening passions. Queer he doesn't mention pride as a factor.

Turning in book 3 to Moral distinctions—virtue vs. vice—Hume asks whether these are based on reason or passions. We know what his answer will be, having seen the corresponding answer to the question of the determinant of the will. Passions. In this he contradicts much of the tradition, including Locke, who comes in for anonymous mention p. 463: <“There has been an opinion very industriously propagated by certain philosophers, that morality is susceptible of demonstration; and tho' no one has ever been able to advance a single step in those demonstrations; yet 'tis taken for granted, that this science may be brought to an equal certainty with geometry or algebra.”> Locke had asserted (Essay, 1.3) that moral rules are capable of strict demonstration, but never got around to it, despite pressure e.g., of Molyneux: <“Writing to . . . <Locke> on the subject of the Essay in general . . . Molyneux observes: ‘One thing I must needs insist on to you, which is, that you would think of obliging the world with a treatise of morals, drawn up according to the hints you frequently give in your Essay, of their being demonstrable according to the mathematical method. . . .’ To which Locke replies, ‘Though by the view I had of moral ideas, while I was considering that subject, I thought I saw that morality might be demonstratively made out; yet whether I am able so to make it out is another question. . . . to show my readiness to obey your commands, I shall not decline the first leisure I can get, to employ some thoughts that way. . . .’ With this half-promise Mr. Molyneux was not content, but in a letter written shortly after again urges the philosopher to set about a system of ethics. ‘There remains only,’ he says, ‘that I again put you in mind of . . . Practica, or ethics: you cannot imagine what an earnest desire and expectation I have raised in those that are acquainted with your writings, by the hopes I have given them, from your promise of endeavour<i>ing something on that subject. Good sir, let me renew my requests to you therein; for believe me, sir, <'twill> be one of the most useful and glorious undertakings that can employ you. . . .’ Locke, however, after further solicitation from his friend, finally excused himself . . . for not entering upon the undertaking.”>⁹²

92. <Locke, *Works*, 1: 155, 156 n.>

But Hume argues it can't be done. Roughly his argument is that he sees no way it could be done. He breaks this argument into two parts:

- (1) Distinction between good and evil belongs to actions of mind in relation to external objects—obviously: willed action in world situations. But (p. 465) he is at a loss to find such a relation.

Lamely put—but of the essence of his really soundest arguments all down the line: necessary connection, substance, rational willing; that there is nothing that anyone can show, just declamation.

- (2) Even if there were such relations, and the distinction between good and evil were thus drawn by reason, how could good gain its *obligatory* character? This would still have to be irrational.

His point in brief is that *should* has to be irrational, because reason can only push it back: *should in order that*. But why *should* the latter in turn be achieved? Reason can only say: *in order that*. Infinite regress. We have to cut the series: an ultimate *duty* not dictated by reason must be the end of the line. Point of view regarding will, book 2, is exactly the same: reason can answer why do you *want* to do this? only by saying "Because it will bring that about." Why do you want that brought about? Same old regress. We must cut the sequence, saying "Just because I want to;" "just because I like it." Sheer taste, finally, that which reason serves as an engineer. And similarly for right and wrong: non-rational moral sense, finally, which reason serves as engineer.

Case is parallel even in Hume's doctrine of cause and induction. Scientific method can answer "Why is such-and-such a method of inquiry or statistics sound?" "Because it has always worked;" or, "Follows from such and such general principle—uniformity of nature, etc.—which is attested by experience." Asked how past experience can give us any assurance that this method will continue to work, or that the uniformity of nature will continue, scientific method can only say: "Generally things long observed to work have continued." Same question can be asked again. We must cut the series, saying "I've become used to expecting this sort of thing." Sheer custom. Reason in science goes back to *custom*, which is ultimate. Reason in volition goes back to *taste*. Reason in ethics goes back to *moral sense*.

This attitude of irrationalism regarding the basis of will and virtue, Hume seems to have got from Francis Hutcheson; and Kemp Smith argues pretty convincingly that this part of the *Treatise* was done before book 1.

He argues that book 1, Hume's significant and revolutionary theory of knowledge, came as an extension of Hutcheson's theory of value: theory that all value depends finally on feeling, not reason.⁹³ Now what is this moral sense? "To have the sense of virtue, is nothing but to *feel* a satisfaction of a particular kind from the contemplation of a character" (p. 471). It is a form of pleasure. Good music, Hume says, affords one kind of pleasure, and is called *harmonious*; good wine affords a second kind, and is called *savory*; and good character affords a third kind, and is called *virtuous*. Three kinds of pleasure among many. Actually, even music may afford more than one kind of pleasure, being both harmonious and lively; and wine may afford more than one kind, being both savory and intoxicating; and likewise a character may afford more than one kind, being both virtuous and amusing, or picturesque, romantic, etc. So he must single out what pleasure it is in particular which is produced by contemplation of a character insofar as it is *virtuous*. Can't single it out except by pointing to things characteristic of it: the peculiar kind which makes us *praise*; and vice *condemn*.

When we praise or blame, we refer only to *motives*, not *actions* (p. 477). Therefore those aspects of character which admit of *virtue* (and *vice*)—which excite that *kind* of pleasure (and pain) in us—are *motives*. Therefore "no action can be virtuous . . . unless there be . . . some motive to produce it, distinct from the sense of its morality" (p. 479). Why distinct? Otherwise circularity: sense of its morality is sense of the morality of the motive. If the motive *makes* the action moral, the morality can't underlie the motive.

—To here August <9>

This last quotation was from p. 479—beginning part 2. But now, to see the mechanism of the moral sense at its clearest, let's skip to part 3; afterward returning to part 2, which, dealing with "artificial" virtues, is less direct. Saw that the traits of character that can be called *virtuous* or *vicious* (a) give pleasure or pain, but also (b) have to do with *motives*. *Actions* are relevant to moral judgments only as evidence of motives (p. 575). Examples of virtues (p. 578): meekness, beneficence, charity, generosity, clemency, moderation. The *actions* issuing from such virtues are such as to give us pleasure;

93. <Smith, *The Philosophy of David Hume*, 12–51. Francis Hutcheson (1694–1746), *An Inquiry into the Original of Our Ideas of Beauty and Virtue* (London, 1725); and *An Essay on the Nature and Conduct of the Passions, with Illustrations on the Moral Sense* (London, 1728).>

but the special meaning of *virtue* is that it is the possession of *motives* for such actions; such motives, evincing an enduring disposition, serve to make us value the person's character as a lasting source of these benefits. The pleasures coming from the actions themselves are of two kinds: *direct* (e.g., generosity and clemency work to our gain when directed on us) and *sympathetic*: pleasure from the pleasure which these bestow on other people when directed on them. Our pleasure is greater, of course, when we are the direct beneficiary; and greater, through sympathy, when beneficiary other than self is near, or loved; and greater when agent has power to *act* on his motives than when he hasn't. But the *virtue* is the *same*; and insofar as our pleasure is to be classified as *moral sense*—approbation—it is the same. But this is only because of the meaning of “virtue:” general traits of character: *that part* of our valuing of a person which corresponds to rules of motivation of his actions. (Such is trend of pp. 581–85.) Differences of pleasure in other respects are just non-moral; moral sense is, by the mere meaning of the word, marked off in a more inclusive sense.

—To here August 12

Case is same (as Hume says, p. 582) with *beauty*: the functionality which often underlies beauty gives more pleasure when we are the direct beneficiary, or someone with whom we have strong sympathy, than otherwise; indeed the pleasure afforded by the beautiful object is, insofar, greater; but the object is said to be equally beautiful regardless—as if what we call *sense of beauty* were, by the usage of the word, limited to *that part* of our pleasure which remains common under such changes of direct beneficiaries. Except, of course—which Hume omits to mention—the identity of the beneficiary *in relation to agent* (not in relation to me as judge of virtue). Thus a man is the more magnanimous for helping someone who has injured him.

Sense of beauty and sense of virtue are alike in this abstracting from special circumstances. The residue rests on recognition of *causal efficacy* by general rules, and not counting in special additional circumstances such as identity of beneficiary (friend or foe) or actual applicability (whether people are *allowed* to sit in the functionally beautiful armchair; whether the virtuous man is able to execute his good purposes [p. 585]). All that is said of virtue of course applies equally and oppositely to vice: vicious motives lead to actions giving pain, either by direct action or sympathetically by giving pain to someone else. But, insofar as vicious, considered in abstrac-

tion from identity of victim (except in relation to agent—e.g., matricide). Also, pain and pleasure must here be taken algebraically: lessening of pain counts as pleasure, and makes for virtuousness of a motive; and vice versa.

And now Hume takes up the question of *pride and humility* from the moral standpoint; previously his concern had been merely with them as passions, respectively pleasant and unpleasant, but without regard to right and wrong. Now why is humility a virtue—or, let us say, pride a vice? I.e., why does pride (on the part of others) lead to actions painful to us or others? Because, as Hume stressed in discussing passions, our impressions are relative: heightened by contrast, diminished by lack of contrast. (After-image.) Manifestation of pride by another lowers, by contrast, our self-esteem; he leads our attention to the alleged good points of himself, and by contrast our own good points are less vivid to us. So *we* suffer pain of humility. (Not *only* by contrast, but also because our pride depended also on few people *sharing* the thing that caused our pride; one of Hume's old four points on the near relation between subject and object of pride.) Pride in others is vicious, because painful to us, because conducive to humility in us, which is a painful *passion*. But two moral precepts come out of this analysis: (1) Pride, when backed up by genuine possession of the qualities of which one is proud, is less vicious, because our humility-pain may be partly offset by *love* of the person for those qualities, or *sympathy* with his being so fortunate. But not entirely; still room for envy. (2) Pride *concealed* is no vice at all, because what we don't know won't hurt us. So the only humility which is a virtue, by *these* particular considerations, is that to which we are enjoined by good breeding.

So far, a virtue is a pattern of motivation which, in others, causes pleasure in us; algebraically; directly or sympathetically; and a vice, pain. But, Hume points out (pp. 587–90 and 606–614), even this is broader than customary usage, which distinguishes between moral virtues and *natural abilities*: prudence, temperance, frugality, industry, assiduity, enterprize, dexterity. This subsidiary distinction can be drawn in this way, mainly: the natural abilities give pleasure directly to him who has them; and to us by sympathy with him; while the moral virtues give pleasure directly to others than him who has them—and thus to us directly or by sympathy with them. But this is not quite coincident with the common classification, because e.g., assiduity or dexterity on the part of another may be *also directly* to our benefit. And there is one so-called natural ability, *wit*, which *primarily and directly* gives pleasure to others than him who has the ability.

Hume's main conclusion is that the division between moral virtues and natural abilities is a fuzzy and inessential one. As to "good or ill desert of virtue or vice" (p. 591), the mechanism is via love and hatred. Hume's treatment of the latter passions provided that virtue is a cause of love and vice of hatred (because of causing pleasure and pain); and that wanting to confer pleasure on object of love and pain on object of hatred is "original" with those passions.

Having done with virtues and vices of the basic kind for Hume, called "natural," let's turn back to part 2 to consider others which present additional problem. The prototype is the virtue of *justice*, as Hume calls it, or *equity*—in the special sense of respect for property. What *virtuous motive* for an honest act? Public interest? Too remote. Benefit of individual object of the act? No, because honesty differs from generosity and benevolence. Solution: no "natural" motive for equity, no simple passion. Equity is one of the *artificial* virtues, couched in social custom, learned from our fellows. An imposed norm. Natural only in that a society is a natural phenomenon; artificial in that it is not instinctive, but taught to us.

Following in its main lines the social contract theory of Thomas Hobbes (a century before Hume), and Locke, he accounts for the morality of honesty as an institution developed by society for mutual benefit—protection of the property of each. So it is a case of engineering: reason enters as a means, to serve the general end of security of property. The means devised is "a convention enter'd into by all the members of the society" (p. 489). So, "*'tis only from the selfishness and confin'd generosity of men, along with the scanty provision nature has made for his wants, that justice derives its origin*" (p. 495). Among people benevolently disposed to one another, or in plenty, or both, justice is no virtue.

How do the senses of virtue and vice come to attach to these artificial or conventional norms? Here Hume invokes (p. 499) a general statement on virtue and vice: "every thing, which gives uneasiness in human actions, upon the general survey, is call'd Vice, and whatever produces satisfaction, in the same manner, is denominated Virtue." We sense the uneasiness, in the case of injustice, by *sympathy*. Including sympathy in other people's estimate of us. And this latter is the channel through which the *artifice* works: making people's awareness more lively by indoctrination.

Idea of justice is inseparable from that of *property*. Inculcation of the ideal of justice is in order *that property be stable*, for everyone's average gain. When idea of property first developed in pre-history, idea probably

was that everyone keep what was already in his physical possession. Because *custom makes us fond of things*; harder to relinquish than not to have had; therefore greater average happiness by this rule. For continuing administration of property institution, four main principles:

Occupation: first possession (discovery of island).

Prescription: long possession (whereat former claims lapse).

—To here August 14

Accession: adjuncts possession property $\supset$ property.

Succession: primogeniture, etc.

All these are cases of natural association of ideas. But what is (physical) possession (presupposed throughout)? Cause-effect; power to use, alter, destroy. Important refinement of property institution is *transfer by consent*. Basis of commerce. Reason: ecology; division of labor. Utility of this desire is multiplied by trading in *futures*, so as to distribute advantages not only of ecology and division of labor but also of *seasons*. And this needs a new auxiliary institution: *promise*. So, three artificially devised relations to reduce conflicts of self-seeking for collective benefit:

- (1) *ownership* (property), having its legal symbol in a *title*;
- (2) *transfer*, having its legal symbol in a *deed*;
- (3) *promise*, having its legal symbol in a *note*. *Money* is an alternative method.

What is a promise? Not a *willing to pay*, for it doesn't immediately precede the paying. Nor a *desire* to pay in the future, for we can promise while explicitly desiring the opposite. Nor a *resolution*, for this would carry no obligation—no morality of complying. He says it is a sheer fiction, *as if* 'willing an obligation' (p. 519). Willing obligation is temporally all right, since obligation does come right after. But absurd as real obligation, like willing a taste; for natural morality is decided by feeling. A fiction. So *fidelity*, like *justice*, is an artificial virtue arising through social convention as a prop for the institution of property. Or, to restate it in the jargon used a century later by Marx, these are bourgeois virtues.

These forces operate in both cases:

- (1) Reason shows us the advantage, to us and everyone, of general honesty (*justice* in sense of respect for property; *fidelity* in the sense of respect for property).

- (2) *Sympathy* with welfare of others makes us esteem these virtues also in ourselves, or in others not affecting our welfare.
- (3) "Artifices of politicians" and "education" reinforce those motives by inculcating general rules, accentuating peoples' awareness.
- (4) Resulting praise and blame strengthen our attitude again, through sympathy of attitudes.

Artificial on the score of (3). And this is essential, (a) because sympathy alone, with public in general, would be too weak to offset own interests; (b) because general rules would dissolve in favor of sympathy for special individuals (natural virtue of humanity). But the general rules are necessary to avoid chaos.

Actually the artifice of indoctrination, even, is insufficient, so we resort also to *government*: enforcement. This is (Hume says) a mechanism of substituting *present* evils (punishment) for the remote evils of deteriorating society—thus offsetting the tendency to discount remote goods and evils (the basic psychological law that the proximate is most vivid). There is a perhaps more essential need for government, or police, which Hume doesn't consider: even if we could take into consideration with full vividness all the remote social consequences of *our* act of dishonesty, including what little effect this *one* deviation might have on the deterioration of other peoples' honesty, it would commonly turn out that the harm to us, both direct and through sympathy with others, would be less than the gain. Like burning theater. Yet to each man's *greater* advantage for *everyone* to stay in line. So contribute to the police force. This consideration opposes the view that everyone would be moral if he could see all consequences; on the contrary, consequences themselves have to be adjusted to this end, and law and punishment are needed no matter how shrewd the people.

Out of this device there arises a *third* artificial virtue, along with *justice* (or honesty toward property) and *fidelity* (or honesty toward promises); viz., *loyalty* to the state. Because, *insofar* as government is needed for social ends just noted, and continues to serve those ends, general loyalty to it is in the public interest. But here again, as in the other two artificial virtues, education (or propaganda) and *policing* are needed—and for the same reasons. Note that there is *no* policing for *natural* virtues! So these three virtues (justice [to property], fidelity [to promises], loyalty [to state]) all go back to maintenance of property. A fourth artificial virtue is, in Hume's analysis, separate from the property concept viz., *chastity in women*; but

similar to the other three in being an artificial virtue inculcated for the good of society. The particular good in this case is care of the young, which connects with Hume's theory (in the *Passions*⁹⁴) of love of relations. This force, love of relations, toward making a man care for his children, is weakened if he is not sure they are his own. He might also have mentioned a connection with property, under succession (e.g., primogeniture)—this having a basis in an association of ideas (son to father) which is weakened if the blood relationship is in doubt.

Keynote of Hume's doctrine in theory of knowledge and ethics alike: Vain to seek a rational foundation. Ultimately we can only describe psychological behavior, not justify it. The basic tenet underlying all this can be phrased in Kant's words thus: *no synthetic a priori judgments are possible*. Kant's words, but not Kant's doctrine; Hume's doctrine, which Kant opposed. Doctrine that, apart from reports of immediate experience itself, the only judgments which are certain are those which follow merely from the meanings of the terms and have no factual content. For *there are no necessary connections between matters of fact*. This view was essentially present already in Locke's rejection of innate ideas; but it was left to Hume to take it in full seriousness and ruthlessly draw the consequences: the consequence that there is no rational basis for prediction, even probable prediction; no rational basis for scientific law, even probable law; and no rational basis for right and wrong.

Hume in turn didn't see the fundamental maxim clearly in terms of a distinction between analytic and synthetic. Unclear on the dependence on *meanings* of what he called "relations of ideas," as opposed to "matters of fact," in *Enquiry*. But it fits his point of view. All that is certain (apart from isolated immediate experience) is *analytic*. And this is the creed likewise of the empiricists of today. And it was in denying precisely this that Kant undertook to save science and ethics. Kant tried to provide for *a priori* (i.e., certain apart from experience) *synthetic* (i.e., non-analytic) judgments. This he did by his elaborate doctrine (which there is certainly no time for here) to the following effect: the mind imposes a certain general *mold* on experience, not merely by compounding ideas, but by sifting or conditioning the train of possible *impressions themselves*. The mind is *susceptible* only to *impressions* that conform to certain preconceived conditions or patterns; and the mind can *know*, by reflecting on itself, what those condi-

94. <Hume, "Dissertation on the Passions," in *The Philosophical Works*, 4: 156–57.>

tions are; and therefore the mind can know *something* about the course of any possible train of future impressions themselves; so there *are some* necessary connections among the *impressions*, knowable in advance; and so *not* all certain knowledge in advance of experience is analytic. And in this way, in particular, knowledge of *causality* or *uniformity of nature* was to be provided for; Hume's skepticism in this regard thus overcome, and science saved. A ridiculously obscure and difficult theory as set forth by Kant, and largely indefensible in its details; but ingenious in its general conception. However, its basic shortcoming is suggested in Kant's own words: his program was to show how *a priori* synthetic knowledge is *possible*. Not to show that we in fact have it, but how we might construct a *myth* coherently according to which man would have such knowledge if the myth were true. And clearly he can't do more.

So there has remained the Humean point of view, down into modern empiricism. In fact it develops into two divergent modern trends: modern *constructive empiricism*, and *pragmatism*. By no means exhaustive of Hume-inspired trends; nevertheless affords an illuminating contrast.

Constructive empiricism: explain all meaningful scientific discourse by contextual definition on the basis finally of reference to direct experience.

Pragmatism: abandon such a project as impossible, and say that our discourse is merely variously conditioned by experience without being wholly reducible to empirical terms. Abandon, therefore, empirical criticism of concepts; instead, judge any form of discourse in terms of its *utility*—this utility being measured within empirical science by ordinary empirical methods. Just as Hume didn't undertake to translate references to physical objects into discourse about experience (as a constructive empiricist would), but said it was psychologically a mere confusion, or false identification. But no hint that it lacks utility.

What is common to the Humean tradition, whether constructive empiricism or pragmatism or something else, is the negative: no metaphysical truth behind experience. While it is a skepticism, it is not a doctrine of despair and inactivity. The same old drive to science and induction exists, and is applauded; but it is a natural drive, its methodology is ultimate and irreducible to deductive logic, and the effort to find a formulation below and beyond science itself is vain and doomed to failure. Skepticism as a

counsel of despair and inaction exists in Hume only with regard to this latter point. Indeed, Hume points out quite eloquently that skepticism in this sense, far from being antithetical to science, is decidedly in the scientific spirit. *Enquiry*, foot p. 608: <“Nothing . . . can be more contrary than such a philosophy to the supine indolence of the mind, its rash arrogance, its lofty pretensions, and its superstitious credulity. Every passion is now mortified by it, except the love of truth; and that passion never is, nor can be, carried to too high a degree. It is surprising, therefore, that this philosophy, which, in almost every instance, must be harmless and innocent, should be the subject of so much groundless reproach and obloquy. But, perhaps, the very circumstance which renders it so innocent is what chiefly exposes it to the public hatred and resentment. By flattering no irregular passion, it gains few partisans: By opposing so many vices and follies, it raises to itself abundance of enemies, who stigmatize it as libertine, profane, and irreligious.”>⁹⁵

95. <Hume, *Enquiry*, in Burt, ed., *The English Philosophers*, 608.>

— 4 —

The Importance of Logic for Philosophy

Harvard Philosophy Club, December 2, 1947

The little group of skeptics who run this club have asked me to explain why in the world anybody who is intent on philosophy should allow himself to be sidetracked into taking a course in logic. I know better than to try to satisfy these people by suggesting that a logic course might help a student pass the prelims. For, being philosophers, these skeptics will pursue the issue to its fundamentals, and ask: Why should there be a preliminary exam in logic? The doubt implicit in such a question could never be a genuine doubt, of course; it must be a purely methodological doubt, a Cartesian doubt of the most fanciful order; but the inquisitors in question are quite capable, I believe, of Cartesian doubt.

This gives me no cause for complaint. If it is difficult and unnatural to doubt the value of logic, it should be correspondingly *easy* and *natural* to *establish* the value of logic. I'm grateful for having been assigned an easier task than any I might have taken on if left to my own devices.

To illustrate one important way in which logic contributes to philosophy, let me begin with an over-simple instance of philosophical confusion: "If an event has a cause at all, it must have *something* as its cause, rather than *nothing*, for *nothing* cannot be a cause. But if an event has no cause, *nothing* causes it. Therefore it is impossible for an event to have no cause". This, substantially, is the unsympathetic version which Hume gives to an argument in Locke. Whether it is an argument that Locke or any other sensible person could have intended, at least it will help me to begin schematizing one of the uses of logic.

This argument makes use of the fact that the word "nothing" behaves in ordinary language like a noun. "A causes the event" means that something,

viz., this A, is the cause. Analogously it is argued that “*nothing* causes the event” means that something, viz. this *nothing*, is the cause.

Something similar seems to have been going on in Heidegger’s mind when he asked “What about Nothing? What does Nothing do?” and decided that Nothing *nothings*, and that we have a special way of sensing *nothing*, this sense being *anguish*.

Such treatment of the word “nothing” can lead to ridiculous results, as Lewis Carroll has shown; and similar fallacies can be devised with “something” and “everything”.

Now, returning to the first example, let’s see how it fares in logical symbols:

Nothing causes x
 $(y) \sim (y \text{ causes } x)$

“Whatever y may be, y does not cause x ”. Here there is no term, no substantive corresponding to “nothing”, and so no opportunity to draw the false inferences that depend on handling “nothing” like a name. In logical notation the same thing happens also to the false substantives “everything” and “something”; they disappear.

Now let’s see what this sort of therapeutic translation really comes down to. A troublesome or misleading part of ordinary language has been translated into another part which is well-behaved. Another part still of ordinary language:

Whatever you may select, *it* does not cause x .

(Symbols are inessential.) All the serious and proper uses of “nothing” can be translated, sentence by sentence, into this well-behaved part of language. But the translation operates only on the sentences as wholes; the word “nothing” has *no* translation by itself; and hence this part of language contains no misleading substantive, simple or complex, corresponding to the misleading substantive “nothing”. Therefore the perplexities and fallacies attaching to the word “nothing” are by-products of a dispensable part of our language; and in this sense they may be said to be purely linguistic confusions.

The so-called *quantification* idiom used in modern logic is prompted by technical counteractions of efficient manipulation and deduction. Its efficacy in eliminating the fallacies of “nothing”, and the similar fallacies of “everything” and “something”, is incidental. But it illustrates the first point

which I want to make this evening: that logical analysis resolves certain confusions and philosophical perplexities by showing how to paraphrase language into peculiarly clear and straightforward subsections of language. It resolves the perplexities by showing them to be purely verbal: products of gratuitous idioms.

I don't maintain, as Wittgenstein does, that all philosophical problems are verbal in that sense. But some are. I've illustrated the point with a trivial type of example; a more serious example is afforded by Russell's theory of descriptions, in its bearing on perplexities of non-existence. The perplexity, briefly, is this: How can we consistently say there is no such thing as Pegasus, if there is no Pegasus about whom to say it? Or that there is no such thing as the king of France? How can any such statement of non-existence be true?

One answer sometimes resorted to is this: There are such things as Pegasus and the king of France after all; they merely happen not to be actual. Actuality, like redness, is merely one special quality of things.

Two reasons why some of us find this unsatisfactory: one, the recognition of a realm of unactualized possibles offends our taste for a tidy, economical ontology; two, the method still avails us nothing if we change the example and speak not of Pegasus or the king of France but of the round square. To accommodate such cases as this we should have, analogously, to recognize not only a realm of unactualized possibles but a realm of impossibles. Few of us, certainly, have so little taste for tidiness as not to be offended by an ontology of this kind. Various awkward problems arise, moreover, apart from mere questions of taste.

Russell goes back to the beginning of the trouble and protests that when we say there is no such thing as Pegasus, or the king of France, there need be no Pegasus or king of France about whom to say it. He first paraphrases all statements containing so-called singular descriptions: *the* king of France, *the* author of Waverley: idioms using 'the' in the singular. He does not paraphrase these descriptions themselves, but the statements as wholes which contain them; and the statements into which he paraphrases them contain such words as 'author' and 'Waverley', or 'king' and 'France', scattered through them, but no single phrase that can be equated with 'the king of France' or 'the author of Waverley'. Thus these translations are statements still about France and perhaps kingship, or perhaps Waverley and perhaps authorship, but there is nothing in them that would tempt us to say that they are about *the* king of France or *the* author of Waverley.

‘Pegasus’ can be handled similarly if we read it as short for a description: ‘the winged horse captured by Bellerophon’. So for all proper names.

The same process is at work here as in the treatment of words like ‘nothing’: paraphrase the *context* so that no single segment survives answerable to the troublesome substantive in question. The substantive ‘nothing’ itself received no translation, but simply disappeared on translation of the contexts as wholes; and similarly now for the substantive ‘Pegasus’ or ‘the king of France’ or ‘the author of Waverley’. The terms surviving in the translations are terms not giving rise to the perplexities that prompted the analysis; and those perplexities are thereby shown to be by-products of gratuitous, avoidable portions of language.

Logic is needed to guide such paraphrasing; and the supplier and subterfuge of our logic is, the farther it can carry us in philosophical enterprises of this type.

Now there is also another conspicuous domain of philosophy in which such logical paraphrasing plays a central rôle. This is in the empiricistic program of constructing the world from experience.

Hume’s notion was that all our ideas are composites of so-called simple ideas, each of which is a straight copy of some simple sense datum. Essentially this was also Locke’s point of view. According to this doctrine, every *meaningful word*—meaningful in the sense of expressing an *idea*—evidently ought to be translatable into a complex of words each of which, individually, names a simple sense impression.

But this is an intolerably strict standard of meaningfulness, as becomes apparent the minute you try to put it into practice at all. Whatever our own criterion of a simple impression may be, certainly a mere roster of names of simple impressions isn’t going to be enough. By merely putting such names together, and taking new words as abbreviations of the compounds, we can’t begin to build a serious language; and men like Hume and Locke can scarcely have thought that we could.

At the very least we are going to need various auxiliary connectives; and in fact we shall need much more.

Without trying to decide what Hume and Locke wanted, I expect we can agree among ourselves that the most the empiricist can *reasonably* demand is definition of words in *context*, sentence by sentence. We suppose a basic language which talks, in specified ways, about direct sense data; and then all further meaningful *sentences* are to be translatable *as wholes* into

sentences of that basic language. Individual words of the derivative language need not go over directly into complexes of words of the basic language. The word “nothing” did not go over into any continuous segment of quantification notation, but the sentences containing that word were translated systematically as wholes; similarly for descriptions; and the same may be expected when we undertake constructions upon a basic language of empiricism.

This being the case, there is full scope again for all the resources of the supplest and subtlest logic we can contrive; the better we are at logic, the farther we can hope to get in our constructions, and the more of everyday and scientific language we can hope to reduce to that *basic* language, whatever it may be, which empiricism allows us as a starting point. For, a surprising amount can be done through contextual definition if one has techniques for finding his way through the mazes.

This is strikingly brought out by reading Carnap’s rare and little known *Der logische Aufbau der Welt*. Although Carnap by no means succeeds in reducing the whole of everyday or scientific discourse to the very meager basic language of experience which he allows himself, he does get startlingly far along the way. Many of his constructions are complex and ingenious; and few of them could have been achieved without high proficiency in modern logic.

Much had been said in an empiricistic vein in recent centuries, but the *Aufbau* is so far as I know the first serious effort to put the program of epistemological reduction rigorously into effect. The problem couldn’t even have been seriously tackled without the benefit of modern advances in mathematical logic. As Professor Sheffer remarked an hour ago, the importance of logic to philosophy is simply the importance of basic experience.

Even so, there is reason to suppose that the program of straight reduction to a basic language of experience can never be carried through to completion—not even granted all the liberality of contextual definition and all the earnestness of modern logic. I expect that empiricists are bound, eventually—as most, including Carnap, already have done—to relax their demand of *translatability*, and to require only that every meaningful statement be *related* to the basic language, in some way short of out and out *translatability*.

Just what the basic language should be is itself a question, but this I shall

pass over. Another question is: what relations *should* we demand between the statements of science and the basic language, if full translatability is not to be insisted on?

One answer which readily suggests itself is this: any meaningful statement should at least have some statements of the basic language as consequences, or else be a consequence of some statements of the basic language.

But on closer scrutiny even this requirement begins to appear too strong. I won't here go into the technical reasons why; but merely say that the closer scrutiny in question is considerably aided by, if not dependent upon, methods of modern logic.

The most that can in general be demanded by way of connection is perhaps this: there must be some *degree of confirmation*—some degree of probability, in a sense—relating any meaningful statement to some statements of the basic language.

Then comes the problem of formulating what is meant by degree of confirmation. This is a vast problem, but, according to the considerations I have just sketched, it would seem to be an *essential* problem for empiricism; for, degree of confirmation seems to be the one surviving connection between sciences, or common sense, and the direct experience which is the empiricist's bedrock.

Carnap, Hempel, Helmer, Oppenheim, and others have been hard at work trying to formulate degree of confirmation, availing themselves liberally of the resources of modern logic. I for one am far from content with the results to date, but this much is clear: if a satisfactory formulation of confirmation is in the making, modern logic is playing an indispensable part in its creation; and if we are to decide rather, in time, that such an attempt is doomed to failure, any *conclusive* arguments to this negative effect will probably have to exploit the resources of modern logic in subtle ways that none of us have yet envisaged.

Logic figures no less importantly, and perhaps yet more obviously, in another central problem of philosophy which is closely connected with that of empiricistic reduction or confirmation; namely, the problem of the nature of *a priori*, or *analytic*, statements.

When we make explicit how to reduce statements to basic ones about experience, or how to confirm them, we thereby implicitly determine the *meanings* of those statements and of their constituent words—insofar as they may from an empiricistic point of view be said to have meanings at all. Empiricists and pragmatists from Peirce through Wittgenstein to Pro-

fessors Lewis and Bridgman have agreed, in varying idioms, that the meaning of a statement consists in the methods of its empirical confirmation. Now as by-products of these methods there will be some statements, notably Stevenson's old stand-by "No spinster is married," which are automatically true by virtue of the meanings alone—by virtue, in other words, of the mere methods of empirical confirmation themselves, no matter what the content of experience may happen to be. Such statements are, in Kant's terminology, analytic: true by virtue of the meanings alone.

All those statements which are theorems of pure logic, or instances of theorems of pure logic, are generally viewed as falling into the category of analytic statements; so it is not surprising that a training in logic should contribute to an understanding of analytic statements, if anything will.

In fact the development of modern logic has had an important bearing on Kant's doctrine of *a priori synthetic* judgments: the doctrine that not all statements which are true independently of experience are analytic. Kant felt that arithmetic was largely *a priori synthetic* in this sense, as contrasted with logic, which is *analytic*. But Frege, in 1879–1888,¹ showed that arithmetic was derivable from logic alone.

Kant might still maintain that logic in Frege's sense, being stronger than the traditional logic, goes beyond the bounds of the analytic. If so, however, it would be up to Kant to show just where Frege, in his logic, crossed the line. In any case the problem of delineating the bounds of the analytic becomes one which can hardly be profitably discussed any longer without reference to technicalities of modern logic.

The question of analytic or *a priori* truth is, in the terms of the Aristotelean tradition, very close to the question of *properties*, or *essential* attributes, as opposed to accidents. I mention this only to emphasize how central the topic is to philosophy. Actually I have no doubt that it is far better to think of these matters in terms of *a priori* or analytic truth of statements than in terms of essentiality of attributes. For, attributes are attributes of the *thing*, not adjuncts of their name; but whether something attaches *essentially* or *accidentally*, better *a priori* or *a posteriori*, depends not on the thing but on the particular name or description through which the thing is referred to. It is essential that the speaker at the Philosophy Club have the

1. While 1879 was the year of Frege's *Begriffsschrift*, it is unclear why Quine mentions 1888 here. He may have thought of Frege's *Grundlagen der Arithmetik*, which came in 1884. A more reasonable guess, however, is that he wanted to refer to Frege's *Grundgesetze der Arithmetik*, whose first volume appeared in 1893 and the second in 1903.

faculty of speech, but accidental that anyone of the name Quine have that faculty, as opposed to being a deaf mute; so it is strictly meaningless to speak of that faculty as essential or accidental to *me*, who am at once Quine and the speaker at the Philosophy Club. Here incidentally is a simple case of the kind of clarification which is the business of a proper theory of the *a priori*; and such a theory, if it is to be adequate, is bound to lean heavily on logic.

It's hard for me to talk for an hour without bringing up Gödel's theorem; and at this point I want to bring it up as an illustration of how logic, at its very subtlest, may reveal philosophical difficulties which nobody would otherwise have dreamed of.

Gödel's theorem has to do with the elementary arithmetic, and algebra, of whole numbers. What he establishes is roughly this. Consider any system of rules and principles for proving laws of arithmetic. Gödel shows that any such system, no matter how good or bad, is bound to be in one of these two predicaments: either 1) it enables us to prove a falsehood, so it is no good at all; or else 2) it is inadequate for proving certain truths of elementary arithmetic, even though those truths be capable of being written wholly in elementary arithmetical notation.

In a word, every technique for recognizing or generating elementary arithmetical laws is bound to be either inconsistent or incomplete.

From an empiricistic point of view, this is a baffling situation. The laws of arithmetic are not empirical hypotheses; to say they are true at all, then, must mean that they can be *proved*, rather than be verified by observation. Otherwise wherein can their truth consist?

Yet Gödel's argument shows that, no matter how good our technique, there must be some arithmetical truths that cannot be proved. His argument is mathematically rigorous and there is no escaping it. The empiricist could indeed throw up his hands and declare all arithmetic meaningless; but he is likelier to reconsider his empiricism.

It is to be expected, surely, that any statement which is *analytic* should be capable of proof. So maybe Gödel's result may be interpreted that Kant was right in viewing arithmetic as *a priori* but synthetic. Undeservedly right, for Kant could have had no justifiable inkling of the incompleteness of arithmetic.

But then, as I remarked earlier, we have to show where logic stops being analytic and becomes synthetic; for the fact remains that arithmetic is

translatable into modern logic without remainder. Logic itself, like arithmetic, is subject to the Gödel predicament of incompleteness.

Actually there *is* a line which can be drawn across modern logic, such that the logic to one side of the line does admit of complete systematization, and is not adequate to all of elementary arithmetic; while the logic on the other side of the line is incomplete, and contains implicitly elementary arithmetic and the rest of mathematics.

The line is just this: it is the line between that part of logic which is neutral on the question of universals, and the part of logic which implies a Platonistic answer to the question of universals.

The part of logic lying on the one side of the line neither affirms nor denies that there are abstract entities. The logic on the other side of the line assumes abstract entities and tells us things about them. Among these entities are numbers.

It seems natural enough to classify the Platonistic part of logic as *a priori* synthetic, rather than analytic, if true at all; for this removes the ontological problem of universals itself from the domain of problems whose answers are analytic.

If it seems unnatural to classify part of logic as not analytic, this may merely reflect an unwise extension of the term "logic" itself. As a matter of fact it would be equally in keeping with the *old* tradition to limit "logic" to the hither side of the ontological line I have described, and call the farther side *mathematics*.

These developments suggest an interesting managing of two traditionally quite distinct philosophical theses: the epistemological thesis of *empiricism*, which repudiates the *a priori* synthetic, and the ontological thesis of nominalism, which repudiates universal entities. A third repudiative thesis, less conspicuous traditionally, also appears to merge with them; viz., the repudiation of mathematics.

I have argued that the completable region of logic may conveniently be separated from an incomplete remainder of logico-mathematical science by drawing a boundary between the non-Platonistic and the Platonistic. My method thus far, in arguing that point, has been the single method of assertion. I must leave it at that; to elaborate it would involve me in another somewhat recondite result of Gödel's, the completeness of quantification theory, and various other matters too technical for this evening.

One point involved is, of course, a criterion for deciding what theories *are* Platonistic and what ones are not. That is, a criterion for deciding whether a theory involves a commitment to abstract entities, or merely uses abstract words without presupposing corresponding entities.

The criterion which I use is this: a theory is committed to abstract entities just in case its variables of quantification—say the ‘ x ’ of ‘ $(x)(\dots x \dots)$ ’—are sometimes used to refer to abstract entities.

I have sometimes been misconstrued as meaning that the question whether there *are* abstract entities is a question of language. I almost wish I could agree to this, if only in order to sustain a *bon mot* due to Dr. James Grier Miller, who undertook to sum up the first of my papers on this topic with the formula *ontology recapitulates philology*.

Actually I mean no such thing. My criterion is for deciding not whether there are universals, but whether a given theory, a given form of discourse, *commits* itself to the doctrine that there are universals; and here philology has a place.

I haven’t tried to allow time to defend my criterion here, because I talked on that topic before this club last spring. Apropos of the prescribed theme of the present talk, I may merely add that this criterion of ontological commitment, if it is as adequate as I think it to be, is one more example of the use of modern logic in philosophy.

Before I stop I want to correct a wrong impression which I must have given when I suggested that empiricism, nominalism, and the repudiation of mathematics all converge. I must have given the impression that I thought empiricism and nominalism were involved in absurdity.

On the contrary, I am partial to empiricism and nominalism, and my taste for each is enhanced by the support which it proves to gain for the other.

True, I also have a certain taste for mathematics. But repudiation of classical mathematics as a body of meaningful statements need not involve abandonment of mathematics altogether.

For one thing, we may still pursue classical mathematics, and examine its structure, without us philosophers according genuine meaning to its formulas. And we may, as philosophers, still inquire into the reasons why such a system of meaningless formulas should be useful to science.

Secondly, we may explore the possibilities of constructing a fragmentary neo-mathematics devoid of Platonistic commitments. However useful the classical mathematics may be, it would be important philosophically—

from an empiricistic or nominalistic point of view—to show that a fragmentary non-Platonistic mathematics can be *theoretically* adequate for all applications to natural science.

In such a program, whose motivation is purely philosophical, techniques of modern logic would have to be exploited to the utmost.

Where Logic Is Going

Visiting Committee dinner, Harvard University, December 13, 1947

Logical proof, ordinarily so-called, may accomplish either of two things. It may prove a statement to be *logically valid*, like ‘If Boston is a city then Boston is a city’; or ‘Boston = Boston’; or ‘Everything is identical with itself’.

Or, it may show that one statement follows logically *from* another statement or set of statements—where the individual statements themselves are perhaps ordinary statements of fact, not logically valid individually at all. This second function of logic is of course the important one when logic is used as a practical tool; it constitutes *deduction*, and is needed, to varying extents, in all disciplines.

But this second function of logic is obviously reducible to the first, that of picking out statements which by themselves are logically valid. For, instead of proving that statement B follows logically *from* statement A, we can prove that the *compound* statement ‘If A then B’ is logically valid. So we may conveniently, and without loss of generality, think of logical proof as *always* doing the *first* kind of thing: showing that some statement is logically valid.

Now the idea of logical theory, so far as proof is concerned, might, to begin with, be set at this: to discover a purely mechanical method of computation whereby we can always decide whether a statement is logically valid.

Consider equation in ordinary arithmetic, where specific Arabic numerals are combined on each side of the equation by operations like ‘plus’, ‘times’, and ‘power’. Every schoolboy has a mechanical method for deciding whether *any* such equation is true, i.e., arithmetically valid, or false: he merely multiplies out, and adds out, and powers out, and then sees whether both sides look alike in the end. Adults do it by machine.

Now what if a comparable technique for testing logical validity could be devised?

Frege showed, in the latter part of the past century, that all mathematics could be reduced to logic. (There may still be room for quibbling on the score of geometry and a few minor branches, but this is of little theoretical consequence, because, in any case, disciplines exactly *similar* to these—in one-to-one correspondence in every detail—*can* be constructed within logic.)

So if we had a mechanical technique for testing logical validity, we'd have a mechanical technique for checking every mathematical proof.

Mathematicians would stop publishing proofs of theorems. They'd state their theorems and talk about their utility, but leave it to the reader to test their truth. Mathematical journals would reduce to leaflets.

Mathematical problems which have withstood solution for generations or centuries, such as the celebrated Fermat problem, or the four-color-map problem, or Goldbach's problem, would be solved by an infallible computation by a clerk or a machine.

So we're ready, perhaps, to expect that a general mechanical technique of computation for testing logical truth is impossible. It has, as a matter of fact, been *proved* impossible.

Mathematical logic divides along natural lines into four departments or levels of increasing complexity. I'll name them, but won't stop to explain what they are: first, truth-functional theory; second, monadic quantification theory; third, general quantification theory; fourth, general set theory.

The first two levels include already, in effect, the traditional logic of the syllogism, and a great deal more. Now it happens that there *is* a general mechanical technique for computing validity for these first two levels. It is due to Löwenheim and dates from 1915.

But it has been proved that there *can* be *no* such technique adequate to the *third* level, let alone all four.

It is surprising that such an *impossibility* should be *provable*. The proof was accomplished by Church in 1936, building on earlier results of Gödel.

Here, then, is an important contrast between the lower two and the upper two levels. For the lower two we have an infallible method for mechanically computing validity; for the third and fourth we know that such a method can never be attained.

For logic as a whole, then, since no mechanical technique of testing is possible, we must settle for some general method of *proof* which falls short

of an out-and-out *test*. The difference is this: *proof* may be discovered by luck or ingenuity, or may even happen to resist discovery forever though in principle it be discoverable. When it *is* discovered, though, its *correctness* must be mechanically testable. That is, a spark of creativity in writers of mathematical papers may be irreducibly required, but not in editors or readers. Proofs as usually published do often call for ingenuity also in the reader, but in principle the gaps can always be filled in so that the validity of each *step* of the proof can be checked mechanically. But the proper *series* of steps had to be discovered, and cannot, in general, be got by a pre-scribable mechanical process.

For logic as a whole, then—which includes all mathematics—the most we can require is a complete theory of *proof*, such that every logically valid statement *has* a proof which, if logically *discovered*, can be checked mathematically to see that it *is* a proof.

Such a complete theory of proof is *in fact at hand* covering the first three levels. It was proved complete by Gödel in 1930.

But even a second-best type of theory such as this is impossible when we come to the fourth level, set theory. Gödel showed in 1931 that no consistent theory of proof for set theory can be complete.

No matter how carefully we devise a general consistent system of logical proof, there will be logically valid statements of set theory that not merely will not, but *can* not, be proved in it.

More particularly, Gödel shows it is impossible to get a complete consistent system of proof even for a quite elementary algebra of whole numbers.

This discovery came as a shock, for it had seemed natural to suppose that in algebra, or in logic generally, validity *consisted* in provability. But not so.

Here, then, is an important contrast between the lower three levels and the fourth. For the first three levels we have a technique adequate to proving any valid statement, though in the third level the proofs are not mechanically discoverable. But in the fourth level we can't have even that.

Another contrast between the third and fourth levels, to which I attach special philosophical significance, is this: the fourth-level logic, in its standard interpretation of course, is *Platonic*, in this sense: it requires that some or all of the entities to which the variables of the theory refer be *abstract* entities. No such presupposition is involved for the first three levels.

So there are *two* very basic differences between the fourth level and the lower ones. So basic that I'm somewhat inclined to think of this as the

boundary between logic and mathematics. This is a terminological question. If we do draw this distinction, we can no longer say that mathematics is part of logic, or derivable from logic. Logic, in this narrower sense, has a complete theory; and logic in this narrower sense is free from ontological commitments. Mathematics, on the other hand, becomes identical with set theory. And it becomes at the same time (under the normal interpretation of the signs, of course, and I know no better interpretation) a general theory of abstract entities. And it is forever incapable of being given a complete set of true axioms, or any other complete and consistent proof machinery.

My present interests, insofar as they run to technicalities of logic rather than to problems of semantics and theory of knowledge, tend to cluster around the *boundaries* I have described, and various problems relating to them.

Animadversions on the Notion of Meaning

Philosophy Colloquium, University of Pennsylvania, December 6, 1949

Common error to confuse meaning of a word with thing the word refers to. 'Boston'. Meaning of word = what the word is name of.

Leads quickly to difficulties:

- 1) Every linguistic form is name of something or meaningless.

Leads to swollen universe. There has to be such a thing as Pegasus, entity of some sort. And strange entities for words of *all* sorts to name: and, than, . . .

- 2) Empirical identifications become unallowable.

Astronomy powerless to identify Morning Star and Evening Star unless these phrases have same meanings; and if they do, astronomy needn't bother.

Especially evident from this example (Frege) that two things are being confused: *meaning* and *reference*. (Sinn und Bedeutung). No need to mean same in order to name same. No need to name at all in order to mean. Relatively few forms name, or even purport to (singular terms).

What philosophers now call semantics better called two: *theory of reference* and *theory of meaning*.

Theory of reference meager? Not wholly. Besides naming there is *truth*. Also denotation, or "truth of". Semantics as of Tarski (and poorly distinguished part of Carnap) is theory of reference. Here the so-called semantic paradoxes.

But my topic: theory of meaning. Synonymy (= sameness of meaning), analyticity (= truth by meaning), significance (= possession of meaning).

Distinction between meaning and reference no novelty. Corresponds to traditional “intensional vs. extensional”, “connotation vs. denotation”, and still older “essence vs. accident”.

For Aristotle, essential in men to be rational; accidental to be two-legged. But by same token: cyclist.

Improved stage of Aristotelian essentialism is theory of meaning: Instead of properties being essential to certain things, they are involved in *meanings* of certain *words*.

So notion of meaning comes of splitting essence off from reference.

What are these meanings? Advisable to limit problem:

- 1) Meanings of statements only
 - a) Only statements are concerned by analyticity.
 - b) Contextual definition (Russell’s advance on Hume) treats parts as understood only in whole
 - c) Also linguistics takes statement as basic vehicle of meaning.
- 2) Cognitive meaning only: i.e., such as relevant to truth value of the statement. (Not expressive poetic.)

Empiricism treats (such) meanings as the conditions of verification in experience. “Criterion in Mind” (Lewis). Therefore:

statement is significant $\equiv$. . . (familiar empiricist doctrine = operationism)

statements are synonymous (cognitive) $\equiv$ same criterion in mind.

statement is analytic $\equiv$ criterion met ipso facto by *any* experience.

This is much the state of empiricism; but a state more typical of metaphysics generally than empiricism has any right to be in.

Notion of criterion in mind is too vague and unempirical. “Everything colored is extended”?

Better to seek standards directly of synonymy, significance, analyticity. Theory of meaning ceases to be theory of meanings.

Interdependent:

significance = synonymy with self

analyticity = synonymy with ‘0 = 0’

synonymy = analytic biconditional

So enough to explain *analyticity*.

Truth depends in general on language *and* fact. “It rained on December 1 in New Jersey”. Analytic = factual component nil.

This notion fundamental to logical empiricism. Their empiricist standards of significance have a special exemption for logic and mathematics: these statements needn’t meet empiricist standards, because analytic or contradictory.

Fundamental also to Cambridge analysts, “Entailment”.

Still, no objective criterion. “Null factual component”? “True in all possible worlds”? These are repetitions at best.

“True by virtue of meanings”? There are two problems: “by virtue of” and “meaning”. “By virtue of” = “logically deducible from”? Then thus: Becomes logical truth when synonyms substituted, e.g., “No bachelor is married”; “all brothers are male”. Therefore presupposed:

a) logical truth (from logical vocabulary and truth [theory of reference])

↑

not so bad

b) synonym for synonym; but circular, and even away from restriction to statements.

They say also: true by definition, i.e., definitional abbreviation of logical truth. “Bachelor = _{definition} man not married”, etc. But “definition”—what?

Legitimate appeals to definition are two:

- 1) Convention of abbreviation for convenience: imagine expanding always. Then our example *isn’t* “No bachelor is married”, but “No male not married is married”. *Mere* logical truth, not raising question of synonymy. Or—
- 2) Assertion of synonymy between two given forms. Then “true by definition” = “result of putting synonym for synonym in logical truth”—preceding problem.

Most confusing of all is Carnap: “true by semantical rules”. How do we find them? “Can’t in natural languages, because vague. But in $L_0, \dots$ ”.

Wrong to blame the difficulty on imprecision of natural language. “Blond”.

Saw analyticity as “null factual component”—which seemed sensible since there are the two components. But couldn’t find criterion. Maybe whole separation of linguistics from factual components is groundless metaphor.

Another view, not distinguishing the components: we have our sense experience, and our own system of beliefs. Latter does *more* than enumerate sense data; talks of physical objects, electrons, forces, genes, mutations, number, etc. Such talk is, if you like, myth; anyway, a social growth. Survival value: it enables us conveniently to plot the uniformities of sense experience, for prediction.

But it is underdetermined by experience. System as a whole must conform to experience along periphery; but disconformities can be repaired each by any of many changes of the system. We choose by two canons:

- 1) maximum elegance of whole system,
- 2) maximum conservatism.

By 2), the more central principles resist change the more. These might be called the *more analytic*: matter of degree.

In a way they are linguistically true or true by convention: it is of the nature of our own artificial scientific construction to change them last (i.e., to involve them centrally). But no *boundary* here between linguistic component and factual component.

The central, less casually mutable statements needn't be general. Complex structure of interdependence *by* general principles, between general and very casual statements (e.g., "No bachelor is married"). Or even the evident *worthlessness of change* of truth value of "No bachelor is married" can operate.

Room for analysis, and need for criterion, remain even when analyticity is matter of degree.

If analyticity has degrees, so do cognitive synonymy and significance. Plausible for significance: statements of length more significant than of electrical charge; statements of electrical charge more significant than of age of universe, or position of an electron. Proximity to operational criterion, plus degree of operability of criterion (supposing criterion at hand to make sense of such comparisons of criteria).

And cognitive synonymy comes in degrees? Clearly poetic does: better and worse translations. Maybe degrees of cognitive synonymy are an intermediate zone.

Or maybe (Williams) the degrees of synonymy have several dimensions, of which the cognitive is one.

None of this makes sense pending the framing of empirical criteria for

at least the comparing of degrees of synonymy. (Not yet done by linguists even, for synonymy of any sort.)

For this and relative degrees of analyticity, criteria are needed comparable to the field linguist's implicit behavioristic criterion of sentencehood for A.

Curiously, professedly scientific philosophers feel no need of such criterion and are even willing to assume absolute concept of analyticity and the rest.

Challenged, they sometimes say: let's hope for clarification, but meanwhile postulate these concepts subject to certain laws which we should require any ultimate clarification to realize.

But scientific philosophy is notoriously at a standstill, and maybe the facile postulation of improbable concepts is to blame.

As a novel experiment in logical empiricism and scientific philosophy, I propose talking sense: adhering to terms for which we can fashion at least rough criteria on the occasion of using them. Maybe this policy will lead to a clarification of analyticity and cognitive synonymy, at least in terms of degree. Or better, maybe it will steer us away from such notions altogether, away from the last vestiges of Aristotelian essentialism.

Maybe there are more fruitful ways of anatomizing the behavior of talking animals, the methodology of science. Our chances of falling into such ways could be improved by freeing ourselves from the dead hand of entailment, analyticity, and meaning. Anyway, little is to be hoped for from a theory of meaning which is at present meaningless by its own standards—or would be if it had standards by which to judge.

The Entangled Philosophies of Mathematics

Harvard Philosophy Club, October 2, 1950

It has become traditional to speak of the philosophies of mathematics as three: *logicism*, *intuitionism*, *formalism*. Impressionistically stated:

Logicism: Mathematics reducible to logic, so that mathematical truth is grounded in logical necessity.

Intuitionism: Mathematical truth is grounded in certain infallible mathematical intentions—e.g., the intuition of a next integer after any given one.

Formalism: No meaning is to be sought in theorems of mathematics, but only in the rules of proof—those being rules for the manipulation of mere notational patterns.

Consider logicism. Must distinguish two senses in which mathematics is sometimes said to be reduced to logic.

- (1) Lazy way: mathematics is committed not to truth of its postulates but only to *logical implications* between postulates and theorems.

Espoused by many mathematicians, including *early* Russell: “Mathematics is that subject where one never knows what one is talking about nor whether what one is saying is true.”

This isn’t logicism; merely a form of change-the-subjectism, or illogicism. [Explain.]

If it is more than change-the-subjectism, and says that the postulates and theorems of mathematics are to be allowed no meaning at all (rather than merely taking mathematics as not concerned with those meanings), then we have *formalism*; still not logicism.

True logicism says rather that mathematics is reducible to logicism in

(2) serious way.

Positive integers and zero $\rightarrow$ classes (Frege)

Rationals $\rightarrow$ ratios of positive integers (Peano)

Reals $\rightarrow$ classes of ratios (Dedekind, Russell)

Functions $\rightarrow$ relations (Peano)

Relations $\rightarrow$ classes of ordered pairs (Peirce)

Ordered pairs $\rightarrow$ classes of classes (Wiener)

Illuminating. But it is a reduction to “logic” only in very broad sense—better “set theory,” which is most speculative branch of all mathematics! Misleading to say that mathematical truth comes to be grounded in “logical” necessity, if this connotes bedrock. Logic in this sense is committed to a towering ontology of universals.

If one is to take the reduction to “logic” as a justification of mathematics, one might as well have declared mathematics as in no need of justification in the first place—thus waiving the problem.

If we are to say that “logic” itself is justified by some direct intuition of forms, we might as well say this of unreduced mathematics outright; for set theory is no more intuitive than the rest. Actually, since the advent of the logical paradoxes, the hardest Platonists must recognize that at the level of set theory intuition is bankrupt. [Explain.]

Might defend set theory, or mathematics generally, on a *pragmatic* basis. Certainly pragmatism does seem to be our method in foundations of set theory. But logicism has become irrelevant; justification whether by intuitive considerations or by pragmatic ones, is no better nor worse for “logic” than for mathematics outright. Reduction of mathematics to set theory is important, but not for the giving of the grounds of mathematical truth.

What is the pragmatic justification of mathematical principles? Utility of mathematics is as auxiliary to empirical science. From a pragmatist’s point of view mathematics is neither more nor less right than it is convenient as an adjunct of science. The function of science as a whole may be taken to be prediction of experience. A hypothesis or law in physics or biology is good only insofar as the total system of which it forms a part enables us to anticipate experience; and the same may be said of pure mathematics itself, thought of as a functioning part of the total scientific conceptual scheme.

But this gives mathematics no different ground, in principle, from physical hypothesis. If the purpose of logicism was to ground the mathematical

truth in logical necessity, we have wandered far from the purpose. First the logic needed is so broad that “logical” necessity is no real reduction of mathematical necessity. Second, mathematical necessity has now appeared to have no better footing than the necessity of physical hypothesis; no necessity about it. But this last line came of a special turning: despairing of intuitive grounds for general set theory and classical mathematics, we settled for pragmatic grounds.

There is an alternative: insist on intuitive grounds, repudiating whatever in mathematics can’t be thus grounded. This is *intuitionism*—the second of the three philosophies of mathematics as listed. Brouwer, Heyting, Weyl; forerunners Poincaré, Kronecker. Their repudiations cut deep—even into such elementary logic as quantification theory, as applied to such elementary mathematical objects as positive integers. They reject truth even of:

Every even number is sum of two primes or some even number is not sum of two primes pending such time as one or other part is proved.

Intuitionism (or constructivism) admits of different forms of degrees; how much logic and mathematics is to survive depends on what intuitions are available, and therefore on the particular theory of intuition adopted. But any intuitionism is bound to sacrifice much of classical mathematics—portions surely of theory of real numbers and theory of functions.

This is where *formalism* (Hilbert’s school) steps in. The formalist would eat his cake and have it too. He is unwilling to make the sacrifices of classical mathematics in which the intuitionist acquiesces. But *like* the intuitionist he is unwilling to accord *meaning* to mathematics beyond the reach of intuition.

So he says: keep classical mathematics and exploit it where useful, but don’t worry about investing it with meaning and truth.

Therefore no place for standards in mathematics? Every theory legitimate? Can stop trying so hard?

No. The standard which remains is: conformity to rules. What rules? Any, provided they are explicit. Or *almost* any; any not leading to self-contradiction.

Because correctness ceases to be a matter of content and becomes a matter of formal manipulation according to rule, the rules must be explicit and *formal*. [Explain.] Therefore Hilbert’s creation of *proof theory* (syntax, metamathematics).

Would formalist say that there is *no* basis of preference between one set of rules and another, barring self-contradictions? Well, not quite; he’d con-

cede *pragmatic* grounds of preference, for he by no means denies utility of mathematics.

Even insistence on freedom from contradiction is an extreme case of pragmatic preference; for, a system with a self-contradiction yields *all* formulas as theories (if contains usual logic) and therefore makes no discriminations; ‘proof’ here surely useless.

Consistency not only a pragmatic feature of extreme importance, but also one that can be investigated formally (i.e., without exploring possible uses). So proof of consistency is central part of Hilbert’s program. (Truth of content gone). We know, since paradoxes, that consistency in non-elementary parts of mathematics *can’t* be assumed as a matter of course, no matter what our philosophy of mathematics.

But now see how our picture of the philosophies of mathematics begins to show a convergence from opposite ends. We started with logicism. It dissolved between our fingers, and we found ourselves faced with a choice of pragmatism or intuitionism or formalism. But now formalism begins to look like pragmatism. Systems of mathematics are chosen pragmatically for development. What difference is there between formalism, as we now see it, and the pragmatism considered earlier?

The pragmatism considered earlier seemed to put mathematical principles and physical hypotheses in same boat: both are fragments of the total scientific network, which, *as* a whole, is fashioned and revised by pragmatic standards of convenience in relation to the business of anticipating experience.

Formalism *seems* to separate mathematics drastically from natural science, by treating theorems of mathematics as empty formulas. But is this a real difference? In what way *aren’t* the hypotheses of physics empty formulas? They are parts of a total conceptual machine which, as a whole, helps us anticipate experiences; but mathematics is part of that machine too.

When a supposed anticipation of experience fails, we can in general change the system at one point *or* another to reconcile. The experience doesn’t tell us just which statements to change and which not.

But in general there are some statements which we are readier to change—perhaps with a pragmatic view to disturbing the whole least.

Many statements, indeed, we keep in abeyance pending further experiences—thus cutting down on revocations.

Because of those large regions for adjustment in our evolving total system of science, we can propose to keep some large bodies of statements

fixed come what may. Here we are free then to describe in advance, by reference to purely *formal* or notational features, whole classes of statements which are to be held to as true. These are the so-called *a priori* statements, and the formal features which mark them out are what make up the formal rules of proof.

It seems reasonable to suppose that the vocabulary of mathematics is that part of the total vocabulary of science such that *all* statements constructed purely of the *part* vocabulary are settled *a priori* as true or as false, by adopting formal rules of proof. This view of math is reasonable because it fits the common-sense principle that in mathematics truth is a matter of demonstrability and not of empirical observation.

Perhaps *formalism* goes beyond ordinary pragmatism in making just this added point then: the mathematical vocabulary is just that part of the total vocabulary of science such that *all* statements of the part vocabulary are decided according to formal rules of proof. Some statements outside mathematics seem also settled *a priori*, e.g., that momentum is proportional to velocity, or equal to mass times velocity; but the mathematical vocabulary is such that *every* combination of this vocabulary into a statement is formally decided: provable or disprovable (negation provable) by rules.

But if this is wherein formalism improves on ordinary pragmatism, then formalism is disproved by Gödel! [Explain.]

The formalist might at this point be persuaded to reject parts of mathematics as illegitimate—including even some portions of elementary number theory—in order to get down to a formally completable mathematics. But if he does this he has failed dismally in his original project of eating his cake and having it too. He will have to whittle down his mathematics to intuitionistic size.

Gödel's proof is the fairest flower of proof theory, the discipline which Hilbert founded from formalistic motives; yet it is at the same time a shattering triumph for intuitionism over formalism.

So the philosophies of mathematics seem to be boiling down to two. A straight *pragmatism* seems to sum up what we can retain of logicism and formalism as philosophies; and on the other side we have intuitionism.

The reduction of classical mathematics to set theory is a heritage of logicism which remains important, but it is not a philosophy. And we have an important heritage of formalism in proof theory, which came to maturity in Gödel's bombshell, and has yielded other remarkable results in the two decades since.

As pragmatists we can go on with classical mathematics, obeying formal rules of proof as usual; but we must recognize that the previously suggested boundary between mathematics and the rest of science has been rubbed out by Gödel's discovery. We may try to find some other way of drawing the boundary, if we like; but we don't need a boundary in order to go on with our work.

Pragmatism is the way of eating our cake and having it too. But even if we adopt this attractive way, still intuitionism or something like it—some constructivism—will remain important to us for two reasons.

First, it is obviously convenient to exploit a completable part of mathematics where we can. Any completable part will be very limited and constructivistic in character, though not vice versa.

Second, there is the little matter of proof theory itself, which is vital in understanding the workings and potentialities of our own various formalisms even in the incompletable parts of mathematics.

But proof theory is itself a strictly mathematized discipline, equivalent in fact to a certain fragment of number theory. And this much mathematics needs to be understood more than as a formalism, even if the formalisms to which proof theory is applied are not. We are thus brought back to an intuitionism at second level, second intension, even if we refuse to be hemmed in by the restrictions of intuitionism at the start.

So I seem to have ended up with an overall pragmatism, embodying much of the formalistic point of view; but still with attachment of a special value, a central importance, to that core of mathematics which is acceptable to the intuitionist.

— 8 —

Meaning

Center for Advanced Study in the Behavioral Sciences,
Stanford University, March 1959

When man in the street talks semantics, his paradigm is relation of names to people and places.

Each name has its thing.

Is extended to other words.

Reason: distinct words, e.g., in distinct languages, have “something in common.”

So we get abstract entities—attributes? ideas?—as “meanings” of general terms.

And meaning for all the particles. Ideas?

If they were *just* projections of the words and phrases, they would be idle and therefore harmless.

But they aren’t *just* projections; they are to coincide for translations.¹ And to diverge for identical words “used with different meanings.”

Source of intuition of domestic synonymy.²

They therefore cease to be idle and become harmful: illusion of accounting for correct translation and for domestic ambiguities and communication failures.

Actually mental entities are absurd in linguistic theory, whatever one’s philosophy. For language is learned socially.

Voltaire said “If God hadn’t existed. . .”. I say “If ideas did exist we’d have to disregard them.”

1. This was a reason for assuming them: *distinct* words share them, so they are other than the word.

2. If someone doesn’t admit substitution of ‘married man’ for ‘bachelor’ we don’t have any idea what he will do with the word; communication is threatened. So we are encouraged in our myth of meanings as shared ideas.

So we get a behaviorization of meaning in Wittgenstein and Oxford: meaning is use. But armchair behaviorism.

Difficulty: take any word other than terms teachable by pointing. Its use is in relation to other words.

How say its translation has same use? Only in relation to translations of the other words. Circular.

Obvious first step toward coping with situation. Take not word but sentence as basic semantic vehicle.

Diminishes difficulty of contextual relativity: use of sentences isn't *as* relative to that of other sentences as use of most words is to use of other words.

Objects of translation become sentences. Importance of words is for finite codification.

O.K., what of meaning of sentences?

Somewhere, should mention information theory, and remark on its irrelevance to this problem.

Good in its place. But it has to do with quantity of meaning (in some sense) at best, not identity of meaning.

It supposes a prior binary coding, in effect: *you* may prepare a question list, yesses and noes to which will exhaust what you want to convey. The number of yesses in the executed questionnaire is the *amount* of information.

Nothing is said as to when two executed questionnaires, with different questions, convey *same* information.

Get back to meaning = use, said of sentences. Or, since sentences often have their significant use only in relation to other sentences, take long enough sentences for independent translation—whatever that is.

So that we can talk of sameness of use *not* relative to further *language*. Sameness of extra-linguistic circumstances.

Manageable for sentences of special sorts, that can still be kept short for the purpose. Stimulus meanings = classes of stimulations = classes of irradiation patterns, etc. Occasion sentences.

Primarily observation sentences. But occasion sentences generally, if translator goes bilingual.

But sentences generally? Can't separate relevant non-verbal occasions, unless by taking in whole system of discourse. And what is the occasion of a chapter? How can we compare *such* occasions as basis for translation? We translate the whole relative rather to *interests* (engineering, literary, . . .).

Analytical hypotheses. Indeterminacy.

Osgood suggested sameness of neural hookups, to get sentence correspondence in a bilingual. But this isn't what we *want*; two bilinguals could have different such correspondences, and we don't want to call one wrong.

Similarly for comparison of ways of learning.

What then of Whorf? There is only a question of difficulty of simple³ natural translation; not bizarreness. Or take Lévy-Bruhl.

3. Quine's handwritten text has the abbreviation "pct" between "simple" and "natural." We have not been able to identify a suitable reading of these letters.

The Way the World Is

(For Harvard's 350th), March 1986

Let me begin by stating what I take to be the business of physics. It is the discovery of the ultimate constituents of the world and their regularities. Other sciences that are likewise concerned to discover regularities in the world are for the most part not derivable from physics, but only because their objects are excessively complex aggregates of physical elements. Their regularities have to be discovered afresh rather than just being computed from the behavior of the ultimate physical constituents. They are not reducible to physics in polynomial time, as the computer people would say. But the behavior of these gross objects *is* nevertheless the sum of the behavior of the ultimate physical constituents, however incalculable.

This is the sense in which physics is basic. But what now if an occult phenomenon such as telepathy were conclusively confirmed; would the physicists then have to move over and make room for a second basic science? They would not. They would repair in glee to their desks and linear accelerators. The mechanism of the new phenomenon would cry out for investigation, and one more revolution in physics would be on the way. The theory eventually emerging might not seem very corporeal, but then neither does current physics. The objective is just full coverage of the ultimate ingredients and forces of the world, whatever they may be. The buck stops in physics.

Now to the way the world is. I gather from physicists that the world is a four-dimensional multitude of microphysical states or events. In large part these are strung out in one-dimensional manifolds called particles; one-dimensional, that is, in space-time. The world tapers back in the time dimension to a point, and it spreads forward without limit. Or maybe it

swells to a maximum and then tapers in that direction too; I am awaiting further word from the physicists.

My reason for deferring to the physicists is not that I am dazzled by their achievements, though I am. I would defer to them in these matters anyway, for they are the specialists whose business it is. But their account remains tentative, for they are prepared to correct themselves.

My sketch of our four-dimensional world does want supplementation in one respect: we must also recognize an infinitude of abstract objects, notably numbers, functions, and classes, out of space, out of time. Many of our hard-headed friends will dismiss such objects as figments of man's creative imagination, but such dismissal itself usually dissolves into empty verbiage when we try to give it objective meaning. When physics is explicitly codified with help of the notations of logic and mathematics, it makes use irreducibly not only of variables that range over particles or other physical objects, but also of variables that range over numbers and other mathematical objects. Physical theory thereby assumes not only the physical objects, which occupy space-time, but also those mathematical objects, which do not. It assumes both equally if the assuming of objects consists in admitting the objects as values of one's variables; and this is the only way I see, apart from empty rhetoric, to speak of the admission or exclusion of objects. Thus it is that my very acceptance of physical theory carries with it a recognition of these insubstantial but indispensable mathematical auxiliaries.

Granted then that it is the business of physics and other sciences to find out how the world is, how does the scientist check his theory? He checks it by deducing from it that a certain observable event should occur in certain observable circumstances, and then checking his prediction. And what counts as an observable event? There are two requirements. One is that the observer's verdict on whether the event is happening or not must be responsive to the concurrent stimulation of his sensory receptors: he must be disposed to say "Yes" under some stimulations and "No" under others. The other requirement is intersubjectivity; all witnesses to the event must be prepared to agree in their verdict, provided that they know the pertinent technical vocabulary and their sensory receptors are in order.

Prediction of observable events, then, is the test of science. It is how science keeps contact with its subject matter, namely, the world. What I am calling science, moreover, for want of a better name, includes various pur-

suits in which prediction is not at all common: it includes economics, history, even mathematics. Ultimately and indirectly, however, prediction is what also keeps these sciences empirically significant, because of systematic connections between these sciences and other domains where prediction does prevail.

One central consideration to this effect is what has been called *holism* and linked to the name of Pierre Duhem. The point here is that when a prediction is made in order to test some hypothesis, in physics say, usually the predicted event is not simply implied by that one theoretical hypothesis plus the observable experimental condition. If the logical connection were all spelled out, other purported truths of physics and common sense would figure in the argument along with the questioned hypothesis, and so probably would a fair chunk of pure mathematics. It is only the whole combination, strictly speaking, including even the pertinent mathematical laws, that is tested by the prediction. If the prediction failed, the questioned hypothesis could still be saved by retracting one of these other purported truths instead. We could even retract one of the mathematical laws.

We would not in fact choose thus to disrupt mathematics, because that disruption would reverberate too severely throughout science. We are restrained by a maxim of minimum mutilation. That maxim perhaps explains why mathematical truth is regarded as necessary truth. The maxim applies also, in a less extreme way, to other parts of science. The tenet that we choose to retract, to accommodate an unexpected observation, will be one whose retraction threatens the overall system least, other things being equal. Other things are not generally equal, however, and other considerations than minimum mutilation—simplicity, for one—are brought to bear. The aim, of course, is a revision that will best predict further observations; and the inarticulate intuitions that govern the groping are not well understood.

The lesson of holism, however, remains intact for all that. Namely, what implies a predicted observation is the observable experimental condition plus not just some one theoretical sentence, but a great cluster of them. And thus it is in particular that mathematics shares the empirical content, the predictive content, of the sciences in which it is used.

In economics and history there is indeed occasional direct prediction; prediction of a recession, prediction of an archaeological find. But the main dependence of such sciences upon prediction is, again, indirect. They depend heavily on tacit everyday knowledge, which is continually molded

and controlled by fulfilled and unfulfilled expectations, hence tacit predictions. Also they share the logic and some of the mathematics that imbibe indirect predictive support from their in the more overtly predictive sciences. All of science coheres as a continuous fabric, and a more tenacious one than we tend to suppose.

Various sciences have outlying portions still, whose relation to prediction is more remote than any of this. Higher set theory is one case. These outliers have a legitimate place in our system of the world, if at all, only in the interest of overall simplicity. They deal with questions that can be formulated in the same vocabulary and the same idioms that are also useful, in other combinations, elsewhere in science. We accordingly accept those questions as meaningful, because it would complicate matters more to excise them as meaningless; it would mean unduly gerrymandering the overall language of science.

There are further reaches of discourse, however, for which not even these claims to a scientific status can be made. One thinks here of bad metaphysics and benighted forms of religion. Also there is great poetry and fiction. For all their greatness these do not describe reality, for the most part, nor are they true. John Keats's first law of aesthetics, "Beauty is truth," and his second law, "Truth beauty," are both false.

Poetry and other arts sometimes heighten our sensitivity to certain parallels and contrasts in nature, and in this way they might at some point even spark an idea for a hypothesis within science. But even then they would not be supplementing science or second-guessing it. They in no way participate in the logical links of implication that relate the fabric of science to the predicted observations. Prediction is the language game, as Wittgenstein might have put it, that distinguishes science; and it is one language game among others.

Prediction is not the main purpose of science; not any more. Perhaps it was in remote times, when natural selection was first molding the rudiments of what we now call scientific method; for successful prediction has survival value in eluding predators and catching prey. Today the major motivations of science are technology and intellectual curiosity. But prediction of observable events remains the bottom line, the check that keeps science honest, honestly factual.

To understand this connection between observation and scientific theory is to understand the nature of evidential support: the nature of our evidence for how the world is. Let us examine the connection. What I am

calling observable events, as I explained, are events to which the scientists are prepared to attest on the strength of their sensory stimulation at the time. The sentence that reports the event is directly associated with the sensory stimulation. The association may have been implanted by direct psychological conditioning, or it may have come about indirectly through past training in scientific theory itself; what matters is just that assent to the sentence has now come to be a direct response to observing the event. Such a sentence I call an observation sentence, for the community of scientists concerned.

The rest of the connection between observation and scientific theory is logical. When a theory is checked by experiment and prediction, what has happened in principle is that a conditional sentence 'If p then q ' has been deduced from the theory, where ' p ' and ' q ' are observation sentences expressing the experimental condition and the predicted consequence. A maze of interlocking sentences of theory and common sense, some logically implied by combinations of others, is what connects high-level scientific theory with these observation conditionals. It is by osmosis upward, from the implied observation conditionals to the impliers, that science imbibes its empirical content. Scientists do not of course delineate all the intervening links, the logical steps.

We see then that what are concerned first and last are sentences: sentences at the bottom that are conditioned as wholes to the stimulation of the scientists' sensory receptors, and sentences of high scientific theory at the top. Where then do objects get into the act, and the terms that refer to them? The terms lurk inside the observation sentences and the theoretical sentences. Reference to objects, which has always seemed so central to science and knowledge, plays only a subordinate and auxiliary role in the relation of theory to evidence. It is part of the apparatus that links the sentences in implication. By dipping into the logic of quantifiers, one can show how the positing of objects serves to tighten the logical links. And one finds further that the contribution is purely structural: any new objects would serve as well as the old, as long as we can relate the old and the new ones one to one.

We are bound to conclude that two global scientific systems of the world can fit all possible observations equally well and yet differ utterly in the objects of which they purport to treat. Save the structure and you save all.

I began this account with an avowal of naïve realism. I deferred to science on how the world is. I may seem now to have argued myself rather

into a free-floating idealism, where even the physical objects attested to by science can be reconstrued at will.

The apparent conflict dissolves when we distinguish between *ratio essendi* and *ratio cognoscendi*; between questions of fact and questions of evidence. Let us consider how the one domain fits into the other. Here we sit in the material world, and within it we find the scientists themselves, generating their theories for us on the basis only of such information as reaches them through the impact of molecules and light rays on their sensory receptors. We can study and analyze the relation between that sketchy input of data and that torrential output of theory. It is the very study that has been occupying us here for the past ten minutes. Clearly there is a lot of slack between that input and that output, and much scope for creativity. In particular there is utter slack, we found, between the mere structure contributed by an ontology of objects and the particular ontology adopted for the purpose. This is a finding regarding evidence: a finding regarding the relation between the stimulation undergone by the scientist and the theory produced by him. It is not a finding regarding what objects there really are in the world. For that we look, as always, to the theory produced by the scientist, in all his fallibility. Our reflections on evidence taught us that other answers preserving the same structure would fit all evidence equally well; hence also that they would fit it no better. We settle for what we have, in full scientific tentativity, there being no higher tribunal.

Pressing Extensionality

Føllesdal Festival, Oslo, June 13, 1992

Predicate logic: framework for science at its clearest. [State its syntax.]

Its grammatical particles are its logical particles. [Enumerate.]

Lexicon: predicates ($\geq$ one place)

If ‘ $\in$ ’ is in lexicon, can define (1) ‘=’, (2) description, (3) names, (4) term functors, (5) all classical math.

Virtues: clarity; deductive efficiency; completeness.

Source of its clarity: extensionality. [Explain.]

No theory is fully clear unless extensional.

Propositional attitudes not extensional, but needed.

Root: ascription of perceptions, in handing down language. [Explain.]

Other attitudes by analogy.

They violate substitutivity of ‘=’. Unreasonable.

Frege’s way.

Better idea: empathy, as in ascribing perception.

Especially clear in indirect discourse.

Content clause is what *ascriber* would have said.

But its ontology is *subject’s*, supposedly.

Hence failure of substitutivity. Cicero, Tully.

Quantifying in: also upset by ontology.

Quantifier and variable range over different worlds.

Failure of extensionality: subject’s ignorance of coextensiveness.

What are *objects* of belief, etc.? Facts? Propositions? Sentences! What the ascriber *would* have said. Adjustments: “*I am Napoleon.*” No presumption that subject can speak.

‘So that’ is a quotation mark. Indirect quotation vs. direct? ‘Say’ vs. ‘utter’. Quotation merely names a string. Failure of substitutivity becomes vivid. Failure of quantifying in becomes vivid. Quotation is analyzable by *spelling*. Thus reconciled with predicative logic and extensionality. Quotation made the failures vivid; spelling obliterates them.

Together they *digest* the propositional attitudes, syntactically and logically.

Semantically, propositional attitudes remain low grade.

That was *de dicto*. Ascriber plays subject’s role.

In *de re*, ascriber steps out of role.

There is someone whom R believes to be a spy.

$\exists x(R \text{ believes that } x \text{ is a spy})$.

Can’t quote. Ascriber’s ontology invades the belief.

Føllesdal localizes the *de re* quality in the *term*: R believes that *Ortcutt* is a spy.

Term can be taken *de re* (ascriber’s) or *de dicto* (subject’s).

If *de re*, amenable to substitutivity and quantification.

I pin the *de re* quality on *position* (“referential”). Position available to substitutivity and quantification. Advantage: applies to primitive termfree notation.

Modal logic differs: all positions are quantifiable, and substitutivity varies with the term.

So Føllesdal’s way is wanted here.

“*Genuine*” singular term obeys substitutivity and existential generalization.

Can say: term is genuine iff defined by *essential* traits of its object.

For modal logic invites essentialism.

Tenets of essentialism:

- (a) Some objects have some of their traits necessarily, others contingently.
- (b) Some traits are necessary in some objects, contingent in others.

Modal logic is committed to (a), on pain of collapse.

Modal logic without (b): much ado about little. Every modal clause would be true for all or no values of its free variables.

I have no use for metaphysical necessity. [Explain.]

Literary *fiction*: more akin to propositional attitudes. Story :: content clause; factual preface :: R. believes that . . . Its internal logic and its ontology can be assessed independently.

All is digestible, again, by spelling.

The *de re* virus can infect fiction.

Are the London and Gladstone real? or similar? Matter of degree; can be blithely disregarded.

But, to go back: I can't be so blithe about 'There is someone whom R. believes to be a spy.' A specific suspect in real world. What counts? A floating glimpse of unidentified in suspicious circumstances? Just any non-empty singular description acceptable to R?

SLEIGH'S REFUTATION

≥ three grades of useful cognitive discourse:

- 1) extensional, with fairly precise predicates.
- 2) extensional, with imprecise predicates (*de dicto* propositional attitudes).
- 3) irremediably intensional (*de re* propositional attitudes).

Number three can alert us to seek data *de dicto*.

Literary fiction and indexicals, including necessity, are off this hierarchy.
[STOP]

What further threats to extensionality?

SUBJUNCTIVE CONDITIONALS

A strict conditional? No! *Literary fiction*.

Suppress or supplant some knowledge, to suit purposes. Then reason seriously.

Indispensable to science, but outside it.

Literary fiction and science hand in glove.

Subjunctive conditional dominates ethics (Tännsjö). But can't handle ethics as *consecutive* fiction. Tännsjö finds modal logic convenient. Anyway ethics is not included in science, for all its importance.

Dispositions seemed based on subjunctive conditionals. Wrong. Differ only in name from other physical properties. Name alerts us to useful symptom.

Extensionality is challenged by causality.
Føllesdal's example:

Drinking poisoned water caused his death
" from that well " " "

Yes, no? But same event. So $\sim$ extensional. Restore extensionality by recognizing the event however described. Causal *laws* still relate described *classes*.

Probability? of events? Restore extensionality as above. Of states of affairs? Treat quotationally, like *de dicto*. How much remains, e.g., in quantum mechanics? Maybe (as Føllesdal suggested) must settle for intrinsic propensities, but still extensionally.

The Innate Foundational Endowments

March and May 1996

I.

Descartes thought we had innate knowledge and innate ideas. Locke thought not. I despair of sharpening the issue by defining the term ‘idea’. Definition even of ‘knowledge’ is in trouble since Gettier’s challenge of the definition of knowledge as true and warranted belief. However, we need no such sharpening of the issue to see that the evidence favors Descartes over Locke. It is reported that a child in early infancy will show anxiety when glowered at, and will even imitate a frown or a smile. The infant will show anxiety also after seeing an object steadily approach a screen, pass behind it, and then fail to emerge in due course at the far side. Further there is our spontaneous identifying of different perspectives of an angular object as all somehow the same thing. All this is innate, or, as we say, instinctive. Hence innate knowledge, we might say, and no innate words to express it.

The next imposing of the instincts having to do with knowledge is our *inductive* instinct, for it serves in its faltering way to generate knowledge. It is our own disposition to expect similar events to have sequels that are similar to each other. This is the basis of all expectation, all learning, indeed all habit formation, all conditioning. One act of ours has a pleasant consequence, so by induction we expect a repetition to have a similar consequence; such is conditioning. Since induction thus underlies all learning, it cannot itself have been learned; so it is innate.

Philosophers were long baffled by the power of induction. Our expectations occasionally fail, but not like random guessing. Philosophers have tried to account for this by pure logic, but without avail. The good batting average of induction is explained rather by Darwinian natural selection, as

we shall see, and the crux of the matter is *similarity*; for the inductive principle was that similars tend to have similar sequels. So bear with me while I dwell for a while on similarity.

I spoke of similarity of events, but let us stick to our data. As Locke and earlier empiricists appreciated, our ongoing clues to what is going on around us are limited to our ongoing sensory intake and owe nothing to telepathy or clairvoyance or other extrasensory perception. To put it in philosophical terms, our clues to events around us are limited to the triggering of our neural receptors by the impact of light rays and molecules.

It turns out, even so, that there is a skew in our responses to this sensory input. It is illustrated by the last of the examples that I cited of innate knowledge, namely our interpretation of the various perspectives of an object as one and the same thing. Different perspectives trigger very divergent sets of neuroceptors in the retina, but we account them similar. In my terminology, these neural intakes are receptually dissimilar, but *perceptually* similar.

Now we can get back to induction; for it is perceptual similarity, not receptual, that matters to induction. Induction disposes us to expect perceptually similar neural intakes. Since the inductive impulse is innate, then, one's standards of perceptual similarity must be largely innate too, though they can change with experience. Moreover, there has been survival value down the ages in correct expectation; helped our forebears elude their predators and catch their prey. Natural selection has consequently favored standards of perceptual similarity that tended to mesh with trends in the environment.

The perceptual similarity of perspectives of an object to one another is a vivid case in point: it behooved our ancestor to recognize a sabre-toothed tiger on sight, irrespective of orientation and perspective.

Perceptual similarity is subjective, in the sense that it relates one's neural intakes only to one another and not to another's. It is objective, however, in being experimentally testable. We can test it by testing his inductive expectations, and this we can do by reinforcing and extinguishing the subject's responses to stimulations. To test whether a neural intake A is perceptually more similar to B than to C, we see whether a response that was rewarded as a response to B but penalized as a response to C is still evoked as a response to A.

II.

Natural selection affords us no assurances of the continuing success of inductive reasoning, beyond whatever conviction we have that nature will continue in the even tenor of her ways. But it does explain why induction has served us so well up to now.

“Explain” in what sense? Descartes already had his purported explanation, in divine guidance. Natural selection is a better explanation, because it offers an intelligible causal structure; as follows. Correct expectation promotes survival; hence the better one is at induction, the likelier one is to live long enough to reproduce his kind. But inductive prowess is largely inherited. Therefore only random variations in this trait, from parent to child, will tend to be handed down and down if it is an improvement, and otherwise not. Inductive prowess thus grew by the selective retention of random variation.

This causal account raises questions in turn. The survival value of good induction is clear enough, but wherein does this inductive prowess consist? This we answer by highlighting perceptual similarity, along with its experimental test, and then explaining inductive prowess as conformity of perceptual similarity to environmental trends.

The vagueness of this does not bother me, but we are still left with questions about the mechanism of perceptual similarity standards, about the inheritance of them, and about the random variations, down the generations, that enhanced their conformity to environmental trends.

The similarity mechanism is presumably to be sought in the configuration of the neural network, mainly in the brain, and in the selective receptivity of the various synapses. Much of this clearly is innate, and much acquired. Research into the neurology of perception proceeds apace. The questions of inheritance and random variation bring us to the lively frontiers of genetics. The gene made its tentative *début* as a hypothetical entity a century ago, and in recent decades has been fleshed out in a bewildering richness of chromosomes and double helices of DNA. We can count on a continuing proliferation of questions, step by step with the discoveries. No end is in sight, for want perhaps of any end to see. Such is the way of scientific theory: explaining the familiar causally in terms of the ever more remote and conjectural.

May 17–18, 1996

The Copernican revolution transformed cosmography into a budding science of cosmology, which flowered in Newtonian mechanics. The Darwinian revolution struck yet deeper, illuminating the theory of knowledge and indeed metaphysics itself. Ever since Aristotle, efficient cause and final cause had been complementary ways of accounting for events. Efficient cause pushed for the past, while final cause or purpose, to the future. Efficient cause was exerted by nature, while final cause resided in human desire and effort, grandly supplemented by divine providence. Darwin's theory of natural selection dissolved final cause into efficient cause. The manifestations of purposeful design throughout the living world are causally accounted for by natural selection.

That much is familiar. But natural selection also has an uncanny way of clarifying and integrating other familiar points of traditional philosophy too, quite apart from final cause. One such matter is induction. Another is Descartes' doctrine of innate ideas. A third is Leibniz's doctrine of preestablished harmony. Another is the Protagorean doctrine of Man as the means of all things, and therewith the issue of realism vs. idealism. These are the five ramifications that I propose to survey . . .

10 A.M., 27 May

- C ① Leibniz: preestablished harmony
- D ② Protagoras: Man the measure
- F ③ Man and nature, idealism and realism, cunningly interleaved. Differences that we can discern and that matter causally. They cause differences in our perceptions *and elsewhere*. Is it so great that differences causally significant in the world should be causally significant in perception? Natural selection had a head start.

It gives us phenotypic taxonomy and well it might, for it is guiding the subject matter in its own self-propagation as well.

It creates the subject matter as well as guiding us to it.

- A ④ Natural selection the great solvent. Dissolved final cause into efficient cause. Dissolves realism and idealism into each other. Dissolves disparateness of neural input into a like output. Kinds are projected for terms, and terms are vague, applying in degrees. Relative similarity is what is objective; not natural kinds.

Biological species are atypically clean-cut.

Similarity is projected on the world from our associations.

B ⑤ *Innate idea*: similarity

E ⑥ Descartes: our kind God would not deceive us. Read rather: our cruel God eliminates poor expectors.

The five interlock. I shall have to zigzag among them.

To begin somewhere, I begin with the instinct of induction. It is basic to all learning, all expectation, so it cannot itself have been learned; it is instinctive. It is the tendency to expect similar events to have sequels that are similar to each other. Up to now it has stood us in good stead, for whatever reason. Up to now our inductive expectations have been overwhelmingly more successful than random guessing. Natural selection, accordingly, has fostered it down the ages; for successful expectation increases one's chances of living long enough to reproduce one's kind.

Let us keep close track here. We reflected that induction has been working quite well, but did not see why. We saw only that because the instinct *has* been working well, natural selection has handed it down the generations. So the question remains: *why* have similar events in nature tended to have similar sequels? We must now look to the similarity relation.

Similarity, as a personal and objective relation between events in nature, is notoriously indefinable. Since it is a matter of degree, the best we can hope for is a triadic relation of comparative similarity: x is more similar to y than to z . Does it mean that x shares more properties with y than z ? This strands us with the intractable notion of a property. The neutral notion of a class, or set, will not serve here; for x and y have joint membership in countless classes, and so do x and z . A property is a special set of class whose members are united in some distinctive kinship; and this kinship is just the similarity that we were hoping to define.

Despairing, then, of a definition of objective similarity, even in a comparative sense, let us turn to subjective similarity. This is objectively definable, oddly enough, and experimentally testable. It is *perceptual* similarity, and it is testable by reinforcement and extinction of responses. A stimulus x is perceptually more similar to y than to z , for a given individual, if it evokes a response that had been reinforced as a response to y , despite having been penalized as a response to z .

Each of us has his inarticulate seals on a standard of perceptual similarity. Perhaps my triadic rendering is an over-simplification. Patricia Church-

land has argued somewhere that it breaks down irreducibly into respects: x is more similar to y in this respect, less in that. But this need not concern us now. Whatever their structure, we are born with standards of perceptual similarity: they develop and change with experience, but are largely innate, since they implement the instinct of induction.

Now that we are armed with perceptual similarity, we can deal with our outstanding question of induction: the question why similar events have tended to have similar sequels. The answer is that natural selection has fostered successful induction, down the generations, by warping our standards of perceptual similarity into alignment with environmental trends.

In philosophizing about induction there is a familiar warning against circular reasoning: to defend induction on the grounds that it has worked well is to defend induction by induction. True enough. We have shown and meant to show only why it *has* worked well: natural selection has molded our similarity standards in recognition of inductive successes down the remote generations, but happily the environmental trends have not changed too much. Abrupt changes in our own time are taken in stride, for our standards of perceptual similarity do change with experience. But these changes do not get directly into the genes; they get handed down for a while by word of mouth, printing press, fax, e-mail, and internet.

We were at a loss for a definition of similarity in nature, but successful with perceptual similarity; and then we noted that natural selection has promoted a certain consonance between it and nature. By projection, then, we get something of an idea of natural similarity after all: events out there are similar if our perceptions of them are similar. Here is where I find Protagoras rearing his venerable head: man is the measure of all things.

Here also I find Descartes, with his innate ideas about the world; for a perceptual similarity is innate, and so then is its projection onto the world. We talk freely and uncritically of similarity of this and that in the external world, despite despairing of any general analysis of the idea; and now we see that it is an innate idea, projected only from within us.

Descartes, however, credited our innate gifts to a kindly God who would not deceive us. The credit goes rather to cruel natural selection, which refines our similarity standards by eliminating our less insightful brethren before they breed.

The general relation of similarity, that we thus project onto nature from perception, is of course relative at best: x is more similar to y than to z .

The Growth of Mind and Language

University of Oldenburg, Germany, June 5, 1997

I propose to delineate a succession of upturns in the development of the infantile and the prehistoric mind. I shall zigzag somewhat between development of the child and development of the species. This is convenient because of the degree to which ontogeny recapitulates phylogeny and also, as James Grier Miller observed, the degree to which ontology recapitulates philology. I have no experiments or excavations to report, but there has been room for insights by just raising new questions against existing knowledge.

To begin with there are the infant's immediate responses, instinctive and distinctive, to pleasant and unpleasant stimulation. If it was something he ate, moreover, he will accordingly accept or reject more of the same the next time around. We say that he has been conditioned. It is the basis of all learning. It embodies induction, which is the expectation that similar simulations will have similar sequels.

Induction is instinctive, since it underlies all learning. Though it is notoriously fallible, induction is much more successful than random guessing. Philosophers have wondered why this should be.

The answer to this question must be sought in the relation of similarity involved; for induction has us expecting subjectively similar stimulations, or neural intakes, to be followed by stimulations that are in turn similar to each other. It is a question of our subjective standards of what I call perceptual similarity. The success of induction suggests that our subjective standards of perceptual similarity are innately attuned to trends in nature, in such a way that events that qualify for us as similar will indeed commonly have effects that likewise so qualify. This happy harmony would be

due to natural selection, since successful expectation is conducive to survival.

Perceptual similarity diverges widely from mere sameness or near sameness of the neuroceptors triggered on two occasions—what I call receptual similarity. A square on the floor, seen from various angles, triggers very unlike sets of retinal receptors as it projects various trapezoids on the retina, seldom a true square; but these stimulations all qualify as perceptually similar, however dissimilar receptually. Or again, what is more to the point for natural selection, the projections of a tiger from various angles and in various contortions qualify as perceptually similar despite their receptual diversity. Natural selection has schooled us and culled us, down the ages, to reckon our similarities along survival-friendly lines, however innocently on our part.

So we have here, thanks to Darwin, a biologically explicable case of something like innate knowledge, though unconscious, of the uniformities of nature. It is implicit in our standards of perceptual similarity. We also have an answer to the nagging philosophical problem of induction, and it assures us that induction will continue to behave if nature stays on an even keel. And while we are on Darwin let us recall that it was he also that reduced Aristotle's final cause to efficient cause, at least for nature. His contribution to philosophy was overwhelming.

The evolutionary account of induction has a nice corollary. Heredity being what it is, there will be minor variations in the innate standards of perceptual similarity from individual to individual. Those variations veering more in tune with the trends of nature will tend to survive the more, and thus the efficacy of induction will tend ever upward except as nature herself takes too abrupt a turn.

Each of us is launched on society with his heritage of innate standards of perceptual similarity and with the predilection to induction that they implement. But each of us also, in the course of experience, revises those standards to some degree. Science is the great source of revision here, revising the very standards to which it owes induction. Those improvements will go down the generations in books but not in the genes, there being no inheritance of acquired characteristics. Each baby starts with the same old heritage, except for the occasional mutation.

So much, then, for innate endowments. The next consideration is the child's acquisition of language. Here I have nothing to add to parents' ob-

servations and professional studies until we get to the mother's involvement in the child's acquisition of expressions.

At this point the dog comes in, the mother tells the child it's a dog, the child says 'Dog', and the mother approves. Next day the dog comes in again, the child again says 'Dog', and the mother again approves. The stimulations of the child's retinas on the two occasions were perceptually similar enough, by the child's subjective standards, for the later one to elicit the response that was conditioned to the earlier one. The stimulations of the mother's retinas on the two occasions were likewise sufficiently similar by her standards to elicit her approval. Insofar, their subjective standards of perceptual similarity are evidently in intersubjective harmony.

Congruence of input does not explain the harmony, for they were viewing from different angles, and the dog was not still. Moreover, the same problem recurs with regard to the word 'Dog' itself: its sound is perceptually similar to the child's ear and to the mother's when spoken by the child and by the mother and when spoken on each of the two days.

One is tempted still to assume, in partial explanation, an approximate homology of nerve nets in the two individuals. But even this is ruled out by Darwin himself, who wrote in the opening page of his *Origin of Species*, that he found wild variation in the neural patterns even in simple insects arbitrarily chosen from a single swarm.

If Darwin closed one avenue, he opened another. As we saw in accounting for the successes of induction, natural selection has engendered a degree of harmony between standards of perceptual similarity and regularities of nature. Moreover, standards that harmonize with the same regularities are apt to harmonize pretty well with each other. So, in tailoring our similarity standards to the environment, natural selection has accorded us not only a key to nature in our instinct of induction, but also, as a side effect, it has encouraged our meeting of minds in the harmony of our similarity standards.

In all this there has been no need to suppose a qualitative resemblance between one person's sensations and another's, whatever that would mean, nor even a homology between one person's neuroreceptors or nerve net and another's. Our sense of perceptual similarity is evidently seated somewhere deep in the brain.

The expressions that we have pictured the child as acquiring at the mother's knee are what I call, *observation sentences*. My example 'Dog' was indeed a noun, but as learned it amounted to a sentence, 'That's a dog.'

What qualifies it as an observation sentence is that it has been conditioned to a range of perceptually similar neural intakes. Like the apes' cries and some bird calls, observation sentences announce current observation. But they contrast With the apes' limited repertoire in the abundance in which the child acquires them.

Observation sentences are *occasion* sentences, true only intermittently on the occasions of their utterance. It was in the subsequent course of evolution, to *standing* sentences, that we left our furry and feathered friends much farther behind and moved into language full blown. I despair of a plausible scenario of this development, but it issued in a rudimentary theory of nature. We came to recognize scenes that could be revisited at will, interspersed with experiences that recurred unbidden. Some of these could be predicted, thanks to induction, and some avoided.

We saw observation sentences as the child's port of entry to cognitive language. However, one goes on acquiring further ones at all ages. For a given speaker, an occasion sentence is an observation sentence, as I use the term, if he has learned to give it unhesitating assent under appropriate stimulation. The sentence can be one which in his younger days he would affirm only tentatively after deliberation and patient examination of a situation, but which in his subsequent expertise he can check at a glance.

Understood thus broadly, observation sentences serve two vital purposes. One is in the acquisition of one's language and the other is in the testing of one's theories about the world. The theory to be tested is expressed in standing sentences, but it is found that if a certain observation sentence is adjoined to those standing sentences, another observation sentence can be deduced. The experimenter then brings-about fulfillment of the one observation sentence and predicts fulfillment of the other. Thus it is that theory gains empirical content, and thus it is that theory supplements induction.

Our prehistoric theory must have had a logical structure in which the observation sentences and standing sentences were loosely linked by implication and lively expectation. I suppose the primitive early logic included the truth functions, for they call only for acquisition of the particles 'and' and 'not' or their equivalents. Learning logic at this level is just part of learning language. Affirmation of a contradiction of the form ' p and not p ' would simply bespeak misuse of the particle 'and' or 'not', like calling a spade a club. Failure to infer ' q ' from ' p ' and 'if p then q ' would betray mere failure to master the particle 'if', where 'if p then q ' is not (p and not q).

A vital early development, surely, was man's dissection of reality into objects grouped under general terms manipulated by quantifiers 'some' and 'every'. This I see as acknowledging existence of the objects, whatever ones there may be, that fall under the term. Seen from within the growing theory, it is discovery of objects. Seen from the point of view of theory construction, it is reification, the positing of objects. It is of vital importance to the construction of theory, for it complements the frail logic of truth functions with the powerful logic of the quantifiers 'some' and 'every'. This logic received only a partial and sketchy regimentation in Aristotle's monadic logic of the syllogism. It underwent explicit and complete formalization by Frege and others in the familiar polyadic logic of quantification.

The quantifiers 'some' and 'every' induce an inconspicuous exaltation of the pronoun. As its name suggests, the pronoun is thought of as a mere convenience to avoid repetition of some cumbersome expression that stands as its grammatical antecedent. When its antecedent bears a quantifier, however, the pronoun becomes essential: its replacement by its antecedent spoils the message. 'Someone lost his wallet' becomes puzzling when transcribed as 'Someone lost someone's wallet', and 'Everyone loves his mother' does not imply that everyone loves everyone's mother.

The essential pronoun appears in logical notation as the bound variable. Quantifiers are not the only variable-binding operators: there are also the operators of class abstraction, function abstraction, singular description, differentiation, and integration. But all these idioms go over into existential quantification under appropriate paraphrase, and into universal quantification under equivalent paraphrase. Thus it is that, as I have been saying for 58 years, to be is to be a value of a variable. A bound variable, or essential pronoun.

To appreciate the crucial role of reification in the structuring of theory, consider again the feeble truth-functional conditional 'if p then q ', or 'not (p and not q)'. Taken as a premise along with the premise ' p ', it enables us to infer ' q '. This is negligible progress. On the other hand ' $\forall x$ {if Fx then Gx }', the universal quantification of the truth-functional conditional, conjures up a veritable thicket of implications at a single stroke.

Mastery of time and tense was another major step in the development of language. Derek Bickerton speculates on it in his *Roots of Language*,¹ where he has drawn on 23 pidgin and creole languages for possible analogues of

1. Ann Arbor: Karoma, 1981.

early language. Pidgin arises from the spontaneous efforts of heterogeneous groups to communicate. It is a faltering jargon based on venturing one's own words and guessing at the interpretation of another's. Creole is the somewhat systematic development of pidgin by subsequent generations. These studies lead Bickerton to conjecture that the notion of cause came early, but the more general notion of past and future emerged only later. He wonders if this crucial development might account for the rise of agriculture and towns 9000 years ago.

Even before that breakthrough into the limitless past and future, there would have been recognition of temporal order within the specious present, and there would have been words like 'before' and 'after' to reflect it. Also of course a recurring acquaintance or landmark would be remembered and recognized after longer absences, for this is just the work of perceptual similarity, and shared by other animals. But the breakthrough to a limitless time dimension was quite another matter. It endowed our ontology with serious substance, as we shall see, and in so doing it vastly extended and reinforced the structure that man's theory of nature had already gained from reification within the specious present. Bickerton's coupling of it with the dawn of civilization, the rise of farms and villages, becomes plausible indeed.

Predictably. I shall illustrate these matters with my threadbare old example, "If a cat eats spoiled fish and sickens, it will thereafter avoid fish." The thought expressed here could not even be entertained until our gifted ancestor opened up the time dimension. The statement applies to every cat, known and unknown, but to each separately throughout the rest of its life. It imposes upon us the responsibility of making sense of sameness of cat over time. We meet the responsibility, in the case of cats and other bodies, by appealing to gradualness of change of form and substance. This answer presupposes a substantial theory of nature, involving the unseen trajectory of the cat or other body between one appearance and the next. We are suddenly committed to a theory of predominantly unobserved but temporally continuous bodies and substances.

These repay us by structuring our theory of the world in the time dimension. Causal regularities in the cat's largely unobserved career relate her behavior on one regrettable occasion to her avoidance movements on various occasions in subsequent years. The dynamics of cats and other bodies thus link times and places into a four-dimensional causal network that is our theory of the world.

Counting imposes further ontological refinement. In counting the members of the Drones Club we tick off the members and end up with a hundred, which is a trait of the club today. The word 'hundred' denotes classes, the Drones Club for one, which is the class of its current members. Classes are abstract objects, and 'hundred' is accordingly an abstract general term when used in counting.

Concrete objects, and preeminently bodies, are surely what were reified first in the development of language. What with our prey and other comestibles on the one hand and our predators on the other, bodies must have predominated as subject matter of most sentences, most quantifiers, most pronouns. I suppose they still do. To the layman they are the existent objects, the things *par excellence*.

The nominalists, medieval and modern, repudiated abstract objects and recognized only concrete ones. Abstract terms, however useful, were rated as syncategorematic, denoting nothing. The old nominalists went on using abstract terms, so what are we to make of their disclaimer? What counts, as reference, or existence?

Rather than await the nominalist's answer, let me speak for myself. Primarily the question belongs in the plural: the existence not of Socrates or Pegasus individually, but of lions or unicorns in general. It is a question of truth of ' $\exists x Fx$ '. Existence in the singular, then comes as the special case ' $\exists x(x = a)$ '. To be, once again: to be the value of a variable.

For ' $\exists x Fx$ ', finally, all we can ask for as a warrant is that the sentence be logically implied by our going theory of the world or of the subject matter with which we are concerned. The evidence for the theory, in turn, is as usual: its resistance to experimental tests, its success in prediction. Fallibility of course prevails.

We can still commend the nominalist's taste for economy, and even his taste for concrete objects. Let us pursue it as best we can. For nominalists and all of us, concrete objects comprise more than bodies; certainly any masses of matter. I favor a broad notion of concrete physical object; namely, the material content of any portion of space-time, however diffuse and discontinuous. In particular, then, not only does each member of the Drones club qualify as a physical object, but also any number of members, taken together, add up to a physical object in turn. The entire club can thus be construed as a physical object. It still qualifies as a hundred, but no longer as a class of a hundred members; rather as a physical object with a hundred separated but internally continuous parts.

This strategy can liberate biological taxonomy from its notorious dependence on classes. Each species, each genus, each order, in short each taxon, can be construed simply as the mereological sum of its members. Something then qualifies as a *member* of its species, rather than a mere part of it, if it is a separate but internally continuous part of it. Every taxon, like every club, can be construed as a physical object. However, the strategy has its limits. It does not work for a class whose members abut or overlap. It does not work for the class of the states of New England, nor of the counties of New England. The mereological sum of the states is identical with that of the counties: it is simply the whole of New England. But the class of the states has only six members, while that of the counties has 67.

Already at the humble level of counting, therefore, we are committed after all to classes as abstract objects. The needs of primitive commerce subsequently drive us past the whole numbers and into fractions, and the needs of engineering carry us on into the irrationals. Systematizing this, with an eye always to structural simplicity as well as ontological economy, we are driven to reifying our old whole numbers themselves as classes of classes. Advancing technology makes increasing demands, and we find ourselves countenancing classes of classes of classes and so on up without end.

In my youth I thought of the question of existence, or what there is, as perhaps the most basic question of philosophy and science. In the fullness of time the scales fell from my eyes. Any two ontologies are equally supported by all possible data if we can express a one-to-one correlation, what I call a proxy function, between them. We can just substitute the proxies for the old objects as values of the variables, and then compensate by reinterpreting all the predicates conversely. The double shift reinterprets ' x is an F ' as 'the proxy of x is the proxy of an F ', which is simply equivalent, since proxy is one to one.

The question what there is seems searchingly objective, but my present perspective brings the objective pole alarmingly close to subjective sources. The raw input that has been least subject to human creativity is the triggering of our sensory receptors. *Alles ubrige*, as Kronecker wrote in another connection, *ist Menschenwerk*.

The drive for ontological economy, that had been exercising not only nominalists but also the rest of us in our more temperate way, has now been sapped by the proxy function, for it leaves all ontologies on a par that are known to admit of one-to-one correlation with one another. Since

one-to-one correlation is the test of sameness of size, we are left with only reduction in size as the goal of ontological economy.

We observed that natural science calls for reification of numbers without end, so we are resigned to an infinite abstract ontology. But modern set theory apprises us of smaller and larger infinities, so we can still try to get by with the cosiest possible infinity.

Here we must note a divergence between the aims of the natural scientist and those of the pure mathematician. Basic mathematics is not only indispensable to natural science, but integral to it. The chunks of theory that imply experimentally testable consequences are apt to include purely mathematical truths, usually tacit, along with physical hypotheses; and the mathematical truths quantify freely over pure numbers. Thus far the natural sciences and the pure mathematician are seeing eye to eye. But then the mathematician may add further assumptions and explore them for their own sake, soaring into the empyrean of indenumerable infinities and inaccessible numbers. Here we must just recognize the divergence of concerns.

Subtleties even of natural science have driven theory beyond the reach of the denumerable infinite. The irrational numbers, indispensable in measurement, are indenumerable as now defined. In limning a philosophy of nature we might include just enough pure mathematics for the needs of natural science, and round it off by making the presumably more economical choice whenever an otherwise undecidable hypothesis such as the axiom of choice or the continuum hypothesis rears its head.

Hermann Weyl years ago, Hao Wang more recently, and today Solomon Feferman have concerned themselves with reworking the foundations of mathematics, notably of the real numbers, in such a way as to meet the needs of natural science without exceeding the denumerable, or lowest, infinity. Feferman is optimistic, but is still checking on how far those needs extend.

In this survey I have stayed in the mainstream of science, neglecting intensional and mentalistic usages. I have pictured science as couched in the syntax of modern logic, namely predication, truth functions, and quantification. This excludes intensions and ensures *extensionality*. What extensionality means is that if two predicates are true of just the same things then substitution of one predicate for the other anywhere in a true sentence always results in a true sentence. Extensionality, in short, is interchangeability of coextensive predicates *salvé veritate*.

Derivatively, objects are called extensional if admitting them as values of

variables in otherwise extensional discourse preserves extensionality. It is the one difference between clauses and properties: classes are extensional, properties not. Classes with just the same members are identical; properties may be felt to differ though they be true of just the same things. They just happen, we say, to be true of the same things. It comes down to the difference between essence and accident, whatever that is. There is no place for properties in our logic of truth functions and quantification, where extensionality prevails. There is no visible benefit to science in reifying properties: classes do the job.

The modal logic of necessity—necessary truth and necessary consequence—is an extravagant departure from extensionality. Necessity itself can instead be clearly and usefully construed in one and another relativistic way: relativized to logic, or to one or another deductive system of some special subject matter. But these matters can be handled for the most part in extensional terms by what I call *semantic ascent*: by rising from the sentences concerned to sentences about them. Instead of affirming a sentence with the adverb ‘necessarily’ prefixed to it, we name the sentence by quotation and then say *of it*. That it is a theorem of the system. Similarly, instead of joining two sentences in a necessary conditional, we say of the second sentence that it is logically deducible from the first and the special assumptions of the system. These expedients fail to accommodate some formulas of modal logic, but this is good riddance, for they are ones involving dubious iterations of necessity.

Violation of extensionality is rife in the idioms of propositional attitude: ‘believes that’, ‘hopes that’, ‘says that’, ‘regrets that’, and the rest. These, moreover, are not to be lightly dismissed. But I have seen my way to extensionalizing them in most of their forms, again by semantic ascent, and to finding the teat of them questionable anyway on other counts. But I must leave you to the last chapter of my recent thin book *From Stimulus to Science* for this.

— II —

Previously Published Articles

Relations and Reason

How Mathematical Logic Throws Switches
for Institute Engineers, Outlines Clauses
for Actuaries, and Provides Amusement
for the Puzzle-minded

*The Three Joiners of Berea and What They Did—
Subtlety That Becomes Simplicity as Symbols
Are Put to Work—Mathematics and Logic
Cöoperate to Devise New Methods
With Varied Technological Values*

1939

Can a man be his own uncle? half uncle? Can two men be each other's uncles? half uncles? If we construe "can" in the sense of physical possibility, we have four problems here; if we construe it in the sense of legal possibility, we have four more. These eight problems are simple, and with a certain amount of fumbling we can solve them. We can satisfy ourselves that two of the four relationships suggested are physically impossible, one is possible physically but not legally, and one is possible both physically and legally (Figure 13.1).

But these problems stand in marked contrast to the little algebraic exercises which treat of men rowing upstream and downstream. In the uncle questions there is nothing to call x , no quantity to solve for; we are depen-

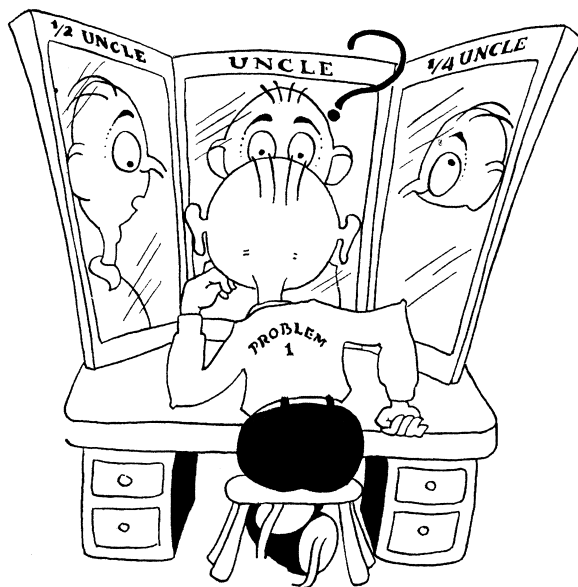


Figure 13.1.

dent on unimplemented thought. The difference between the two kinds of problems is not, however, a difference between the nonmathematical and the mathematical; the answers to the uncle problems are connected with the familiar principles of parentage and legal matrimony in a way which is just as mathematical, just as inexorably deductive, as the connection between the value of x and the data about current and rowing time. The difference is one, rather, between sections of mathematics—the section which treats of relations and the section which treats of numbers.

It is in the second of these sections that the science of mathematics happens to have developed most brilliantly—in arithmetic and the derivative disciplines. Even in geometry, which has no primary concern with number, the greatest advances of technique have been made by tapping the well-developed resources of numerical theory through the medium of analytical geometry. Where number is irrelevant, regimented technique is commonly lacking; our approach to the uncle problems is likely to be about as unmethodical as our approach to the rowboat problems would be if we had never studied algebra or arithmetic.

Thus it is that the progress of natural science has always depended so

largely upon the discernment of measurable quantity of one sort or another. Measurement in physics or any other science consists in correlating our subject matter with the series of real numbers. Such correlations are desirable because, once they are set up, all the well-worked theory of numerical mathematics lies ready at hand as a tool for our further reasoning.

But no science can rest entirely on measurement, and many scientific investigations are quite out of reach of that device. For those scientific problems to which measurement is not germane, techniques provided by the numerical parts of mathematics are useless. Either we must turn to those less known and less developed parts of mathematics which treat of non-numerical matters—relations, for instance—or else we must simply forego mathematical technique and proceed by unaided common sense, as in the uncle problems.

Among the nonnumerical bits of mathematical theory which have proved useful to natural science, perhaps the best known is the theory of groups. Where a set of movements or other operations is under consideration, the theory of groups provides a systematic technique for dealing with the effects of applying these operations one after another. But non-numerical techniques such as this have come to the scientist's aid only at too rare intervals.

It is on this score that the scientist stands to gain a good deal from the new mathematical logic. This discipline had its crude beginnings, eight decades ago, in Boole's effort to put Aristotle's logic of the syllogism in algebraic dress. The dress, as it turned out, was fitted with wings, and in the course of the next generation logic soared to heights of generality and effectiveness undreamed of by men who knew only the pedestrian logic of tradition. The new logic is indeed a systematization of the fundamental relationships from which *all* mathematics can be derived. From the basis which it provides we can proceed into arithmetic and the derivative disciplines by a series of definitions and elementary steps of inference. Equally, starting off from the same basis and working out our definitions and inferences in other directions, we can wind up in group theory or abstract algebra or any other of the countless parts of pure mathematics, known and unknown. To the scientist longing for nonquantitative techniques, then, mathematical logic offers relief in two ways: It provides explicit techniques for manipulating the simplest ingredients of discourse; and it provides a clear and systematic basis upon which to construct further techniques appropriate to such special scientific needs as may arise from time to time.

The simplest part of mathematical logic is the calculus of statements. Just as elementary arithmetic governs the manipulation of such connectives as $+$ and $\times$, so the statement calculus governs the manipulation of those connectives *and*, *or*, and so on, which are used in ordinary discourse for compounding sentences. Another operator of the statement calculus is *not*, which is, for convenience, written initially, contrary to grammatical usage: “It is not raining” becomes “not it is raining.”

The rules governing *and* and *or* resemble, in many respects, those which govern $\times$ and $+$ in arithmetic. For example, if we suppose p , q , and r replaced by any statements, it is apparent that $(p \text{ and } q) \text{ and } r$ is equivalent to $p \text{ and } (q \text{ and } r)$; that $p \text{ and } q$ is equivalent to $q \text{ and } p$; and similarly for *or*. The arithmetical law of “multiplying out” has its analogue in the statement calculus: $p \text{ and } (q \text{ or } r)$ is equivalent to $(p \text{ and } q) \text{ or } (p \text{ and } r)$. But in the statement calculus, unlike arithmetic, there is also a law of “adding out”: $p \text{ or } (q \text{ and } r)$ is equivalent to $(p \text{ or } q) \text{ and } (p \text{ or } r)$. Another convenient way in which the behavior of *and* and *or* diverges from that of $\times$ and $+$ is this: $p \text{ and } p$ amounts simply to p , and so does $p \text{ or } p$. For instance, replace p by a statement; “John is a man and John is a man” amounts simply to “John is a man.”

The behavior of *not* resembles that of the minus sign in arithmetic, at least to this extent: *not not p* is equivalent to p alone. But the arithmetical principles $(-x) \times (-y) = x \times y$ and $(-x) + (-y) = -(x + y)$ do not carry over. Instead we have a more symmetrical pair of principles, according to which $(\text{not } p) \text{ and } (\text{not } q)$ and $(\text{not } p) \text{ or } (\text{not } q)$ are equivalent respectively to $\text{not } (p \text{ or } q)$ and $\text{not } (p \text{ and } q)$.

Another simple part of mathematical logic—the part, indeed, which Boole himself began—is the calculus of classes. Where α and β are replaced by general terms “particle,” “man,” “skilled,” “employed,” and so on—each of which may be thought of as naming a single collection, or class, of objects—the notation $\alpha \cap \beta$ is used to denote the common part of the two classes. Thus *skilled* $\cap$ *employed* is the class of skilled job holders. Again, $\alpha \cup \beta$ is used to denote the total class of all those objects which fall into at least one of the classes α and β ; thus *skilled* $\cup$ *employed* is the class which comprises all skilled persons, employed and unemployed, together with all job holders, skilled and unskilled. The notation $\bar{\alpha}$ is used to denote the class of all objects which fall outside the class α . Thus $\bar{\text{skilled}}$ is the class of the unskilled. Now the principles governing $\cap$, $\cup$, and $\bar{}$ closely resemble those which govern *and*, *or*, and *not* in the statement calculus. Thus

$\bar{\alpha} \cap \bar{\beta} = \overline{\alpha \cup \beta}$, $\bar{\alpha} \cup \bar{\beta} = \overline{\alpha \cap \beta}$, $\alpha = \bar{\bar{\alpha}} = \alpha \cap \alpha = \alpha \cup \alpha$, $\alpha \cap \beta = \beta \cap \alpha$, $\alpha \cup \beta = \beta \cup \alpha$, $\alpha \cap (\beta \cup \gamma) = (\alpha \cap \beta) \cup (\alpha \cap \gamma)$, and so on. A relation which plays a prominent role is that of inclusion: α is *included in* β means that all members of α are members of β . This is written $\alpha \subset \beta$, which may be treated as an abbreviation of $\alpha = \alpha \cap \beta$.

A notion which is fundamental to class theory—though not explicit in Boole’s class calculus—is abstraction. The notation is $\hat{x}(\dots)$, where the blank is filled by any condition containing the variable x ; the whole denotes the class whose members are just the objects satisfying the condition in question. Thus $\hat{x}(x \text{ chews its cud})$ is the class of ruminants. Similarly, supposing an object y given, $\hat{x}(x = y)$ is the class of all objects x satisfying the condition $x = y$. In other words, it is the class whose sole member is y (Figure 13.2). This class, by the way, figures prominently in mathematical logic. It is called the unit class of y , symbolically ιy . Again, $\hat{x}(x = x)$ becomes the class of all self-identical things—in other words, simply the class of everything—commonly written V .

Another part of mathematical logic, an imposing part, is the theory of



“. . . the class whose sole member is y ”

Figure 13.2.

relations. One sample notion of relation theory is the distributive referent $R''\alpha$. Where R is any relation and α is any class, $R''\alpha$ is the class of all objects which bear R to one or more members of α . Thus, where R is the relation of parent to child and α is the class of criminals, $R''\alpha$ is the class of parents of criminals. Note that $R''ix$ becomes automatically the class of parents (two in number) of the individual x . $R''V$ becomes simply the class of parents.

Suppose α is a period of time, thought of as a class of moments, and R is the relation which a man bears to each moment in which he lives. It is readily seen, then, that $R''\alpha$ becomes the class of all men living at all in the period α ; $R''\bar{\alpha}$ becomes the class of men living at all outside the period α ; $\bar{R}''\alpha$ becomes the class of men who do not live clear through the period α ; $\bar{R}''\bar{\alpha}$ becomes the class of men not living through all time before and after α (hence, all men); $\overline{R''\alpha}$ becomes the class of men whose lives do not touch the period α ; $\overline{R''\bar{\alpha}}$ becomes the class of men whose lives are confined to α ; and so on.

Of the theorems of relation theory, one elementary illustration is this: $R''(\alpha \cup \beta) = (R''\alpha) \cup (R''\beta)$. But relation theory is useful also in making clear what principles do not hold. For example, it is not in general the case that $R''(\alpha \cap \beta) = (R''\alpha) \cap (R''\beta)$. The class of benefactors of crippled violinists does not coincide with the class of all those people who are at once benefactors of cripples and benefactors of violinists.

Mathematical logic is now becoming a major field of mathematical investigation. There is already a large body of literature, and production is accelerating. But the motive of this productivity has been, so far, almost wholly theoretical rather than technological: analysis of the foundations of mathematics. One question dealt with here is the question of minimum essential notions. For example, instead of countenancing both $+$ and $-$ in arithmetic, one could get along with just $-$, for $x + y$ can be viewed as an abbreviation of $x - ((x - x) - y)$. Now the question is: How far can such reduction be carried? It has long since appeared, mainly through the work of Frege, Peano, Dedekind, Whitehead, and Russell, that all notions of pure mathematics can be reduced by definitions to the notions of mathematical logic alone. The latter can be reduced, in turn, by further definitions to a slender bundle. In fact, all we need are two: inclusion and abstraction. Every statement of mathematical logic or arithmetic or function theory or abstract algebra or any other part of pure mathematics can be translated into a statement composed only of the notations $\subset$ and $\hat{x}(\dots)$, of inclusion

and abstraction, applied over and over. Such translation is not, indeed, convenient. The theorem $1 + 1 = 2$ would run to a length of some 217 typewritten feet when thus expanded, and the binomial theorem would reach from Pole to Pole. But the fact remains that inclusion and abstraction are the only notions ultimately presupposed by mathematics.

Further motives of mathematical logic have been reduction of the principles of mathematics to a minimum of axioms and deductive rules, discovery of mechanical procedures for testing the truth and falsehood of mathematical statements, inquiry into the scope and limits of mathematical methods of proof. In this last direction, a discovery has been made which ranks among the most sensational of modern times: Gödel has shown that it is impossible to frame a set of axioms and deductive rules adequate to the whole of mathematics, or adequate even to the whole of arithmetic! There will always be certain demonstrably indemonstrable arithmetical truths.

Despite the fact that mathematical logicians have hitherto been moved thus by theoretical aims, there is reason to expect that the elaborate systematizations which those aims have evoked may prove no less valuable in direct application to natural science. The yield of mathematical logic in this direction may be expected to consist largely in techniques for solving problems, but largely also in a contribution of rigor and clarity—a sharpening of the concepts of science. Such sharpening of concepts should yield practical benefits of two kinds: disclosure of hitherto hidden consequences of given scientific hypotheses and obviation of subtle errors which may stand in the way of scientific progress (Figure 13.3).

Hopes of this sort can be made to appear warranted, I believe, by discussion of a few fictitious instances of concept building. Suppose that in a sociological study occasion arises to treat the relation A of close acquaintanceship. It might prove adequate for given purposes to define A by the criterion of reciprocal use of first names. Thus xAy , x bears A to y , means that x and y call each other by their first names. Now perhaps we are interested in examining one or another feature of those groups of society—cliques, we might say—whose members are bound to one another by the relation A . But we do not want to define a clique simply as any class α satisfying the condition: (i) “All members of α bear A to one another.” For where a, b, c, d, e , and f all call one another by their first names, we should prefer not to regard a, b, c , and d as forming a clique; we should prefer to regard them as forming merely part of a clique to which e and f also be-

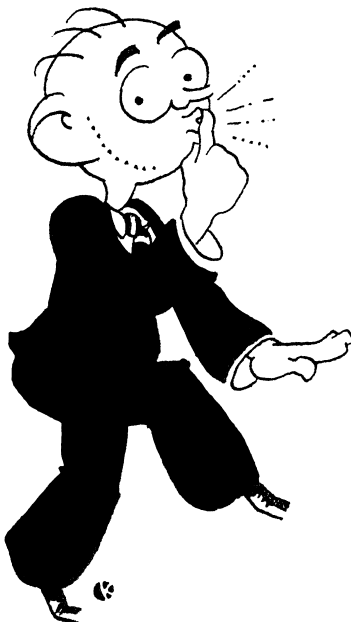


Figure 13.3.

long. Thus we will define a clique not as any class α satisfying (i) but as any largest class of the kind. The use of the word “largest” here is a bit vague. What is meant, however, is that the class must not exclude any persons whose inclusion would still be compatible with (i).

So far, little subtlety. This much by way of concept building might reasonably be expected without benefit of mathematical logic. Our definition of clique is, perhaps, somewhat off toward the periphery of intuitive logic, but the form of construction used is an important one, capable of application in other connections as well. Suppose we were to study the relation F of college fraternity brother rather than the relation A ; i.e., let xFy mean that x and y are fraternity brothers. Then what defined the notion of a clique comes now to define the notion of a college fraternity. A fraternity is any class α which satisfies the condition: (ii) “All members of α bear F to one another,” and which excludes no person whose inclusion would still be compatible with (ii).

Now to the subtlety. A third application of our form of construction would appear to be the following. Take L as the relation of lodge brother and define lodge as any class α which satisfies the condition: (iii) “All

members of α bear L to one another,” and which excludes no persons whose inclusion would still be compatible with (iii). Actually, however, this is wrong! Suppose that Elson, Mallow, and Ellow are citizens of Berea; that Elson is an Elk and a Mason, Mallow is a Mason and an Odd Fellow, and Ellow is an Elk and an Odd Fellow; and that no other Berean belongs to more than one lodge. Now let α be the class which has these three men as members. It is clear, to begin with, that α fulfills the condition (iii); Elson and Ellow are lodge brothers (through the Elks), so are Elson and Mallow (through the Masons), and so are Mallow and Ellow (through the Odd Fellows). Furthermore, α excludes no persons whose inclusion would still be compatible with (iii), for such a person would have to be a lodge brother of Elson, Mallow, and Ellow. To this end he would have to belong to at least two of the three lodges, but we know that there is no such joiner in all Berea apart from Elson, Mallow, and Ellow (Figure 13.4). We see, therefore, that our class of three men would constitute a lodge according to the definition proposed. The definition is thus a failure, for the three men do not actually constitute a whole lodge nor, indeed, even part of a lodge.

And why isn't our definition of fraternity subject to the same difficulty? The answer is that fraternities, unlike lodges, are mutually exclusive. This precludes any situation analogous to that which obtains among the lodges of Berea. The mutual exclusiveness of fraternities can also be phrased as a property of F , namely, transitivity; i.e., if xFy and yFz , then xFz . If x and y are fraternity brothers, and y and z are, then so are x and z ; for y cannot have been linked to x and z through different fraternities.

But there is a further subtlety; mutual exclusiveness, or transitivity, is not the whole story. Consider again the relation A and the cliques. Cliques are not mutually exclusive, nor is A transitive. From the facts that x and y call each other by their first names and that y and z do so, it by no means follows that x and z do so. Despite this, however, the proposed definition of clique remains sound. Now why should the construction work for clique as well as for fraternity yet fail for lodge? We see that the crux of the problem is not transitivity, or mutual exclusiveness. What it is may be left to the mathematical logician.¹

It is perhaps needless to say that mathematical logic does not leave these definitions in such awkward form. Relation theory enables us to transform

1. The essential problem under consideration has been pointed out and investigated by H. N. Goodman, in unpublished writings.



Figure 13.4.

them to terms whose manipulation follows clear laws. Transformed, the definition of clique (and similarly of fraternity) is as follows: A clique is any class α such that $\alpha = \hat{x}(\alpha \subset ix \cup A \text{“} ix)$. In words a clique is any class α whose members are all and only those objects x such that all members of α other than x bear A to x . But the symbols are easier than the words are.

By way of application of mathematical logic to the solution of specific technical problems, some striking cases have been turned up by Claude E. Shannon, a graduate student at Technology. In the statement calculus, the simplest part of mathematical logic, Shannon has found a tool for the solution of numerous problems in electrical engineering.² The efficacy of the tool can be made apparent with the help of Fig. 13.5, which shows a two-terminal circuit containing 14 positions for the making and breaking of contact. These positions are connected variously in series and in parallel as

2. C. E. Shannon, *Transactions of the American Institute of Electrical Engineers* 57 (1938): 713–723. I am indebted to this for essential features of the appended example.

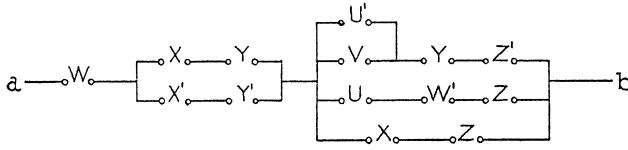
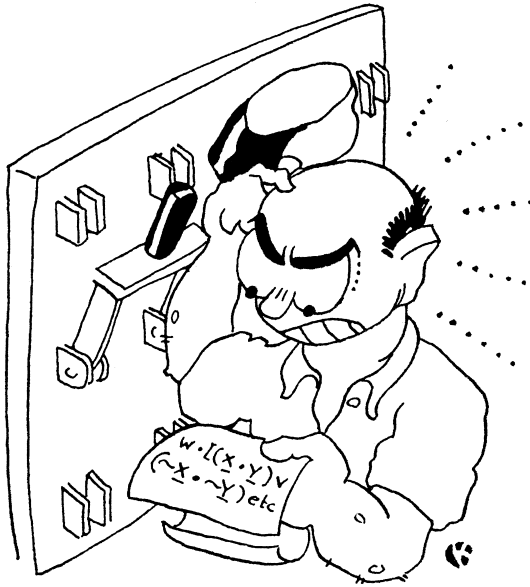


Figure 13.5.



Engineer throwing logical switches

Figure 13.6.

depicted, and each of them may be thought of as controlled from one or another of six hand switches u , v , w , x , y , and z , not appearing in the diagram. The contact marked U is made and the one marked U' is broken when the handle of switch u is up; the two contacts marked X are made and the one marked X' is broken when the handle of switch x is up; and so on. Now there are 2^6 , or 64, possible ways of leaving our six switch handles: none up, u alone up, v alone up, u and v alone up, and so on. From among these 64 situations just certain ones will close the whole circuit from a to b .

One problem which might be proposed new is the determination of the

particular switch situations which will close the circuit (Figure 13.6). A further problem, however, is this: What hookup simpler than that of Fig. 13.5 would yield the same result? What is the simplest hookup which would call for exactly the same switch situations to close the circuit as does the given hookup?

Now if current is to pass from a to b , clearly the contact marked W must be made. Again, either the first contact marked X and the first contact marked Y must both be made or else those marked X' and Y' must both be made, for the current must pass through one or other of these two routes. Examination of the right-hand half of Fig. 13.5 gives the further conditions that must be fulfilled, but let us defer these and consider the simple conditions so far derived: The W contact must be made, and either the first X contact and the first Y contact or else the X' contact and the Y' contact. This means that switch handle w must be up and x and y must both be up or both down. If we write W hereafter to mean " w is up," X to mean " x is up," and so on, our conditions can be put together thus: (iv) W and $[(X \text{ and } Y) \text{ or } (\text{not } X \text{ and } \text{not } Y)]$. Here, then, is the left segment of a statement of the total switch conditions that close the circuit.

But observe that (iv) stands in a neat correspondence to the left segment of Fig. 13.5 itself: *And* corresponds to series connection of contacts, *or* corresponds to parallel connection of contacts, and *not* corresponds to the accent. Simply by following this same principle of correspondence through the rest of Fig. 13.5, the formulation begun in (iv) can be finished without further thought: (v) $W \text{ and } [(X \text{ and } Y) \text{ or } (\text{not } X \text{ and } \text{not } Y)] \text{ and } \{[(\text{not } U \text{ or } V) \text{ and } Y \text{ and } \text{not } Z] \text{ or } (U \text{ and } \text{not } W \text{ and } Z) \text{ or } (X \text{ and } Z)\}$.

This tells us exactly the conditions, as regards switches, under which the circuit will be closed. Or rather, it does so if we are agile enough in reading out the piled up *and*'s, *or*'s, and *not*'s to grasp the net import of the whole complex statement—but this is a big order. However, the routine manipulations of the statement calculus enable us to transform (v) into more and more simple equivalent statements, step by step, until the net import does become obvious. Skipping the intermediate stages of these mechanical transformations, I will put down the result: (vi) $W \text{ and } X \text{ and } Y \text{ and } (V \text{ or } Z \text{ or } \text{not } U)$. In this form the condition on the switches is clear: The circuit will be closed just in case w , x , and y are up and either u is down or else v or z is up.

In finding this out we have incidentally solved our further problem of finding the simplest hookup which would have the effect of the original.

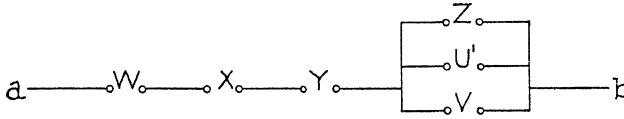


Figure 13.7.

For just as (v) was obtained from Fig. 13.5 by a simple rule of translation, so our reduced statement (vi) can be translated back into a diagram as Fig. 13.7. Here we see a hookup which uses just six contacts instead of 14, yet accomplishes the same result as the original. The machine for which the hookup of Fig. 13.5 was designed can now be built pretty cheaply.

Insurance is another field where an elementary level of mathematical logic has proved useful in solving practical problems. E. C. Berkeley of the Prudential Insurance Company has made extensive use of the class calculus in connection with the formulation and application of complex contracts.³ Not that the Prudential policies will be printed hereafter in logical symbols; the symbols merely make their contribution to insurance within the privacy of Mr. Berkeley's office.

Clause writing is a complicated matter. The regulations involved in the employees' pension plan of a large corporation may run to more than a dozen closely printed pages. Now suppose such a plan has finally been written out in a series of 50-odd laborious paragraphs covering all the special combinations of relevant circumstances that may arise on the part of the employees. What reason have we to believe that the whole could not have been formulated equivalently in half the space and with half the labor, simply by recombining and reclassifying the employees' circumstances along different lines? If there is such a short cut, common-sense thinking will never reveal it unless by luck, for the pension plan is too unwieldy an affair to revolve *in toto* in the mind. Mathematical logic, then, to the rescue: Berkeley schematizes his clauses in the formal notation of logic, then transforms the set of clauses systematically into the simplest possible equivalent by routine techniques of manipulation. The result goes back into English.

The usefulness of mathematical logic to insurance is not limited to the

3. E. C. Berkeley, *The Record* (American Institute of Actuaries), vol. 26, part 2, 373-414.

simplification of contracts. Berkeley uses it also in testing the consistency of a set of clauses; in determining the implications of a contract for one or another random set of circumstances; in distilling out the net differences between two closely related contracts; in all places, in short, where he detects value in being able to manipulate the terms of a contract with mathematical facility.

The usefulness of mathematical logic for Shannon and for Berkeley has resided in the application of prefabricated techniques to preformulated problems. The practical usefulness of a theory is not, however, to be appraised alone on the basis of such cases. We must allow the applicational needs themselves, rather, to play their part in motivating future elaborations of theory—just as the elaboration of tensor analysis and even of the differential calculus was motivated by the needs of physics. The history of mathematics has consisted to an important degree in such give-and-take between theory and application. Much of the value of mathematical logic for technology lies not merely in the direct applications of its already perfected techniques but in its potentialities as a basis from which to construct subsidiary techniques of unforeseen kinds in response to special needs.

On the Reasons for Indeterminacy of Translation¹

1970

My *gavagai* example has figured too centrally in discussions of the indeterminacy of translation. Readers see the example as the ground of the doctrine, and hope by resolving the example to cast doubt on the doctrine. The real ground of the doctrine is very different, broader and deeper.

Let us put translation aside for a while and think about physical theory. Naturally it is underdetermined by past evidence; a future observation can conflict with it. Naturally it is underdetermined by past and future evidence combined, since some observable event that conflicts with it can happen to go unobserved. Moreover many people will agree, far beyond all this, that physical theory is underdetermined even by all *possible* observations. Not to make a mystery of this mode of possibility, what I mean is the following. Consider all the observation sentences of the language: all the occasion sentences that are suited for use in reporting observable events in the external world.² Apply dates and positions to them in all combinations, without regard to whether observers were at the place and time. Some of these place-timed sentences will be true and the others false, by virtue simply of the observable though unobserved past and future events in the world. Now my point about physical theory is that physical theory is underdetermined even by all these truths. Theory can still vary though all possible observations be fixed. Physical theories can be at odds with each other and yet compatible with all possible data even in the broadest sense. In a word, they can be logically incompatible and empirically equivalent.

1. I am indebted to Burton Dreben for helpful criticism of an earlier draft of this paper.

2. The concept of observation sentence that I developed in §10 of *Word and Object* gains perhaps some further clarification in 85–89 of *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969).

This is a point on which I expect wide agreement, if only because the observational criteria of theoretical terms are commonly so flexible and fragmentary. People who agree on this general point need not agree as to how much of physical theory is empirically unfixed in this strong sense; some will acknowledge such slack only in the highest and most speculative reaches of physical theory, while others see it as extending even to common-sense traits of macroscopic bodies.

Now let us turn to the radical translation of a radically foreign physicist's theory. As always in radical translation, the starting point is the equating of observation sentences of the two languages by an inductive equating of stimulus meanings. In order afterward to construe the foreigner's theoretical sentences we have to project analytical hypotheses, whose ultimate justification is substantially just that the implied observation sentences match up. But now the same old empirical slack, the old indeterminacy between physical theories, recurs in second intension. Insofar as the truth of a physical theory is underdetermined by observables, the translation of the foreigner's physical theory is underdetermined by translation of his observation sentences. If our physical theory can vary though all possible observations be fixed, then our translation of his physical theory can vary though our translations of all possible observation reports on his part be fixed. Our translation of his observation sentences no more fixes our translation of his physical theory than our own possible observations fix our own physical theory.

The indeterminacy of translation is not just an instance of the empirically underdetermined character of physics. The point is not just that linguistics, being a part of behavioral science and hence ultimately of physics, shares the empirically underdetermined character of physics. On the contrary, the indeterminacy of translation is additional. Where physical theories *A* and *B* are both compatible with all possible data, we might adopt *A* for ourselves and still remain free to translate the foreigner either as believing *A* or as believing *B*.

Such choice between *A* and *B* in translation could be guided by simplicity. By imputing *B* to the foreigner we might come out with shorter and more direct translations, and with less in the way of elaborate contextual paraphrases, than by imputing *A* to him. That is one possibility. A second possibility is that both choices, *A* and *B*, require forbiddingly circuitous and cumbersome translation rules. In this case we might regard the foreigner as holding neither *A* nor *B*; we might attribute to him rather some

false physical theory which we can refute, or some obscure one which we despair of penetrating, or we might even regard him as holding no coherent physical theory at all. But we can imagine also, third, the possibility that *A* and *B* are both reasonably attributable. It might turn out that with just moderate circuitousness of translation at certain points—different points—*A* and *B* could be imputed about equally well. In this event no basis for a choice can be gained by exposing the foreigner to new physical data and noting his verbal response, since the theories *A* and *B* fit all possible observations equally well. No basis can be gained by interrogation in a theoretical vein, since the interrogation would take place in the foreigner's language and so could itself be interpreted according to either plan. In this event our choice would be determined simply by the accident of hitting upon one of the two systems of translation first.

The metaphor of the black box, often so useful, can be misleading here. The problem is not one of hidden facts, such as might be uncovered by learning more about the brain physiology of thought processes. To expect a distinctive physical mechanism behind every genuinely distinct mental state is one thing; to expect a distinctive mechanism for every purported distinction that can be phrased in traditional mentalistic language is another. The question whether, in the situation last described, the foreigner really believes *A* or believes rather *B*, is a question whose very significance I would put in doubt. This is what I am getting at in arguing the indeterminacy of translation.

My argument in these pages has been and will remain directed to you who already agree that there can be logically incompatible and empirically equivalent physical theories *A* and *B*. What degree of indeterminacy of translation you must then recognize, granted the force of my argument, will depend on the amount of empirical slack that you are willing to acknowledge in physics. If you were one of those who saw physics as empirically underdetermined only in its highest theoretical reaches, then by the argument at hand I can claim your concurrence in the indeterminacy of translation only of highly theoretical physics. For my own part, I think the empirical slack in physics extends to ordinary traits of ordinary bodies and hence that the indeterminacy of translation likewise affects that level of discourse. But it is important, for those who would not go so far, to note the graduation of liabilities.

Gavagai, whose troubles I shall now review, lay at an extreme of the scale. It was an observation sentence. Its stimulus meaning was inductively

well established, we supposed, coinciding with that of 'Rabbit'.³ Where indeterminacy threatened was in trying to settle upon the divided reference of *gavagai* as a term: whether rabbits or rabbit stages or undetached rabbit parts. Readers have responded with suggestions of how, with help of screens or other devices, we might hope to give the native informant an inkling of the desired distinctions and so settle the reference.

Ingenuity in this vein proves unrewarding because of vagueness of purpose. The purpose cannot be to drive a wedge between stimulus meanings of observation sentences, thereby linking *Gavagai* rather to 'Rabbit' than to 'Rabbit stage' or 'Undetached rabbit part'; for the stimulus meanings of all these sentences are incontestably identical. They comprise the stimulations that would make people think a rabbit was present. The purpose can only be to settle what *gavagai* denotes for the native as a term. But the whole notion of terms and their denotation is bound up with our own grammatical analysis of the sentences of our own language. It can be projected on the native language only as we settle what to count in the native language as analogues of our pronouns, identity, plurals, and related apparatus; and I urged in *Word and Object* that there would be some freedom of choice on this score. Once such choices are settled, on the other hand, however arbitrarily, the question whether the *gavagai* are rabbits or stages or parts can be settled too, by interrogation.

The most to hope for from the screens and kindred aids, then, is an indirect hint as to which of various analytical hypotheses regarding pronouns, identity, plurals, etc., might in the end work out most naturally. When this kind of hint is available, should we say that the supposed multiplicity of choices was not in fact open after all? Or should we say that the choice is open but that we have found a practical consideration that will help us in choosing? The issue is palpably unreal, and the doctrine of the indeterminacy of translation depends in no way upon it.

The *gavagai* example was at best an example only of the inscrutability of terms, not of the indeterminacy of translation of sentences. As sentence, *Gavagai* had a translation that was unique to within stimulus synonymy; for the occasion sentences 'Rabbit', 'Rabbit stage', and 'Undetached rabbit part' are stimulus-synonymous and holophrastically interchangeable. The

3. Strictly speaking, even this induction presupposes something like an analytical hypothesis in a small way: the decision as to what to take as signs of assent and dissent. See *Word and Object*, 80; also D. Davidson and J. Hintikka, ed. *Words and Objections* (Dordrecht: Reidel, 1968), 312, 317 or *Synthese*, xix. 1/2 (December 1968): 284, 289.

gavagai example had only this indirect bearing on indeterminacy of translation of sentences: one could imagine with some plausibility that some lengthy nonobservational sentences containing *gavagai* could be found which would go into English in materially different ways according as *gavagai* was equated with one or another of the terms 'rabbit', 'rabbit stage', etc. This whole effort was aimed not at proof but at helping the reader to reconcile the indeterminacy of translation imaginatively with the concrete reality of radical translation. The argument for the indeterminacy is another thing, as seen earlier in this paper.

Over the inscrutability of terms itself there is little room for debate. A clear example from real life was seen in connection with the Japanese classifiers.⁴ This example makes it pretty clear, moreover, that the inscrutability of terms need not always bring indeterminacy of sentence translation in its train, however the case may be in particular with *gavagai*. Again the questions raised by deferred ostension (*op. cit.*), e.g., as between expressions and their Gödel numbers, are strictly a matter of inscrutability of terms. This, not the indeterminacy of translation, is the substance of ontological relativity.

There are two ways of pressing the doctrine of indeterminacy of translation to maximize its scope. I can press from above and press from below, playing both ends against the middle. At the upper end there is the argument, early in the present paper, which is meant to persuade anyone to recognize the indeterminacy of translation of such portions of natural science as he is willing to regard as underdetermined by all possible observations. If I can get people to see this empirical slack as affecting not just highly theoretical physics but fairly common-sense talk of bodies, then I can get them to concede indeterminacy of translation of fairly common-sense talk of bodies. This I call pressing from above.

By pressing from below I mean pressing whatever arguments for indeterminacy of translation can be based on the inscrutability of terms. I suppose Harman's example⁵ regarding natural numbers comes under this head, theoretical though it is. It is that the sentence ' $3 \in 5$ ' goes into a true sentence of set theory under von Neumann's way of construing natural numbers, but goes into a false one under Zermelo's way. But a limitation of

4. *Ontological Relativity and Other Essays*, pp. 35f. Also in *The Journal of Philosophy*, LXV, 7 (April 4, 1968): 191ff.

5. Gilbert Harman, "An Introduction to Translation and Meaning," in *Words and Objections*, 14. Also in *Synthese*, *ibid.*, 14.

this example, as Harman recognizes, is that ' $3 \in 5$ ' rates as nonsense apart from set-theoretic explications of natural number.

In these pages I prefer not to speculate on how much better one might do from below, or from above either. My purpose here is to separate the issues and identify the arguments; and this may be managed most effectively by leaving the reader to consider what more might be proved.

Methodological Reflections on Current Linguistic Theory

1970

I want to make some broadly methodological remarks on a variety of issues. To begin with I'll talk of *rules*, and dwell a while on the distinction between *fitting* and *guiding*.

Imagine two systems of English grammar: one an old-fashioned system that draws heavily on the Latin grammarians, and the other a streamlined formulation due to Jespersen. Imagine that the two systems are *extensionally equivalent*, in this sense: they determine, recursively, the same infinite set of well-formed English sentences. In Denmark the boys in one school learn English by the one system, and those in another school learn it by the other. In the end the boys all sound alike. Both systems of rules *fit* the behavior of all the boys, but each system *guides* the behavior of only half the boys. Both systems *fit* the behavior also of all us native speakers of English; this is what makes both systems correct. But neither system guides us native speakers of English; no rules do, except for some intrusions of inessential schoolwork.

My distinction between fitting and guiding is, you see, the obvious and flat-footed one. Fitting is a matter of true description; guiding is a matter of cause and effect. Behavior *fits* a rule whenever it conforms to it; whenever the rule truly describes the behavior. But the behavior is not *guided* by the rule unless the behavior knows the rule and can state it. This behavior *observes* the rule.

But now it seems that Chomsky and his followers recognize an intermediate condition, between mere fitting and full guidance in my flat-footed sense of the word. They regard English speech as in some sense rule-*guided* not only in the case of the Danish schoolboys, but also in our own case, however unprepared we be to state the rules. According to this doctrine,

two extensionally equivalent systems of grammatical rules need not be equally correct. The right rules are the rules that the native speakers themselves have somehow implicitly in mind. It is the grammarian's task to find the right rules, in this sense. This added task is set by demanding not just any old recursive demarcation of the right totality of well-formed sentences, but rather a recursive demarcation of the right totality of trees. The trees used to be mere *ad hoc* scaffolding by the aid of which the grammarians, each in his own way, contrived to specify the objective totality of well-formed sentences. According to the new doctrine, the trees are themselves part of the objective linguistic reality to be specified.

We have all known that the native speaker must have acquired some recursive habit of mind, however unconscious, for building sentences in an essentially treelike way; this is evident from the infinitude of his repertoire. We can all go this far with Postal when, in his review of Dixon, he writes:

The claim that there are linguistic rules is simply the claim that individuals know their language and have not learned each of its sentences separately.¹

His word 'claim', even, seems ill suited to anything so uncontroversial. What is more than trivial, in the new doctrine that I speak of, is rather the following: it imputes to the natives an unconscious preference for one system of rules over another, equally unconscious, which is extensionally equivalent to it.

Are the unconscious rules the same, even, from one native speaker to the next? Let us grant that the generated infinitude of well-formed sentences is itself the same for two natives. There may then seem to be a presumption of sameness of generating rules—just because any appreciably different but extensionally equivalent system of rules is apt to be prohibitively complex and artificial. However, this suggestion gets us nowhere. Insofar as it is true, the grammarian can just follow his old plan, after all, of settling for *any* system of rules, naturally the simpler the better, that demarcates the right infinite set of well-formed sentences. If the new doctrine of the grammarian's added burden has any content, it owes it to there being appreciably unlike and still comparably manageable systems of rules for generating the same infinite totality of well-formed sentences. From experiences with axiom systems in mathematics, incidentally, we can easily believe in the ex-

1. Paul Postal, review of Dixon, *Language*, 84–93, specifically 88.

istence of such alternatives. In my parable of the Danish schoolboys I have already assumed the existence of just such alternative systems for English; though it should of course be said, if we are to be fussy about the facts, that Jespersen's grammar and that of the old-fashioned textbooks really fall short of extensional equivalence at some points.

We see then that the new doctrine of the grammarian's added burden raises the problem of evidence whereby to decide, or conjecture, which of two extensionally equivalent systems of rules has been implicitly guiding the native's verbal behavior. Implicit guidance is a moot enough idea to demand some explicit methodology. If it is to make any sense to say that a native was implicitly guided by one system of rules and not by another extensionally equivalent system, this sense must link up somehow with the native's dispositions to behave in observable ways in observable circumstances. These dispositions must go beyond the mere attesting to the well-formedness of strings, since extensionally equivalent rules are indistinguishable on that score. It could be a question of dispositions to make or accept certain transformations and not others; or certain inferences and not others.

Certainly I have no quarrel with dispositions. Nor do I question the notion of implicit and unconscious conformity to a rule, when this is merely a question of fitting. Bodies obey, in this sense, the law of falling bodies, and English speakers obey, in this sense, any and all of the extensionally equivalent systems of grammar that demarcate the right totality of well-formed English sentences. These are acceptably clear dispositions on the part of bodies and English speakers. The sticking point is this Chomskian midpoint between rules as merely fitting, on the one hand, and rules as real and overt guides on the other; Chomsky's intermediate notion of rules as heeded inarticulately. It is a point deserving of close methodological attention.

Ironically these same linguists have expressed doubt about the relatively clear and humdrum notion of a disposition to verbal behavior. Chomsky writes:

Presumably, a complex of dispositions is a structure that can be represented as a set of probabilities for utterances in certain definable "circumstances" . . . But it must be recognized that the notion "probability of a sentence" is an entirely useless one . . . On empirical grounds, the probability of my producing some given sentence of English . . . is indistin-

guishable from the probability of my producing a given sentence of Japanese.²

I am puzzled by how quickly he turns his back on the crucial phrase “in certain definable ‘circumstances.’” Solubility in water would be a pretty idle disposition if defined in terms of the absolute probability of dissolving, without reference to the circumstance of being in water. Weight would be a pretty idle disposition if defined in terms of the absolute probability of falling, without reference to the circumstance of removal of support. Verbal dispositions would be pretty idle if defined in terms of the absolute probability of utterance out of the blue. I, among others, have talked mainly of verbal dispositions in a very specific circumstance: a questionnaire circumstance, the circumstance of being offered a sentence for assent or dissent or indecision or bizarreness reaction.

Chomsky’s nihilistic attitude toward dispositions is the more puzzling in that I find it again in the newspaper account of his recent lectures in England, despite an intervening answer of mine³ to the earlier statement. I seem to detect an echo of it also in a footnote in Postal’s review of Dixon.⁴ This rejection of dispositions would be bewildering by itself. It is doubly so when contrasted with the rather uncritical doctrine just previously considered—the doctrine of unconscious preferences among extensionally equivalent grammars. I’d like to think that I am missing something.

Now some more remarks on the task of the grammarian. What I have said suggests, too simply, the following notion of the grammarian’s classical task: that it is the task of demarcating, recursively and in formal terms, the infinite totality of the well-formed strings of phonemes of the chosen language. It would seem from my remarks up to now that this is the basic or classical task, which, then, is added to if one insists further on some distinction between right and wrong rules, right and wrong trees subtending this same superficial mass of foliage. The trouble with thus stating the basic or classical task is that it presupposes some prior behavioral standard of what, in general, to aspire to include under the head of well-formed strings for a given community. What are the behavioral data of well-formedness? Passive observation of chance utterances is a beginning. The grammarian

2. Noam Chomsky, “Quine’s Empirical Assumptions,” *Synthese* 19 (1968): 53–68, specifically 57.

3. W. V. Quine, “Replies,” *Synthese* 19 (1968): 264–321, specifically 280.

4. Paul Postal, *op. cit.*, note 12.

can extrapolate this corpus by analogical construction, and he can test these conjectures on an informant to see if they elicit only a manifestation of bewilderment. But of course the grammarian settles for no such criterion. Traditionally, at any rate, the grammarian has accepted wide ranges of sentences as grammatical which an informant would reject as bizarre. I think of sentences such as Carnap's example, 'This stone is thinking about Vienna.'

A more realistic characterization of the grammarian's classical task is an open-ended one. He does not have a prior behavioral criterion of well-formedness; he just has some sufficient behavioral conditions. Strings heard from natives count as well-formed, at least provisionally. So do sentences which, when tried on an informant, elicit casual and unbewildered responses. What I then picture the grammarian as doing is to devise as simple a formal recursion as he can which takes in all these confirmably well-formed strings and excludes all strings that would bring really excessive bizarreness reactions. He rounds out and rounds off his data. Sometimes of course he will even reject a heard string as ill-formed, thus rejecting a datum, if he can appreciably simplify his system in so doing; but it would be regrettable to do much of this.

In this somewhat melancholy version of the grammarian's task, I have held Chomsky's doctrine in abeyance. Chomsky believes that the linguistic community itself has a sense of grammaticality which the grammarian can and should uncover; that grammaticality is not just the grammarian's rounding off of performance data. Up to a point I agree; the native's disposition to bizarreness reactions is an implicit sense of grammaticality of a sort. But Chomsky would of course credit the native with a full and precise sense of grammaticality, this being of a piece with the native's purported fund of tacit rules—the native's purported bias even among extensionally equivalent grammars. Now this doctrine is interesting, certainly, if true; let me only mention again the crying need, at this point, for explicitness of criteria and awareness of method.

An attitude that is closely linked to this doctrine is a readiness to recognize linguistic universals. The problem of evidence for a linguistic universal is insufficiently appreciated. Someone says, let us suppose, that the subject-predicate construction occurs in all the languages he has examined. Now of course all those languages have been translated, however forcibly, into English and *vice versa*. Point, then, in those languages to the translations of the English subject-predicate construction, and you estab-

lish the thesis; the subject-predicate construction occurs in all those languages. Or is it imposed by translation? What is the difference? Does the thesis say more than that basic English is translatable into all those languages? And what does even this latter claim amount to, pending some standard of faithfulness and objectivity of translation?

To make proper sense of the hypothesis that the subject-predicate construction is a linguistic universal, we need an unequivocal behavioral criterion of subject and predicate. It is not enough to say that if we take these and these as subjects and those and those as predicates then there are ways of so handling the rest of the language as to get general English translations. The trouble is that there are extensionally equivalent grammars. Timely reflection on method and evidence should tend to stifle much of the talk of linguistic universals.

Insofar, on the other hand, as one is prepared to impute to the native a specific and detailed though inarticulate grammatical system, one is apt to conceive of the notions of subject and predicate and similar notions as objective and as unequivocally apprehended by the native himself. To conceive of them thus is no more of a strain, surely, than to suppose that the native favors one of two extensionally equivalent grammars over another. In all this there is no folly, I feel sure, that conscientious reflection on method and evidence cannot cure; but the cure is apt to take time.

I think it is instructive, before leaving this topic, to fit an idea of Geach's into the picture. Besides singling out the well-formed strings, Geach argues, our grammar must distinguish between proper and spurious components of well-formed strings. One of his examples of a spurious component was 'Plato was bald' in the context 'The philosopher whose most eminent pupil was Plato was bald.'⁵ This demand is reminiscent of Chomsky's demand that the grammarian show how to generate not only the well-formed strings but the right trees. Yet Geach is not committed to finding a bias in the native community between extensionally equivalent grammars. I expect Geach's demand is reconcilable even with the humdrum view of the grammarian's task as the task merely of generating the well-formed strings; for the thing that Geach demands, the marking of the proper components of each well-formed string, would doubtless be a valuable auxiliary to the rules for generating further well-formed strings. The same case can be

5. Peter Geach, "Logical Procedures and the Identity of Expressions," *Ratio* 7 (1965): 199–205, specifically 201.

made, more generally, for Chomsky's insistence that the grammarian's proper product is the whole tree rather than just the well-formed strings that it issues in. The argument is simply that rules for generating further well-formed strings (and trees) can then be formulated in terms of past trees and not just past well-formed strings. This is a strong argument, and it does not depend on any obscure doctrine to the effect that the natives tacitly prefer one system of grammar to another that is extensionally equivalent to it. It would be well to sort out these motives and benefits and see whether the obscure points of doctrine might not be cheerfully dropped.

Such an inquiry could, I suppose, convince us that there is indeed an unarticulated system of grammatical rules which is somehow implicit in the native mind in a way that an extensionally equivalent system is not. For me such a conviction would depend in part upon clarification of criteria.

To get down more nearly to cases, suppose again a language for which we have two extensionally equivalent systems of grammar; two extensionally equivalent recursive definitions of well-formed string. According to one of these systems, the immediate constituents of a certain sentence are 'AB' and 'C'; according to the other system they are 'A' and 'BC'. The enigmatic doctrine under consideration says that one of these analyses is right, and the other wrong, by tacit consensus of native speakers. How do we find out which is right?

An unimaginative suggestion might be: ask the natives. Ask them, in their language, whether the real constituents of 'ABC' are 'AB' and 'C'. Does this pose an embarrassing question of translation? Well, then let the native language be English. The essential problem remains; we do not really understand our own English question. We are looking for a criterion of what to count as the real or proper grammar, as over against an extensionally equivalent counterfeit. We are looking, in the specific case, for a test of what to count as the real or proper constituents of 'ABC', as against counterfeit constituents. And now the test suggested is that we ask the native the very question which we do not understand ourselves: the very question for which we ourselves are seeking a test. We are moving in an oddly warped circle.

Better and more imaginative suggestions may be forthcoming for determining, less directly, what to regard as the real constituents of 'ABC' from the point of view of tacit native grammar. I suggested some time ago that it could be a question of dispositions to make or accept certain transforma-

tions or inferences. But I want now to make use of the unimaginative suggestion as a point at which to take off on a tangent, leaving at last this whole question of a native bias toward one of two extensionally equivalent grammars.

The unimaginative suggestion was: ask the natives. The same question, and the same warped circle or one very much like it, are encountered from time to time in semantics. People like me challenge the notion of synonymy and ask for a criterion. What is synonymy? How do you tell whether two expressions are synonymous? Ask the natives. This essentially was Arne Næss's answer some decades ago, as I analyze it.⁶ Moreover he suited the action to the word, disseminating questionnaires and claiming significantly uniform results. This was also essentially the answer more recently of Fodor and Katz,⁷ as I analyze it; and I have sensed suggestions of it in Chomsky. Now a reason for pausing over this oddly warped circle is that an empirical investigation, however odd, that yields uniformities has a claim to attention. Grant for the sake of argument that Næss's questionnaire on synonymy yielded statistically significant uniformities; what do they mean? Do they show that Næss's laymen are pretty much alike on the score of their synonymy pairs, obscure though it be to us wherein synonymy consists? Do they show something also, or instead, about how Næss's laymen use the obscure word 'synonymy' or its paraphrases? Separation of these components presents an odd problem.

Essentially the same question is raised outside linguistics by work of Smith Stevens on subjective magnitudes.⁸ For years he gathered subjective testimony of the pitch and loudness of sounds: whether this was twice as high as that, or half again as loud as that. He plotted these findings against the physical frequencies and volumes, and came out with significant correlations—not linear, but logarithmic. Significant, but of what? Was it uniformity of error in his subjects' effort to estimate physical frequency and volume? Or was it uniformity of subjective experience, coupled with uniformity of meaning attached to enigmatically subjective expressions like 'twice as high' and 'half again as loud'? Or did the subjective experience vary from subject to subject, while the meaning attached to the subjective expressions varied in a compensatory way? The uniformities surprise me

6. Arne Næss, *Interpretation and Preciseness* (Oslo: Dybwad, 1953).

7. Jerry Fodor and Jerrold Katz, "The Structure of a Semantic Theory," *Language* 39 (1963): 170–210.

8. S. S. Stevens, "On the Psychophysical Law," *Psychological Review* 64 (1957): 153–181.

and I am prepared to find them instructive, but I am at a loss to sort them out. It is the same warped circle.

Turning back to synonymy, or to the semantical notion of analyticity which is interdefinable with synonymy, I might mention also a questionnaire experiment which avoided the warped circle. Apostel and others⁹ in Geneva compiled various lists of sentences. One list contained only sentences that the experimenters regarded as analytic. Other lists had varied and irrelevant motifs. Subjects were given these lists, untitled, and were asked to sort various further sentences into the appropriate lists. The experiment, much the same as one proposed more recently by Katz,¹⁰ sought evidence of a felt similarity among analytic sentences, without benefit of title. The outcome was reported as at best indecisive.

A controversy over semantical notions has simmered for twenty years. Some of us have criticized these notions as insufficiently empirical. Others have defended the notions without improving them. Their defense has been visibly motivated by a sense of the indispensability of these notions in various applications. We would have been spared much of this rearguard action if the defenders of semantical notions had taken the criticism of these notions to heart, and sought seriously to get along without them. In one, certainly, of its most conspicuous applications the notion of synonymy is not needed; namely, in the definition of the phoneme. According to the familiar definition, what shows that two sounds belong to distinct phonemes is that the substitution of one for the other changes the meaning of some expression. Surely, however, meaning enough for this purpose is afforded by the innocent and uncontroversial notion of stimulus meaning.

The behavioral definition of stimulus meaning is as follows, nearly enough: the stimulus meaning of a sentence, for a given speaker, is the class of all stimulatory situations in the presence of which he will assent to the sentence if queried. Stimulus meaning is at its best among observation sentences. The behavioral definition of an observation sentence is as follows: an observation sentence is a sentence whose stimulus meaning is the same for just about all speakers of the language. Examples: 'It is raining', 'This is red', 'This is a rabbit'.

Sameness of stimulus meaning is no appreciable approximation to the

9. L. Apostel, W. Mays, A. Morf, and J. Piaget, *Les liaisons analytiques et synthétiques dans le comportement du sujet* (Paris: Presses Universitaires, 1937).

10. Jerrold Katz, "Some Remarks on Quine on Analyticity," *Journal of Philosophy* 64 (1967): 36–52.

general notion of synonymy to which semantics has aspired. Within observation sentences, however, sameness of stimulus meaning is synonymy enough. For distinguishing phonemes, consequently, it is enough; for surely, if two sounds belong to distinct phonemes, the meaning of some observation sentences will be changed by the substitution.

For that matter, phonemes can also no doubt be distinguished by appealing merely to well-formedness of expressions; by appealing, that is, to the capacity of a string of sounds to occur in the native stream of speech. Presumably, if two sounds belong to distinct phonemes, the substitution will render some coherent string of sounds incoherent. This way of defining the phoneme was proposed by Anders Wedberg,¹¹ and was already implicit, I think, in Zellig Harris. I wanted to bring in the definition in terms of stimulus meaning, however, as an example of how stimulus meaning can sometimes do the work that is desired of meaning or synonymy.

I turn, for the remainder of my remarks, to the notion of deep structure and its relation to logical analysis. Take, first, logical analysis. What do we do when we paraphrase a sentence by introducing logical symbols for truth functions and quantifiers? In principle it is the same as when in high school algebra we were given some data about rowing up and down a river; we paraphrased the data into algebraic equations, with a view to solving these for the speed of the river. In principle it is the same also as programming a computer.

I find the phrase 'logical analysis' misleading, in its suggestion that we are exposing a logical structure that lay hidden in the sentence all along. This conception I find both obscure and idle. When we move from verbal sentences to logical formulas we are merely retreating to a notation that has certain technical advantages, algorithmic and conceptual. I mentioned the analogy of the computer; but essentially the same thing is happening in a more moderate way when in natural history we switch to the Latin binominals for genera and species, or when in relativity physics we paraphrase our temporal references into a spatial idiom using four dimensions. No one wants to say that the binominals of Linnæus or the fourth dimension of Einstein or the binary code of the computer were somehow implicit in ordinary language; and I have seen no more reason to so regard the quantifiers and truth functions.

11. Anders Wedberg, "On the Principles of Phonemic Analysis," *Ajatus* 26 (1964): 235–253.

What now of deep structure? If we believe that native speakers have a detailed though inarticulate grammatical system, specific even as between extensionally equivalent systems, then certainly we believe that deep structure, whatever there may be of it, is there to be uncovered. How to tell whether we are getting it right, whether we are matching the inarticulate native analysis or just carving out an extensional equivalent, is a methodological question that I have mentioned already.

If on the contrary we hold every grammar to be as authentic as every extensionally equivalent grammar, and to be preferred only for its simplicity and convenience, then deep structure loses its objectivity but need not lose its place. Deep structure, and the transformations to and from it, might still qualify as auxiliaries to the simplest and most convenient system we know for demarcating the class of well-formed strings. They would stay on in this role just as the trees would stay on, and Geach's discrimination of proper and improper ingredients.

Thus conceived, the grammarian's deep structure is similar in a way to logical structure. Both are paraphrases of sentences of ordinary language; both are paraphrases that we resort to for certain purposes of technical convenience. But the purposes are not the same. The grammarian's purpose is to put the sentence into a form that can be generated by a grammatical tree in the most efficient way. The logician's purpose is to put the sentence into a form that admits most efficiently of logical calculation, or shows its implications and conceptual affinities most perspicuously, obviating fallacy and paradox.

These different purposes, the grammarian's and the logician's, are not in general best served by the same paraphrases; and for this reason the grammarian's deep structure is not to be identified with logical structure, suggestive though the one may be for the other. I have two major examples in mind to bring out the divergence.

One example is the elimination of singular terms other than variables. Let '*a*' represent such a singular term—perhaps a proper name, perhaps a complex singular term—and let '*Fa*' represent a sentence containing it. We can paraphrase '*Fa*', to begin with, as ' $(\exists x)(Fx \cdot a=x)$ '. In this way all singular terms, other than simple variables such as the '*x*' here, can be confined to one specific manner of occurrence: occurrence to the left of '='. Then, as a next step, we can reckon this identity sign to the singular term as an invariable suffix, thus re-parsing the singular term as a general term or predicate.

The advantages of this transformation are specific and limited. Laws of logic become simplified, through not having to provide for the instantiation of quantifications by terms other than variables. The simplification is the greater for the fact that the instantiations thus avoided were ones that depended awkwardly on existence assumptions. Certain gains in philosophical clarity ensue also. Variables, rather than names, come to be seen as the primary avenue of reference. Little puzzles about names that fail to name anything are swept aside.

This elimination of singular terms is not all good, however, even for logic and mathematics. Inference moves faster when we can instantiate quantifications directly by names and complex singular terms, rather than working through the variables and paraphrases. And complex singular terms are in practice vital for algebraic technique. An algebraist who was not free to substitute complex expressions directly for variables, or to substitute one side of a complex equality directly for the other, would soon give up.

The important point thus emerges that logical analysis itself—better, logical paraphrase—may go one way or another depending on one's specific logical purpose. The image of exposing an already present logical structure by analysis is a poor one. And when our interest turns to English grammar, again we are bound to find that the elimination of singular terms is to no purpose. Surely it yields no deep structure that would help to simplify an account of English grammar. Thus take the distinction between the referential and the non-referential use of singular terms. Work of Geach¹² and Strawson¹³ suggests that this distinction is vital to an appreciation of English; but the logical paraphrase obliterates it utterly.

In my view the logical structure and the deep structure, or let me say the logician's paraphrases and the grammarian's paraphrase, differ not in kind but in detail and purpose. They differ in the same sort of way that the logician's two paraphrases differ from each other: one the austere and pellucid paraphrase containing no singular terms but variables, and the other the algorithmically efficient paraphrase bristling with complex singular terms.

The elimination of singular terms was one example of the difference between paraphrasing for logic and paraphrasing for grammar. Now the other example I have in mind is the treatment of time as a fourth dimen-

12. Peter Geach, *Reference and Generality* (Ithaca, NY: Cornell University Press, 1962).

13. P. F. Strawson, "Singular Terms and Predication," *Synthese* 19 (1968): 97–117.

sion. A while ago I referred this to physics, but it is vital equally for logic and philosophy. A logic of tense is a towering triviality which we have no excuse to put up with if our concern is merely with the scientific use of language rather than with the scientific study of it. We program language into the simple neo-classical logic of truth functions and quantifiers, by eliminating tense and treating times on a par with places. The resulting simplification of formal logic may be sensed from this example, which I have used before: George V married Queen Mary; Queen Mary is a widow; therefore George V married a widow. We cease to have to provide against this kind of thing, among others.

Philosophical clarification ensues as well. Thus consider the following puzzles. How can things be related that do not coexist at any one time? How can a variable range now over things that no longer exist? or range ever over things that never coexist? How can a class have members that never coexist? How can a class, which is an abstract object, be said to change, as it must when its members change or cease to exist? We make a clean sweep of all such puzzles by dropping tense and treating all past, present, and future bodies as four-dimensional substances tenselessly scattered about in space-time.

This is a paraphrase which, we see, works wonders for logic, philosophy, and physics as well, but presumably is not wanted for English grammar. A deep structure without tense seems unpromising, at any rate, as a means of simplifying a grammatical account of an Indo-European language. Here again, evidently, is a wide divergence between the structure that the logician is after and what the grammarian wants under the head of deep structure. And yet, reading Postal's typescript 'Coreferentiality and physical objects,' I begin to wonder whether the four-dimensional view might be useful sometimes in grammar too.

My previous example, the elimination of singular terms, spoke for pluralism not just as between logical structure and grammatical deep structure, but within logical structure; one logical paraphrase served one logical purpose, another another. Perhaps now there is a case also for pluralism within grammatical deep structure: one paraphrase might serve one grammatical purpose, another another. A paraphrase into the tenseless idiom of four dimensions might play an auxiliary role in connection with some grammatical twists, while a different deep structure, retaining tense, might still be exploited for other grammatical ends. So let me conclude with a plea against absolutism.

On Empirically Equivalent Systems of the World

1975

If all observable events can be accounted for in one comprehensive scientific theory—one system of the world, to echo Duhem's echo of Newton—then we may expect that they can all be accounted for equally in another, conflicting system of the world. We may expect this because of how scientists work. For they do not rest with mere inductive generalizations of their observations: mere extrapolation to observable events from similar observed events. Scientists invent hypotheses that talk of things beyond the reach of observation. The hypotheses are related to observation only by a kind of one-way implication; namely, the events we observe are what a belief in the hypotheses would have led us to expect. These observable consequences of the hypotheses do not, conversely, imply the hypotheses. Surely there are alternative hypothetical substructures that would surface in the same observable ways.

Such is the doctrine that natural science is empirically under-determined; under-determined not just by past observation but by all observable events. The doctrine is plausible insofar as it is intelligible, but it is less readily intelligible than it may seem. My main purpose in this paper is to explore its meaning and its limits.

This doctrine of empirical under-determination is not to be confused with holism. It is holism that has rightly been called the Duhem thesis and also, rather generously, the Duhem-Quine thesis. It says that scientific statements are not separately vulnerable to adverse observations, because it is only jointly as a theory that they imply their observable consequences. Any one of the statements can be adhered to in the face of adverse observations, by revising others of the statements. This holism thesis lends credence to the under-determination theses. If in the face of adverse observa-

tions we are free always to choose among various adequate modifications of our theory, then presumably all possible observations are insufficient to determine theory uniquely

The holism thesis is less beset with obscurities than the underdetermination thesis, and again it is a thesis that must command assent, with reservations. One reservation has to do with the fact that some statements are closely linked to observation, by the process of language learning. These statements are indeed separately susceptible to tests of observation; and at the same time they do not stand free of theory, for they share much of the vocabulary of the more remotely theoretical statements. They are what link theory to observation, affording theory its empirical content. Now the Duhem thesis still holds, in a somewhat literalistic way, even for these observation statements. For the scientist does occasionally revoke even an observation statement, when it conflicts with a well attested body of theory and when he has tried in vain to reproduce the experiment. But the Duhem thesis would be wrong if understood as imposing an equal status on all the statements in a scientific theory and thus denying the strong presumption in favor of the observation statements. It is this bias that makes science empirical.

Another reservation regarding the Duhem thesis has to do with breadth. If it is only jointly as a theory that the scientific statements imply their observable consequences, how inclusive does that theory have to be? Does it have to be the whole of science, taken as a comprehensive theory of the world?

We should note that the sciences do link up more systematically than people are apt to realize who forget about logic and mathematics; for logic is shared by all branches of science, and much of mathematics is shared by many. People tend unduly to see the logical and mathematical components of science as different in kind from the rest, and hence fail to see these components as something common to all the branches. Ironically, this very neutrality, this fact of being shared by all branches of science, has encouraged people to think of the logical and mathematical components as different in kind from the rest, and hence to fail to recognize the unity that they confer. Thus I see science as a considerably integrated system of the world even now, though the explicit reduction of major branches to theoretical physics is incomplete.

But we can appreciate this degree of integration and still appreciate how unrealistic it would be to extend a Duhemian holism to the whole of sci-

ence, taking all science as the unit that is responsible to observation. Science is neither discontinuous nor monolithic. It is variously jointed, and loose in the joints in varying degrees. In the face of a recalcitrant observation we are free to choose what statements to revise and what ones to hold fast, and these alternatives will disrupt various stretches of scientific theory in various ways, varying in severity. Little is gained by saying that the unit is in principle the whole of science, however defensible this claim may be in a legalistic way.

This is the end of my digression on the thesis of holism. I shall be concerned from here on with the empirical under-determination of natural science. It is a doctrine which, as I said, is plausible insofar as it is intelligible. My purpose will be to examine the meaning of this thesis more closely, and to consider its limits and its consequences.

A notion that is evidently central to the thesis is that of observation. This notion is subject to a curious internal tension. Observation affords the sensory evidence for scientific theory, and sensation is private. Yet observation must be shared if it is to provide the common ground where scientists can resolve their disagreements. The observation must be the distillate, somehow, of what is publicly relevant in the private sensations of present witnesses. This delicate process of distillation is already accomplished, happily, in our most rudimentary learning of language. One learns the word 'blue' from another speaker, in the presence of something blue. The other speaker has learned to associate the word with whatever inscrutable sensation it may be that such an object induces in him, and one now learns to associate the word with the sensation, same or different, that the object induces in oneself. All agree in calling the object blue, and even in calling their sensations blue.

We do well to recognize this crucial role of language, for in view of it we can spare ourselves the finicky task of defining the elusive notion of observation. We can speak rather of observation terms and observation sentences, thus cleaving to linguistic forms: 'blue', 'This is blue'. Observational expressions can be roughly distinguished from others by a behavioral criterion, involving no probing of sensations. For this is characteristic of them: witnesses will agree on the spot in applying an observation term, or in assenting to an observation sentence, if they are conversant with the language. Their verdicts do not vary with variations in their past experience.

It is sometimes objected that a specialist may recognize at a glance what the untrained observer cannot. The above behavioral criterion settles this

discrepancy in favor of the untrained observer. Specialists rest content with the level of evidence that commands their expert agreement but in principle they usually could reduce this recondite evidence to observation terms at the layman's level.

Not always. There is expertise in tea tasting, in wine tasting, and in the recognition of tones, chords, and timbre, that resists conversion to the common coin.¹ We should like to be able to reckon the esoteric terms in these domains as observational for the experts, contrary to the proposed behavioral criterion. And indeed even a common observation term such as 'blue' has its penumbra of vagueness, where witnesses may disagree in their verdicts. The really distinctive trait of observation terms and sentences is to be sought not in concurrence of witnesses but in ways of learning. Observational expressions are expressions that can be learned ostensively. They are actually learned ostensively in some cases and discursively in others, but each of them could be learned by sufficiently persistent ostension. The behavioral manifestation of observationality, then, namely, the ready concurrence of witnesses, serves merely as a rough practical criterion.

Observation sentences are not incorrigible. A witness who has assented to an observation sentence on the spot is permitted to reconsider his verdict later in the face of conflicting theory. And of course the typical observation terms are not subjective in reference, but objective. 'Blue' was one; others are 'water', 'rabbit', 'ball', 'hard'. They recur in theoretical sentences.

Now the doctrine that is up for clarification is that scientific theory is under-determined by observable events. So we shall want to get clear on the relation of theories to observations. Or, now that we have taken to talking of observation sentences and terms, rather than of observations, let us look to the relation of theories to observation sentences.

An observation sentence is an occasion sentence; it commands assent on some occasions and not others, depending on what is happening where and when the sentence is queried. On the other hand the sentences of scientific theory are standing sentences. They are meant to be true or false independently of the occasion of utterance. The observation sentences cannot, as occasion sentences, be implied by theory; we must first change them into standing sentences, by incorporating specifications of place-times. Let us adopt, then, an arbitrary numerical system of spatio-

1. I am indebted to Joseph Cowan here. For valuable criticism of the paper in general, I am indebted to Burton Dreben.

temporal coordinates, and let us contemplate the infinite totality of what I shall call *pegged observation sentences*. Each observation sentence expressible in our language gets joined to each combination of spatio-temporal coordinates. The resulting sentences are standing sentences, some true and some false. The true ones do not depend for their truth on anyone's having made the observations; it matters only that the observable state or event in question occurs, in fact, at the specified place-time. The time and place may be beyond the reach of all sentient life.²

Our move from occasional observation sentences to these pegged observation sentences is already an abrupt ascent from observation into theory. We need to know not only a bit of mathematics but also quite a lot about the physical world in order to establish a system of coordinates. I shall suppose this much achieved, in order to get on better with further questions; but let us not lose sight of the magnitude of our assumptions. The doctrine of under-determination says there is a certain slack between observation and theory; and we have already lost some of that slack by granting the system of coordinates. Just in order to define the slack, we are having to take some of it up.

We have moved from the occasional observation sentences to the pegged observation sentences because we want sentences that theories can imply. However, the goal is not yet reached. Typically a theory does not imply even a pegged observation sentence outright. Typically a theory traffics rather in generalities. Typically a theory will descend to particulars only conditionally upon other particulars, assumed as boundary conditions. I say 'typically' because I do not want to exclude particulars from theories altogether. I do not want to exclude such particular and unconditional conclusions as might concern the moon or Ur of the Chaldees or the cave painters of Altamira. But even such unconditional conclusions, though particular, will seldom be observational; they will seldom be pegged observation sentences. For a just view of the relation of scientific theory to the observations that support or refute it, we must look to the use of boundary conditions. The theory, plus some set of pegged observation sentences that have already been verified, implies some further pegged observation sentence that can now be checked; such is the test of a theory.

2. To dispel a misunderstanding in Harold Morick, "Observation and Subjectivity in Quine," *Canadian Journal of Philosophy* 1 (1974): 109–127, I must stress that a pegged observation sentence is not an observation sentence. It is a non-observational sentence obtained by pegging an observation sentence.

Instead of saying that the theory and the boundary conditions together imply the further pegged observation sentence, we could as well say that the theory implies, outright, a conditional sentence whose antecedent comprises the boundary conditions and whose consequent is the further pegged observation sentence. Such a conditional sentence I shall call an *observation conditional*. Its antecedent is a conjunction of pegged observation sentences and its consequent is a pegged observation sentence.

At last we can state, tentatively, the relation of theory to observation: the theory implies observation conditionals. However, we have still to tidy up the notion of a theory. I have spoken of a theory as implying sentences, as if the theory were itself a sentence or a set of sentences. It will be better to speak of a *theory formulation* as doing the implying. The theory formulation is simply a sentence—typically a conjunctive sentence comprising the so-called axioms of the theory. Currently the theory itself, then, is often identified with an infinite set of sentences, namely, the logical consequences of the theory formulation. Such has usually been my own usage. A single theory, in this sense, admits of many formulations; all that is required is that they be logically equivalent. But we shall find that even this requirement is too stringent to fit the more common and traditional usage of the term ‘theory’; and we shall find also that some such broader notion of theory is wanted for making proper sense of the thesis of underdetermination.

Meanwhile a word about implication, or logical consequence, and logical equivalence. The notions are clear as long as the theory formulations and their consequences are couched in our regimented scientific language, with its explicit logical notation. If we try to accommodate formulations of theory in other languages, however, it is less clear when to say that one theory formulation implies another, or implies a given observation conditional. Problems of translation intrude. I prefer in the present study to set aside the whole question of formulating theories in other languages, and to think of theories only as formulated in our own language, with help of our own regimented logical notation. These parochial limitations will help us move forward to central issues.

Incidentally a yet more drastic limitation has been with us for some time, in the pegged observation sentences that go into our observation conditionals. For these are pegged to a coordinate system, and even to a single arbitrarily chosen coordinate system adhered to throughout the present theory of theories. When problems outside the scope of this paper

call for lifting these arbitrary limitations, that will be soon enough to consider how best to do so.

For now, accordingly, we can say that theories are always formulated within our own language, of standard logical form. But even within these cozy limits I shall not want simply to identify a theory with the logical consequences of a theory formulation. The next consideration will show why.

Take some theory formulation and select two of its terms, say ‘electron’ and ‘molecule’. I am supposing that these do not figure essentially in any observation sentences; they are purely theoretical. Now let us transform our theory formulation merely by switching these two terms throughout.³ The new theory formulation will be logically incompatible with the old: it will affirm things about so-called electrons that the other denies. Yet their only difference, the man in the street would say, is terminological; the one theory formulation uses the technical terms ‘molecule’ and ‘electron’ to name what the other formulation calls ‘electron’ and ‘molecule’. The two formulations express, he would say, the same theory. Someone else might urge, however perversely, that they express very different theories: both of them treat of molecules in the same sense but disagree sharply regarding the behavior of molecules, and correspondingly for electrons. Clearly, in any event, the two theory formulations are *empirically equivalent*—that is, they imply the same observation conditionals. I think, moreover, that we should individuate theories in such a way as to agree with the man in the street: the two formulations formulate the same theory, despite their overt logical incompatibility. This is why I do not want to identify a theory with the logical consequences of a formulation. I do not want to require that two formulations of a theory be logically equivalent, nor even logically compatible.

Certainly two formulations of a theory should be empirically equivalent, in the sense just defined, even if not logically equivalent. Still empirical equivalence must not be the only requirement, unless we are to repudiate the doctrine of underdetermination out of hand; for that doctrine says that empirically equivalent theories can conflict. What is required of two formulations of a theory must be, in short, some relation stronger than empirical equivalence and weaker than logical equivalence.

The unsatisfying example of ‘molecule’ and ‘electron’ consisted, intuitively speaking, in switching the meanings of the two words. When the

3. There is substantially this idea in B. M. Humphries, “Indeterminacy of Translation and Theory,” *Journal of Philosophy* 67 (1970): 167–178, particularly 169f.

man in the street protested that the conflict between the two theory formulations was merely terminological, his point was that it could be resolved by treating the one formulation as not quite English, and translating its respective words 'molecule' and 'electron' into the English words 'electron' and 'molecule'. I want to preserve this insight while avoiding problems raised by the terms 'meaning' and 'translation'. I can do so by appealing to little more than a permutation of vocabulary; in the present case, of course, a mere switching of the predicates 'molecule' and 'electron' throughout one of the theory formulations. I propose that we count two formulations as formulations of the same theory if, besides being empirically equivalent, the two formulations can be rendered identical by switching predicates in one of them.

This criterion needs a little broadening, in obvious respects. Since logically equivalent formulations were in any event to count as formulations of the same theory, we should not require that a switching of terms render formulations identical; we should only require it to render them logically equivalent. Further, we should not limit the permutation to a switching of two predicates; we should allow permutations of many. Finally, it would be arbitrary to require this transformation to carry predicates always into simple one-word predicates. The intuitive notion, after all, was a reconstruing of predicates; and the general way of reconstruing an n -place predicate is by supplying an open sentence in n variables, not caring whether there happens to be a word in our language with the same extension as that open sentence.

By a *reconstrual* of the predicates of our language, accordingly, let me mean any mapping of our lexicon of predicates into our open sentences (n -place predicates to n -variable sentences). Thus the predicate 'heavier than' might be mapped to the open sentence ' x is heavier than y ' an identity mapping changing nothing, while the predicates 'molecule' and 'electron' might be mapped to the respective open sentences ' x is an electron' and ' x is a molecule', producing our example.

So I propose to individuate theories thus: two formulations express the same theory if they are empirically equivalent and there is a reconstrual of predicates that transforms the one theory into a logical equivalent of the other.⁴

4. Avishai Margalit has suggested to me that this amounts to equating theories that can be formulated by the same Ramsey sentence. (F. P. Ramsey, *The Foundations of Mathematics*, ed. R. B. Braithwaite, Routledge, and Kegan Paul (London, 1931): chapter ix(A), "Theories".

I was able to define reconstrual more simply than otherwise only by assuming something about the form of our language that I now ought to make explicit. I am assuming the standard logical form of language at its most economical; there are just truth functions, quantification, and a finite lexicon of predicates. I made no provision for names or for functors, for there are well known ways of serving the purposes of these devices on the more austere basis.

We have now settled the individuation of theories, within our parochial confines. We have said when to count two formulations as expressing the same theory. Given this equivalence relation, it is a routine matter to say what a theory is. The method is artificial but familiar: theories are the equivalence classes of that equivalence relation. The theory expressed by a given formulation is the class of all the formulations that are empirically equivalent to that formulation and can be transformed into logical equivalents of it or vice versa by reconstrual of predicates.

It is usual in the literature to require of a theory that it be deductively closed. In our present terms, what this means is that if you change a formulation of a theory merely by annexing some logical consequences of that formulation, the result will still be a formulation of the same theory. We have insured this by requiring only that the reconstrual of predicates render the formulations logically equivalent, not identical. Thanks to the circumstance that any formulation is equivalent to itself plus any of its consequences, it is easily shown that theories as I have defined them are deductively closed.

So defined, theories are classes of theory formulations. But at this point we must liberalize the notion of a theory formulation, so as not to be limited to the few formulations that are physically available on paper. What are wanted rather are linguistic sequences in the abstract sense, in their infinite variety. Each single word or letter can still be explained as the class of all its tokens, a class of actual inscriptions, since we are assured that these classes all have members variously situated in space-time. Sentences, however, and longer expressions, are to be taken rather as mathematical sequences of their component words or letters. An expression in this sense is a function, or class of ordered pairs; the first word or letter of the expression is paired with the number 1, the second with 2, and so on. In this way we can assure the existence of all expressions however long, all theory formulations as yet unconceived, all texts as yet unwritten; all 'possible' expressions, as one might say. Theories, finally, are classes of formulations, hence classes of expressions in this abstract sense; classes of functions.

Because the question how to define a theory is interesting in itself, I have pursued it farther than required for what I want to say about the thesis of under-determination. For this purpose it is clearer to treat directly of the empirical equivalence of theory formulations, and of the reconstrual of predicates therein, and not to pause over the individuation of theories as such. In these terms, under-determination says that for any one theory formulation there is another that is empirically equivalent to it but logically incompatible with it, and cannot be rendered logically equivalent to it by any reconstrual of predicates.

We had one attempt at an example, and it failed: the trivial example of two theories that differed only in the switch of 'electron' and 'molecule'. A less trivial case, to which I now turn, is due to Poincaré. Here we have one formulation of cosmology that represents space as infinite, and another formulation that represents space as finite but depicts all objects as shrinking in proportion as they move away from center. The two formulations, again, are empirically equivalent. But again the example is disappointing as an example of under-determination, because again we can bring the two formulations into coincidence by reconstruing the predicates. The reconstrual called for here is less simple than the mere switch of 'electron' with 'molecule', but it presents no serious challenge. The two formulations are formulations, again, of a single theory.

Having disqualified these permutations as cases of under-determination, we might in passing consider how they stand as cases of indeterminacy of translation. If in the light of verbal behavior we translate two foreign words as 'molecule' and 'electron', what behavioral evidence could have obstructed the opposite choices? None, surely, except as we invoke what Neil Wilson called the principle of charity: maximize the agreement between the native and ourselves on questions of truth and falsity, other things being equal. Translation is not the recapturing of some determinate entity, a meaning, but only a balancing of various values. An observation sentence and its translation should command assent under similar stimulations; here is one value. Wide concomitance of assent to standing sentences is also a value. Good translation strikes some optimum combination of values, insofar as they can be compared.

Let us return now to the thesis of under-determination of natural science. We saw that what we need for illustration of this thesis are theory formulations that are empirically equivalent, logically incompatible, and irreconcilable by reconstrual of predicates. For substantiation of the thesis of under-determination we need more: we need to show not only that

such branching alternatives exist, but that they are inevitable. Thus suppose we had an adequate theory of nature, and then we were to add to it some gratuitous further sentences that had no effect on its empirical content. By ringing changes on these excrescences we might get alternative theories, logically incompatible, yet always empirically equivalent. This gratuitous branching of theories would be of no interest to the thesis of under-determination, since the adequate original theory was itself logically compatible with each one of these gratuitous extensions; they were incompatible only with one another. What the thesis of under-determination calls for is unavoidable branching.⁵ The adequate original theory, in our imaginary example, would itself have had to be one of several equivalent and logically incompatible theories if it was to illustrate the thesis of under-determination.

In its full generality, the thesis of under-determination thus interpreted is surely untenable. It must fail for weak theories, theories that imply no rich store of observation conditionals. If the implied observation conditionals (redundancies aside) are finite in number, we can simply take the conjunction of them, a single sentence, as our theory formulation. It contains its observation conditionals without remainder; they are all it is. It is implied by every empirically equivalent theory, and can conflict with none of them. Any that it conflicted with would have to be internally inconsistent, and so not empirically equivalent.

So we see that the thesis of under-determination must fail where only finitely many observation conditionals are implied. They, in conjunction, are their own theory formulation. But much the same thing can happen even where a theory irreducibly implies infinitely many observation conditionals; for it may happen that these can all be encompassed by a single universally quantified conditional, or by finitely many. Such a theory formulation, again, affords a tight fit. No theory formulation that implies just those same observation conditionals can conflict with it, unless indeed it is ω -inconsistent. And I think we may reject ω -inconsistent theories.

5. This requirement evidently disqualifies, for our purposes, an example of empirically equivalent and logically irreconcilable theories that is offered by Clark Glymour in his important paper "Theoretical Realism and Theoretical Equivalence," *Boston Studies in the Philosophy of Science* VIII (1971): 275–288. In his example, the empirical evidence is covered by the statement that there are infinitely many objects. This content can be organized indifferently in a theory of dense order and a theory of discrete order, and these two theories are irreconcilable, sharing, as he says, no common model. But this is a case of avoidable branching.

The empirical content of a theory formulation is summed up in the observation conditionals that the formulation implies. These are material conditionals; each is a truth function of pegged observation sentences. If we could check the truth values of all the pegged observation sentences, we could evaluate any observation conditional without consulting the theory formulation; theory could be dispensed with. But we cannot. Most of the pegged observation sentences are pegged to inaccessible place-times, and they are infinitely many. The theory formulation is thus a device for remote control and for mass coverage. The theory formulation serves to specify en masse the observation conditionals that we are rightly or wrongly taking to be true. It specifies them by implying them. Under-determination lurks where there are two irreconcilable formulations each of which implies exactly the desired set of observation conditionals plus extraneous theoretical matter, and where no formulation affords a tighter fit.

The only hope for a thesis of under-determination, evidently, is in application to theories that imply observation conditionals infinite in number and too ill-assorted to be exactly encompassed by any finite formulation; tightly encompassed, that is, without theoretical foreign matter. The thesis needs to be read as a thesis about the world. It needs to be read as saying, for one thing, that the observation conditionals that are in fact true in the world are thus ill-assorted. And it needs to be read as saying, further, that we can encompass more of these true observation conditionals in a loose formulation than in any tight one. And it needs to be read as saying, finally, that for any such loose formulation there will be others, empirically equivalent but logically incompatible with it and incapable of being rendered logically equivalent to it by any reconstrual of predicates.

Here, evidently, is the nature of under-determination. There is some infinite lot of observation conditionals that we want to capture in a finite formulation. Because of the complexity of the assortment, we cannot produce a finite formulation that would be equivalent merely to their infinite conjunction. Any finite formulation that will imply them is going to have to imply also some trumped-up matter, or stuffing, whose only service is to round out the formulation. There is some freedom of choice of stuffing, and such is the under-determination.

One may ask: why insist on a finite formulation? Why not just settle for the desired observation conditionals as they stand, in all their infinite variety? The answer is immediate, in the form of another question: how else are you going to specify this desired class of observation conditionals? But

this response does not conclude the matter; for we must now notice a startling result due to William Craig.⁶ Consider any formulation, and any desired class of consequences of it. For our purposes these consequences would be observation conditionals, but for Craig they can be any sentences. Then Craig shows how to specify a second or Craig class of sentences which are visibly equivalent, one by one, to the sentences of the desired first class; and the remarkable thing about this second class is that membership in it admits of a mechanical decision procedure.

In the cases that matter, these classes are infinite. Even so, the second or Craig class evidently makes the original finite formulation dispensable, by affording a different way of recognizing membership in the desired first class. Instead of showing that a sentence belongs to it by deducing it from the finite formulation, we show it by citing a visibly equivalent sentence that belongs, testably, to the Craig class.

This result does not belie under-determination, since the Craig class is not a finite formulation, but an infinite class of sentences. But it does challenge the interest of under-determination, by suggesting that the finite formulation is dispensable; and indeed the Craig class, for all its infinitude, is an exact fit, being a class of visible equivalents of the desired class. In fact I might say just how excessively visible these equivalences are. Each sentence in the Craig class is simply a repetitive self-conjunction, ' $ppp \dots p$ ', of a sentence of the desired class.

Having said this much, I would do well to finish the Craig story. Why, when the desired class itself is undecidable, should this Craig class of its repetitive self-conjunctions be decidable? The trick is as follows. Each of the desired sentences (each of the desired observation conditionals, in our case) is deducible from the original finite formulation. Its proof can be coded numerically, Gödel fashion. Let the number be n . Then the corresponding sentence in the Craig class is the desired sentence repeated in self-conjunction n times. The resulting Craig class is decidable. To decide whether a given sentence belongs to it, count its internal repetitions; decode the proof, if any, that this number encodes; and see whether it is a proof of the repeated part of the given sentence.

Obviously there is no place for any of this in practice, as Craig was the first to emphasize. Just counting the repetitions and then decoding the

6. William Craig, "Replacement of Auxiliary Expressions," *Philosophical Review* 65 (1956): 38–55.

proof from the Gödel number would require astronomical time, and each sentence in the Craig class would require astronomical space if it were to be written. And, after all that, the old original finite formulation still has to be consulted in checking the proof. Craig's point is of course strictly theoretical, and as such it is important.

As I said before, Craig's result does not refute the thesis of under-determination, since the Craig class, for all its tightness of fit, is not a finite formulation. However, this technicality is rather a frail reed at which to grasp. After all, one could reasonably extend the notion of theory formulations to apply not just to an expression but to a recursive set of expressions. So the thesis of under-determination would seem to be demoted to the status, at best, of a thesis affirming a certain contrast between expressions and recursive sets of expressions.

However, I see the importance of the thesis of under-determination as lying elsewhere. The more closely we examine the thesis, the less we seem to be able to claim for it as a theoretical thesis; but it retains significance in terms of what is practically feasible. A tempered version, the most favorable available, might run as follows. We, humanly, are capable of encompassing more true observation conditionals in a loose theory formulation than in any tight system that we might discover and formulate independently of any such loose formulation. And then the thesis would go on to say, as before, that for each such formulation there will be others, empirically equivalent but logically incompatible with it and incapable of being rendered logically equivalent to it by any reconstrual of predicates.

Even in this form the thesis is moot. It no longer stands to reason, as it seemed at first to do. The question now is whether we are underestimating the power of reconstrual of predicates. It does still stand to reason, overwhelmingly, that any theory formulation we may hope to devise as an adequate system of the world will be a loose one; that there will be others empirically equivalent to it and logically incompatible with it. This much is illustrated by the very trivial example where the words 'electron' and 'molecule' were switched, and by the half-trivial example from Poincaré; but these incompatibilities were reconciled by reconstrual of predicates that preserved empirical equivalence. What is moot is whether there are also bound to be cases not thus reconcilable.

The easy way to recognize empirical equivalence of two theory formulations is by seeing a reconstrual of predicates that will carry the one into the other. So it was with the examples just mentioned. But surely this is not the

only way. We might study two incompatible theory formulations, trying in vain to imagine an observation that could decide between them, and we might conclude that they are empirically equivalent; we might conclude this without seeing a reconciling reconstrual of predicates. This we might; but there still could be a reconciling reconstrual of predicates, subtle and complex and forever undiscovered. The thesis of under-determination, even in my latest tempered version, asserts that our system of the world is bound to have empirically equivalent alternatives that are not reconcilable by reconstrual of predicates however devious. This, for me, is an open question.

Failing that, a last-ditch version of the thesis of under-determination would assert merely that our system of the world is bound to have empirically equivalent alternatives which, if we were to discover them, we would see no way of reconciling by reconstrual of predicates. This vague and modest thesis I do believe. For all its modesty and vagueness, moreover, I think it vitally important to one's attitude toward science. What it says in effect is just that there are undiscovered systematic alternatives much deeper and less transparent than, for instance, the Poincaré example.

It sets one to wondering about truth. Perhaps there are two best theories that imply all the true observation conditionals and no false ones. The two are equally simple, let us suppose, and logically incompatible. Suppose further, contrary to our last conjecture, that they are not reconcilable by reconstrual of predicates, however devious. Can we say that one, perhaps, is true, and the other therefore false, but that it is impossible in principle to know which? Or, taking a more positivistic line, should we say that truth reaches only to the observation conditionals at most, and, in Kronecker's words, that *alles übrige ist Menschenwerk*?

I incline to neither line. Whatever we affirm, after all, we affirm as a statement within our aggregate theory of nature as we now see it; and to call a statement true is just to reaffirm it. Perhaps it is not true, and perhaps we shall find that out; but in any event there is no extra-theoretic truth, no higher truth than the truth we are claiming or aspiring to as we continue to tinker with our system of the world from within. If ours were one of those two rival best theories that we imagined a moment ago, it would be our place to insist on the truth of our laws and the falsity of the other theory where it conflicts.

This has the ring of cultural relativism. That way, however, lies paradox. Truth, says the cultural relativist, is culture-bound. But if it were, then he,

within his own culture, ought to see his own culture-bound truth as absolute. He cannot proclaim cultural relativism without rising above it, and he cannot rise above it without giving it up.

There is a final fantasy to contemplate. Suppose again two rival systems of the world, equally sustained by all experience, equally simple, and irreconcilable by reconstrual of predicates. Suppose further that we can appreciate their empirical equivalence. Must we still embrace one theory and oppose the other, in an irreducible existentialist act of irrational commitment? It seems an odd place for irrational commitment, and I think we can do better. It is the extreme situation where we would do well to settle for a frank dualism. Oscillation between rival theories is standard scientific procedure anyway, for it is thus that one explores and assesses alternative hypotheses. Where there is forever no basis for choosing, then, we may simply rest with both systems and discourse freely in both, using distinctive signs to indicate which game we are playing. This use of distinctive signs leaves us with two irreducible and unconflicting theories.

Mind and Verbal Dispositions

1975

Descartes supposed that man is the only animal endowed with mind; the others are automata. It is held further, and more widely and on better evidence, that man is the only animal endowed with language. Now if man is unique in enjoying these two gifts, it is no coincidence. One may argue that no mindless creature could cope with so intricate a device as language. Or one may argue conversely that no appreciable mental activity is conceivable without linguistic aids.

Most thought simply *is* speech, according to the pioneer behaviourist John B. Watson: silent, repressed, incipient speech. Not all thought is that. A geometer or an engineer may think by means also of little incipient tugs of the muscles that are used in drawing curves or twirling cogwheels. Still, the muscles that play by far the major role, according to Watson's muscular theory of meditation, are the muscles used in making speeches.

Conversely, there is an age-old and persistent tendency to try to explain and analyse the physical phenomenon of speech by appealing to mind, mental activity, and mental entities: by appealing to thoughts, ideas, meanings. Language, we are told, serves to convey ideas. We learn language from our elders by learning to associate the words with the same ideas with which our elders have learned to associate them. Thus it is, we may be told, that approximate uniformity of association of words with ideas is achieved and maintained throughout the community.

Such an account would of course be extravagantly perverse. Thus consider the case where we teach the infant a word by reinforcing his ran-

dom babbling on some appropriate occasion. His chance utterance bears a chance resemblance to a word appropriate to the occasion, and we reward him. The occasion must be some object or some stimulus source that we as well as the child are in a position to notice. Furthermore, we must be in a position to observe that the child is in a position to notice it. Only thus would there be any purpose in our rewarding his chance utterance. In so doing we encourage the child to repeat the word on future similar occasions. But are we causing him to associate the word with the same *idea* that we adults associate it with? Do we adults all associate it with the same idea ourselves, for that matter? And what would that mean?

The moral of this is that the fixed points are just the shared stimulus and the word; the ideas in between are as may be and may vary as they please, so long as the external stimulus in question stays paired up with the word in question for all concerned. The point is well dramatized by the familiar fantasy of complementary colour perception. Who knows but that I see things in colours opposite to those in which you see the things? For communication it is a matter of indifference.

I believe in the affinity of mind and language, but I want to keep the relation right side up. Watson's theory of thought, however inadequate, has matters right side up. A theory of mind can gain clarity and substance, I think, from a better understanding of the workings of language, whereas little understanding of the workings of language is to be hoped for in mentalistic terms.

I shall say a little about how it is that people feel drawn to a mentalistic account of language, despite the conspicuous fact that language is a social enterprise which is keyed to intersubjectively observable objects in the external world. Also I shall speculate on how we might hope to get on with a properly physicalistic account of language. First I must talk a little more about learning.

I mentioned one primitive way of learning a word: through reinforcement of random babbling. Another way, somewhat the reverse, is imitation. In the case of babbling it was the adult that witnessed what was confronting the child when the child chanced to babble the appropriate word. In the case of imitation it is the child, conversely, that witnesses what is confronting the adult when the adult volunteers the word. The child then volunteers the word when similarly confronted, and thereupon the adult proceeds to reinforce the child's behaviour just as in the case of babbling.

The imitation method is more sophisticated than the babbling method. It can still be explained, indirectly, in terms of stimulus and reinforced response; but I won't pause over it.

What we do need to notice is that all language learning at this primitive level is directed only to the learning of what may be called observation terms, or, more properly, observation sentences. The child learns to assent to the query 'red?' in the conspicuous presence of red things. Also he masters the trick of getting the object by uttering the word; 'red' is a poor example here, but 'ball' and 'milk' and 'Mama' are clear cases. Also he masters the word in a passive way, responding in some distinctive fashion on hearing it. He may respond by turning to face the object, or by fetching it.

Now the observation term, or observation sentence, is a ground on which John the rational animal and Fido the automaton can meet and to some degree communicate. The dog learns observation sentences in his passive way. He learns to respond to them by salivating, by running to the kitchen, by turning to face the object, or by fetching it.

Already at this lowly level of observation sentences the small child differs from the dog, it may seem, in that he learns the sentences also actively: he utters them. This still is not a clear contrast. Dogs learn to ask for things, in their inarticulate way. Let us not arrogate to rationality what may be merely superior agility of lips and tongue and larynx. Premack and his chimpanzee have circumvented these muscular obstacles by resorting to plastic symbols, which they push around on a board. Premack succeeds in teaching his chimpanzee to volunteer observation sentences appropriately and to play a passable game of query and assent.

A contrast that has long been remarked, between human language and animal signals, is the combinatorial productivity of language: man's ability to compose new and unprecedented sentences from old materials, and to respond appropriately to such new creations. But Premack reports that his chimpanzee even passes this test, within modest limits. It would thus appear that combinatorial productivity in language affords no sharp line between man and beast. Man may plume himself on having been the first to develop a combinatorially productive language, but the ability to learn it may be more widespread.

Combinatorial productivity, however, is not the only trait that has seemed to distinguish mind-governed discourse from the performance of trained animals. A major factor is the unpredictable spontaneity of speech. Animal drives are still at work behind the torrent of human speech, but

they are seldom clearly to be traced. Even if in our verbal output we differ from Premack's chimpanzee only in degree and not in kind, still it is this overwhelming difference of degree that invites the mentalistic accounts of verbal behaviour. The torrent of words is seen as a manifestation of the speaker's inner life beyond animal drives. Nowadays one is apt to resort thus to a mentalistic semantics not so much because one sees an ontological gulf between man and the apes, as because one despairs of adhering to the standards of natural science in coping with the complexities of intelligent discourse.

The central notion in mentalistic semantics is an unanalysed notion of meaning. It figures mainly in two contexts: where we speak of knowing the meaning of an expression, and where we speak of sameness of meaning. We say we know the meaning of an expression when we are able to produce a clearer or more familiar expression having the same meaning. We ask the meaning of an expression when what we want is a clearer or more familiar expression having the same meaning.

I said to my small son, 'Eighty-two. You know what I mean?' He said, 'No.' Then I said to my small daughter, 'Ottantadue. You know what I mean?' She said, 'Yes. Eighty-two.' I said, 'See, Margaret understands Italian better than Douglas understands English.'

Our ways of talking of meaning are thus misleading. To understand an expression is, one would say, to know the meaning; and to know the meaning is, one would say, to be able to give the meaning. Yet Douglas could rightly claim to *understand* the expression 'eighty-two', despite answering 'No' to 'You know what I mean?' He answered 'No' because he was unable to *give* the meaning; and he was unable to give the meaning because what we call giving the meaning consists really in the asymmetrical operation of producing an equivalent expression that is clearer. Margaret was ready with a clearer equivalent of 'ottantadue', but Douglas was at a loss for a *still* clearer equivalent of 'eighty-two'. In another context he might have ventured, 'Yes, you mean the temperature is eighty-two.'

People persist in talking thus of knowing the meaning, and of giving the meaning, and of sameness of meaning, where they could omit mention of meaning and merely talk of understanding an expression, or talk of the equivalence of expressions and the paraphrasing of expressions. They do so because the notion of meaning is felt somehow to *explain* the understanding and equivalence of expressions. We understand expressions by knowing or grasping their meanings; and one expression serves as a trans-

lation or paraphrase of another because they mean the same. It is of course spurious explanation, mentalistic explanation at its worst. The paradoxical little confusion between understanding 'eighty-two' and knowing or giving its meaning is already¹ symptomatic of awkward concept-building; but where the real threat lies, in talking of meaning, is in the illusion of explanation.

In all we may distinguish three levels of purported explanation, three degrees of depth: the mental, the behavioural, and the physiological. The mental is the most superficial of these, scarcely deserving the name of explanation. The physiological is the deepest and most ambitious, and it is the place for causal explanations. The behavioural level, in between, is what we must settle for in our descriptions of language, in our formulations of language rules, and in our explications of semantical terms. It is here, if anywhere, that we must give our account of the understanding of an expression, and our account of the equivalence that holds between an expression and its translation or paraphrase. These things need to be explained, if at all, in behavioural terms; in terms of dispositions to overt gross behaviour.

Take understanding. Part of the understanding of a word consists in the ability to use it properly in all manner of admissible contexts. Part consists in reacting properly to all such uses. So there is a good deal here to sort out and organize. We must divide and define. To begin with we can set aside the complication of the myriad sentential contexts of a word, by beginning rather with sentences as wholes: with complete little isolated speeches, consisting perhaps of a single word and perhaps of more.

Bewildering variety confronts us even so. One and the same little sentence may be uttered for various purposes: to warn, to remind, to obtain possession, to gain confirmation, to gain admiration, or to give pleasure by pointing something out. The occasions for uttering one and the same sentence are so various that we can seldom predict when a sentence will be uttered or which one it will be. An unpromising setting, this, in which to explore and exploit verbal dispositions. Somehow we must further divide; we must find some significant central strand to extract from the tangle.

Truth will do nicely. Some sentences, of course, do not have truth values: thus questions and imperatives. Those that do may still be uttered for a variety of reasons unconnected with instruction; I just now enumerated

1. Published text read "always"; Quine corrected text on his reprint.

some. But, among these sentences, truth is a great leveller, enabling us to postpone consideration of all those troublesome excrescences. Here, then, is an adjusted standard of understanding: a man understands a sentence in so far as he knows its truth conditions. This kind of understanding stops short of humour, irony, innuendo, and other literary values, but it goes a long way. In particular it is all we can ask of an understanding of the language of science.

We are interested not only in explaining what it is for someone else to understand a sentence, but also in setting a standard for ourselves, as when we try to penetrate a new language and to understand its sentences, or try to teach the language. Our standard, still, is this: give the truth conditions. Hence Davidson's plan for a semantics in the style of Tarski's truth definition.

But when I define the understanding of a sentence as knowledge of its truth conditions I am certainly not offering a definition to rest with; my term 'knowledge' is as poor a resting-point as the term 'understanding' itself.

We were supposed to be getting things down to terms of dispositions to behaviour. In what behavioural disposition then does a man's knowledge of the truth conditions of the sentence 'This is red' consist? Not, certainly, in a disposition to affirm the sentence on every occasion of observing a red object, and to deny it on all other occasions; it is the disposition to assent or dissent when asked in the presence or absence of red. Query and assent, query and dissent—here is the solvent that reduces understanding to verbal disposition. Without this device there would be no hope of handing language down the generations, nor any hope of breaking into newly discovered languages. It is primarily by querying sentences for assent and dissent that we tap the reservoirs of verbal disposition.

This approach applies primarily to terms, or occasion sentences, rather than to standing sentences. For the disposition to assent to or dissent from the sentence 'This is red' is marked by a correlation between assent and the presence of red, and between dissent and the absence of red, on occasions where the sentence is queried. A standing sentence, whose truth value remains fixed over long periods, offers no significant correlation of the kind. Where the method of queried assent and dissent is at its best, indeed, is in application to occasion sentences of the special sort that I have called observation sentences; for the occasions that make the sentence true are going to have to be intersubjectively recognizable if we are to be able to *tell*

whether the speaker has the disposition in question. Even in these cases, of course, we remain at the mercy of the speaker's veracity: we assume when querying him that his assents and dissents are sincere. Happily we live in a moral climate where this assumption generally holds up; language could not flourish otherwise.

Standing sentences can be queried too, but the stimulatory situation at the time of querying them will usually have no bearing on the verdict; and for this reason we cannot identify the understanding of a standing sentence, even approximately, with a disposition to assent or dissent when queried on particular occasions. I do not know how, in general, in terms of behavioural dispositions, to approximate to the notion of understanding at all, when the sentences understood are standing sentences. Perhaps it cannot be done, taking standing sentences one by one.

Once in a while we get a hint of a specifically relevant disposition, still, when we find someone reversing his verdict on a standing sentence in the face of some observation. But with all conceivable luck we cannot hope to correlate standing sentences generally with observations, because the sentences one by one simply do not have their own separable empirical implications. A multiplicity of standing sentences will interlock, rather, as a theory; and an observation in conflict with that theory may be accommodated by revoking one or other of the sentences—no one sentence in particular.

One sees how a semanticist might despair and seek shelter in the jungle of mentalistic semantics. But there are other courses. Perhaps the very notion of understanding, as applied to single standing sentences, simply cannot be explicated in terms of behavioural dispositions. Perhaps therefore it is simply an untenable notion, notwithstanding our intuitive predilections. It stands to reason that a proper semantical analysis of standing sentences, in terms of behavioural dispositions, will be primarily occupied with the interrelations of sentences rather than with standing sentences one by one.

I mentioned two central semantical notions which, in mentalistic semantics, are obscured by talk of meaning. One was the notion of understanding an expression, and the other was the relation of equivalence between an expression and its paraphrase. Afterwards I considered what might be done about understanding. Now what about the other notion, the equivalence relation? Much of what I have said about understanding applies in parallel fashion to equivalence. Here, as there, we can conveniently organize our work by looking first to sentences as wholes, seeking

an equivalence concept for them. Here, as there, we can usefully narrow our problem by focusing on truth conditions and so exploiting a method of query and assent. And here, of course, as there, the sentences that prove reasonably amenable are the occasion sentences, especially the observation sentences. What relates such a sentence to its equivalent is simply a coinciding of dispositions: we are disposed to assent to both sentences in the same circumstances.

Moreover, in a behavioural account of equivalence, just as in a behavioural account of understanding, we encounter difficulty when we move to standing sentences. Since a man is apt to assent to a standing sentence, if asked, in all sorts of circumstances or in none, the coinciding of dispositions to assent to two standing sentences gives no basis for equating them.

I am persuaded, indeed, that a satisfactory equivalence concept is impossible for standing sentences. My view of this matter can be conveyed most clearly if we consider translation between two languages. I am persuaded that alternative manuals of translation can exist, incompatible with each other, and both of them conforming fully to the dispositions to behaviour on the part of the speakers of the two languages. The two manuals would agree on observation sentences but conflict in some of the standing sentences. Each manual, being a manual of translation, purports to specify the equivalence relation between sentences and their translations, and neither manual is right to the exclusion of the other.

This indeterminacy of translation is unsuspected in mentalistic semantics, because of the facile talk of meaning. Sentences have meanings, and a translation is right if it has the same meaning. Mentalistic semantics requires that one of two conflicting manuals of translation be wrong, though it conforms to every speaker's dispositions. Mentalistic semantics thus sets a false goal, which, even though vague and ill defined, tends to obstruct other lines of thought.

Of course, translation must go on. Indeterminacy means that there is more than one way; we can still proceed to develop one of them, as good as any. And, in a more theoretical mood, we must still consider what counts as evidence for *an* acceptable translation relation, even if the relation is not unique. The evidence will be behavioural still, of course, even though the relation is no simple coinciding of behavioural dispositions, as it was in the case of the equivalence of observation sentences. We have to examine relations of interdependence between verbal dispositions: systematic interdependences between dispositions to assent to standing sentences and dispo-

sitions to assent in certain circumstances to observation sentences. Here again, in the problem of equivalence as in the problem of understanding, it would seem that genetic semantics offers a likely approach. But we must expect no simple picture, no easy answers. For it is a question again of the relations of standing sentences to observation sentences, and hence nothing less than the relation of scientific theory to scientific evidence.

Let us then recognize that the semantical study of language is worth pursuing with all the scruples of the natural scientist. We must study language as a system of dispositions to verbal behaviour, and not just surface listlessly to the Sargasso Sea of mentalism.

It has been objected that when I talk of query and assent I am not really escaping mentalism after all, because assent itself has a mental component. It is objected that assent is no mere mindless parroting of an arbitrary syllable; utterance of the syllable counts as assent only if there is the appropriate mental act behind it. Very well, let us adopt the term *surface assent* for the utterance or gesture itself. My behavioural approach does indeed permit me, then, only to appeal to surface assent; assent as I talk of it must be understood as surface assent. This behavioural notion has its powers, however, and must not be underrated. For the syllable or gesture of assent in a community is not identified at random, after all; it is itself singled out, in turn, by behavioural criteria. One partial criterion of what to count as a sign of assent is that a speaker is disposed to produce that sign whenever a sentence is queried in circumstances in which he would be disposed to volunteer the sentence himself. Even surface assent, thus, is not just the parroting of any arbitrary syllable. Granted, some cases of surface assent are insincere, but happily they are rare enough, as I have already remarked, to permit the field linguist still to find laws and translations on the strength of statistical trends.

I have been inveighing against mentalistic semantics and urging in its place the study of dispositions to behaviour. This move could be represented alternatively and more picturesquely as a matter not so much of substitution as of identification: let us *construe* mind as a system of dispositions to behaviour. This version somewhat recalls Gilbert Ryle and Wilfrid Sellars, who have urged a generally dispositional philosophy of mind. Some small further encouragement for it may be seen in the fact that even our most ordinary and characteristic mentalistic idioms already take almost the form of attributions of verbal dispositions. These are the idioms of propositional attitude: 'x believes that p', 'x wishes that p', 'x ex-

pects that *p*, and so on. They all follow the broad pattern of indirect quotation, 'x says that *p*', as if to attribute to *x* the disposition to utter the sentence '*p*' in some mood. Thus *x* believes that *p* if, approximately, he will affirm *p*; he wishes or regrets that *p* if, approximately, he will exclaim 'Oh that *p*!' or 'Alas, *p*!'

I am offering no proper analysis of the prepositional attitudes. People do not volunteer all their beliefs in affirmations. A better criterion of belief is the disposition to assent if asked; and this still leaves no room for questioning sincerity. Also there is the problem of allowable latitude of translation or paraphrase, when the '*p*' clause in 'x believes that *p*' contains words alien to *x*'s actual vocabulary. This question of allowable latitude of course arises acutely in indirect discourse itself, 'x said that *p*', and it plagues all the idioms of prepositional attitude. And finally there are quandaries over the referential opacity of the idioms of prepositional attitude: quandaries having to do with the substitutivity of identity, and with quantifying into opaque contexts. All in all, the prepositional attitudes are in a bad way. These are the idioms most stubbornly at variance with scientific patterns. Consequently I find it particularly striking that these, of all idioms, already describe mental states in a way that hints at dispositions to verbal behaviour. A philosophy of mind as verbal disposition is after all not so very alien to deep-rooted popular attitudes.

I spoke of three levels of purported explanation: the mental, the behavioural, and the physiological. We have just now been contemplating the second, the behavioural. Now the relation of this level to the third and deepest, the physiological, begins to be evident when we examine the notion of a *disposition* to behaviour and consider what we mean by a disposition.

A disposition is in my view simply a physical trait, a configuration or mechanism. It can be a disjunctive physical trait, since like effects can come of unlike mechanisms. What makes it a disposition is no significant character of its own, but only the style in which we happen to specify it. Thus take the classical example, solubility in water. This is a physical trait that can be specified, with varying degrees of thoroughness, in various ways. It can be described quite fully, I gather, in terms of the relative positions of small particles. It can also be described, less fully, by citing a simple test: put an object in water and see if it dissolves. Instructions for this convenient test are compactly encoded, as it happens, in the adjective 'soluble' itself, with its verb stem 'solu-' and its dispositional ending '-ble'. The ad-

jective 'soluble' is a disposition word, and this is an important classification of words; but the physical traits themselves do not divide into dispositions and others, any more significantly than mankind divides into passers-by and others. The term 'disposition' has its significant application rather as a preface, each time, to an actual singling out of some physical trait; thus we may significantly specify some physical trait as the disposition to behave thus and so in such-and-such circumstances. It is this that is accomplished also, and more laconically, by dispositional adjectives such as 'soluble', 'fragile', 'docile', 'portable'. The dispositional way of specifying physical traits is as frequent and as useful as it is because we are so often not prepared, as we now happen to be in the case of solubility, to specify the intended physical trait in other than the dispositional style.

The dispositional way of specifying physical states and traits is indeed pretty generally *the* way of specifying them, except at high levels of scientific theory. The explicit dispositional idiom does not always appear, either as the word 'disposition' or as a suffix '-ble' or '-ile'; commonly the dispositional force is only implicit. Hardness, for instance, is the disposition to resist if pressed, or to scratch. Redness, said of a body, is the disposition to blush in white light. Hardness and redness come finally, like solubility, to be explained in terms of minute structure, but our first access to these physical traits is dispositional. In fact the same may be said of the very notion of body itself; for a body comes to be known, as Kant remarked, by its disposition to present a repeatable sequence of views as we walk around it or revisit it. True to form, even this disposition qualifies as a physical mechanism: *body*. Like the other physical mechanisms, this one also comes in the fulness of time to be explained in terms of small particles.

John Stuart Mill's characterization of a body as 'a permanent possibility of sensation' was meant in an idealistic spirit, as a reduction of matter to sensory disposition. Thanks to symmetry, however, the identity admits also of a materialistic inversion: corporeality, like solubility, is an objective physical arrangement of particles, but known first in dispositional terms.

Dispositions to behaviour, then, are physiological states or traits or mechanisms. In citing them dispositionally we are singling them out by behavioural symptoms, behavioural tests. Usually we are in no position to detail them in physiological terms, but in this there is no anomaly; we also commonly specify ailments *per accidens*, citing gross signs and symptoms and knowing no physiological details.

We now see the relation of the second level of explanation, the behav-

journal, to the third and deepest level, the physiological. At the second level we treat of dispositions to behaviour, and these dispositions are indeed physiological states, but we identify them only by their behavioural manifestations. The deepest explanation, the physiological, would analyse these dispositions in explicit terms of nerve impulses and other anatomically and chemically identified organic processes.

Our three levels thus are levels of reduction: mind consists in dispositions to behaviour, and these are physiological states. We recall that John B. Watson did not claim that quite *all* thought was incipient speech; it was all incipient twitching of muscles, and *mostly* of speech muscles. Just so, I would not identify mind quite wholly with verbal disposition; with Ryle and Sellars I would identify it with behavioural disposition, and *mostly* verbal. And then, having construed behavioural dispositions in turn as physiological states, I end up with the so-called identity theory of mind: mental states are states of the body.

However, a word of caution is in order regarding the so-called identity theory. How does it differ from a repudiation theory? Let us think for a moment about an analogous question elsewhere, concerning the definitions of natural number in set theory. We may say that numbers are defined as sets in Frege's way, or in Zermelo's way, or in von Neumann's way, these ways all being good but incompatible. Or we may say that numbers may be repudiated, dispensed with; that we can get along with just sets instead, in any of various ways—Frege's way, Zermelo's way, von Neumann's way. This repudiation version has the advantage that we no longer seem called upon to adjudicate between three identifications of the natural numbers, the three being incompatible and yet all somehow correct.

Correspondingly, instead of saying that mental states are identical with physiological ones, we could repudiate them; we could claim that they can be dispensed with, in all our theorizing, in favour of physiological states, these being specified usually not in actual physiological terms but in the idiom of behavioural dispositions. This repudiation version has a certain advantage, though a different one from what we noted in the case of number. Its advantage here is that it discourages a possible abuse of the identity theory. For, product though the identity theory is of hard-headed materialism, we must beware of its sedative use to relieve intellectual discomfort. We can imagine someone appealing to the identity theory to excuse his own free and uncritical recourse to mentalistic semantics. We can imagine

him pleading that it is after all just a matter of physiology, even if no one knows quite how. This would be a sad irony indeed, and the repudiation theory has the virtue, over the identity theory, of precluding it.

Until we can aspire to actual physiological explanation of linguistic activity in physiological terms, the level at which to work is the middle one: that of dispositions to overt behaviour. Its virtue is not that it affords causal explanations but that it is less likely than the mentalistic level to engender an illusion of being more explanatory than it is. The easy familiarity of mentalistic talk is not to be trusted.

Still, among the dispositions to behaviour, some are more explanatory than others. The ones that we should favour, in explanations, are the ones whose physiological mechanisms seem likeliest to be detected in the foreseeable future. To cite a behavioural disposition is to posit an unexplained neural mechanism, and such posits should be made in the hope of their submitting some day to a physical explanation.

The Nature of Natural Knowledge

1975

Doubt has oft been said to be the mother of philosophy.¹ This has a true ring for those of us who look upon philosophy primarily as the theory of knowledge. For the theory of knowledge has its origin in doubt, in scepticism. Doubt is what prompts us to try to develop a theory of knowledge. Furthermore, doubt is also the first step to take in developing a theory of knowledge, if we adopt the line of Descartes.

But this is only half of a curious interplay between doubt and knowledge. Doubt prompts the theory of knowledge, yes; but knowledge, also, was what prompted the doubt. Scepticism is an offshoot of science. The basis for scepticism is the awareness of illusion, the discovery that we must not always believe our eyes. Scepticism battens on mirages, on seemingly bent sticks in water, on rainbows, after-images, double images, dreams. But in what sense are these illusions? In the sense that they seem to be material objects which they in fact are not. Illusions are illusions only relative to a prior acceptance of genuine bodies with which to contrast them. In a world of immediate sense data with no bodies posited and no questions asked, a distinction between reality and illusion would have no place. The positing of bodies is already rudimentary physical science; and it is only after that stage that the sceptic's invidious distinctions can make sense. Bodies have to be posited before there can be a motive, however tenuous, for acquiescing in a non-committal world of the immediate given.

"The Nature of Natural Knowledge" copyright © 1975 by Oxford University Press.

1. This paper is meant as a summary statement of my attitude towards our knowledge of nature. Consequently I must warn the more omnivorous of my readers (dear souls) that they are apt to experience a certain indefinable sense of *déjà lu*. The main traces of novelty come towards the end.

Rudimentary physical science, that is, common sense about bodies, is thus needed as a springboard for scepticism. It contributes the needed notion of a distinction between reality and illusion, and that is not all. It also discerns regularities of bodily behaviour which are indispensable to that distinction. The sceptic's example of the seemingly bent stick owes its force to our knowledge that sticks do not bend by immersion; and his examples of mirages, after-images, dreams, and the rest are similarly parasitic upon positive science, however primitive.

I am not accusing the sceptic of begging the question. He is quite within his rights in assuming science in order to refute science; this, if carried out, would be a straightforward argument by *reductio ad absurdum*. I am only making the point that sceptical doubts are scientific doubts.

Epistemologists have coped with their sceptical doubts by trying to reconstruct our knowledge of the external world from sensations. A characteristic effort was Berkeley's theory of vision, in which he sought our clues for a third dimension, depth, in our two-dimensional visual field. The very posing of this epistemological problem depends in a striking way upon acceptations of physical science. The goal of the construction, namely the depth dimension, is of course deliberately taken from the science of the external world; but what particularly wants noticing is that also the accepted basis of the construction, the two-dimensional visual field, was itself dictated by the science of the external world as well. The light that informs us of the external world impinges on the two-dimensional surface of the eye, and it was Berkeley's awareness of this that set his problem.

Epistemology is best looked upon, then, as an enterprise within natural science. Cartesian doubt is not the way to begin. Retaining our present beliefs about nature, we can still ask how we can have arrived at them. Science tells us that our only source of information about the external world is through the impact of light rays and molecules upon our sensory surfaces. Stimulated in these ways, we somehow evolve an elaborate and useful science. How do we do this, and why does the resulting science work so well? These are genuine questions, and no feigning of doubt is needed to appreciate them. They are scientific questions about a species of primates, and they are open to investigation in natural science, the very science whose acquisition is being investigated.

The utility of science, from a practical point of view, lies in fulfilled expectation: true prediction. This is true not only of sophisticated science, but of its primitive progenitor as well; and it may be good strategy on our

part to think first of the most primitive case. This case is simple induction. It is the expectation, when some past event recurs, that the sequel of that past event will recur too. People are prone to this, and so are other animals.

It may be felt that I am unduly intellectualizing the dumb animals in attributing expectation and induction to them. Still the net resultant behaviour of dumb animals is much on a par with our own, at the level of simple induction. In a dog's experience, a clatter of pans in the kitchen has been followed by something to eat. So now, hearing the clatter again, he goes to the kitchen in expectation of dinner. His going to the kitchen is our evidence of his expectation, if we care to speak of expectation. Or we can skip this intervening variable, as Skinner calls it, and speak merely of reinforced response, conditioned reflex, habit formation.

When we talk easily of repetition of events, repetition of stimuli, we cover over a certain significant factor. It is the *similarity* factor. It can be brought into the open by speaking of events rather as unique, dated, un-repeated particulars, and then speaking of similarities between them. Each of the noisy episodes of the pans is a distinct event, however similar, and so is each of the ensuing dinners. What we can say of the dog in those terms is that he hears something similar to the old clatter and proceeds to expect something similar to the old dinner. Or, if we want to eliminate the intervening variable, we can still say this: when the dog hears something similar to the old clatter and, going to the kitchen, gets something similar to the old dinner, he is reinforced in his disposition to go to the kitchen after each further event similar to the old clatter.

What is significant about this similarity factor is its subjectivity. Is similarity the mere sharing of many attributes? But any two things share countless attributes—or anyway any two objects share membership in countless classes. The similarity that matters, in the clatter of the pans, is similarity for the dog. Again I seem to appeal to the dog's mental life, but again I can eliminate this intervening variable. We can analyse similarity, for the dog, in terms of his dispositions to behaviour: his patterns of habit formation. His habit of going to the kitchen after a clatter of pans is itself our basis for saying that the clatter events are similar for the dog, and that the dinner events are similar for the dog. It is by experimental reinforcement and extinction along these lines that we can assess similarities for the dog, determining whether event *a* is more similar to *b* than to *c* for him. Meanwhile his mental life is as may be.

Now our question 'Why is science so successful?' makes some rudimen-

tary sense already at this level, as applied to the dog. For the dog's habit formation, his primitive induction, involved extrapolation along similarity lines: episodes similar to the old clattering episode engendered expectation of episodes similar to the old dinner episode. And now the crux of the problem is the subjectivity of similarity. Why should nature, however lawful, match up at all with the dog's subjective similarity ratings? Here, at its most primitive, is the question 'Why is science so successful?'

We are taking this as a scientific question, remember, open to investigation by natural science itself. Why should the dog's implicit similarity ratings tend to fit world trends, in such a way as to favour the dog's implicit expectations? An answer is offered by Darwin's theory of natural selection. Individuals whose similarity groupings conduce largely to true expectations have a good chance of finding food and avoiding predators, and so a good chance of living to reproduce their kind.

What I have said of the dog holds equally of us, at least in our pursuit of the rudimentary science of common sense. We predict in the light of observed uniformities, and these are uniformities by our subjective similarity standards. These standards are innate ones, overlaid and modified by experience; and natural selection has endowed us, like the dog, with a head start in the way of helpful, innate similarity standards.

I am not appealing to Darwinian biology to justify induction. This would be circular, since biological knowledge depends on induction. Rather I am granting the efficacy of induction, and then observing that Darwinian biology, if true, helps explain why induction is as efficacious as it is.

We must notice, still, a further limitation. Natural selection may be expected only to have encouraged similarity standards conducive to rough and ready anticipations of experience in a state of nature. Such standards are not necessarily conducive to deep science. Colour is a case in point. Colour dominates our scene; similarity in colour is similarity at its most conspicuous. Yet, as J. J. C. Smart points out, colour plays little role in natural science. Things can be alike in colour even though one of them is reflecting green light of uniform wave length while the other is reflecting mixed waves of yellow and blue. Properties that are most germane to sophisticated science are camouflaged by colour more than revealed by it. Over-sensitivity to colour may have been all to the good when we were bent on quickly distinguishing predator from prey or good plants from bad. But true science cuts through all this and sorts things out differently, leaving colour largely irrelevant.

Colour is not the only such case. Taxonomy is rich in examples to show that visual resemblance is a poor index of kinship. Natural selection has even abetted the deception; thus some owls have grown to resemble cats, for their own good, and others resemble monkeys. Natural selection works both to improve a creature's similarity standards and to help him abuse his enemies' similarity standards.

For all their fallibility, our innate similarity standards are indispensable to science as an entering wedge. They continue to be indispensable, moreover, even as science advances. For the advance of science depends on continued observation, continued checking of predictions. And there, at the observational level, the unsophisticated similarity standards of common sense remain in force.

An individual's innate similarity standards undergo some revision, of course, even at the common-sense level, indeed even at the subhuman level, through learning. An animal may learn to tell a cat from an owl. The ability to learn is itself a product of natural selection, with evident survival value. An animal's innate similarity standards are a rudimentary instrument for prediction, and then learning is a progressive refinement of that instrument, making for more dependable prediction. In man, and most conspicuously in recent centuries, this refinement has consisted in the development of a vast and bewildering growth of conceptual or linguistic apparatus, the whole of natural science. Biologically, still, it is like the animal's learning about cats and owls; it is a learned improvement over simple induction by innate similarity standards. It makes for more and better prediction.

Science revises our similarity standards, we saw; thus we discount colour, for some purposes, and we liken whales to cows rather than to fish. But this is not the sole or principal way in which science fosters prediction. Mere improvement of similarity standards would increase our success at simple induction, but this is the least of it. Science departs from simple induction. Science is a ponderous linguistic structure, fabricated of theoretical terms linked by fabricated hypotheses, and keyed to observable events here and there. Indirectly, via this labyrinthine superstructure, the scientist predicts future observations on the basis of past ones; and he may revise the superstructure when the predictions fail. It is no longer simple induction. It is the hypothetico-deductive method. But, like the animal's simple induction over innate similarities, it is still a biological device for anticipating experience. It owes its elements still to natural selection—notably,

the similarity standards that continue to operate at the observational level. The biological survival value of the resulting scientific structure, however, is as may be. Traits that were developed by natural selection have been known to prove lethal, through overdevelopment or remote effects or changing environment. In any event, and for whatever good it may do us, the hypothetico-deductive method is delivering knowledge hand over fist. It is facilitating prediction.

I said that science is a linguistic structure that is keyed to observation at some points. Some sentences are keyed directly to observation: the observation sentences. Let us examine this connection. First I must explain what I mean by an observation sentence. One distinctive trait of such a sentence is that its truth value varies with the circumstances prevailing at the time of the utterance. It is a sentence like 'This is red' or 'It is raining', which is true on one occasion and false on another; unlike 'Sugar is sweet', whose truth value endures regardless of occasion of utterance. In a word, observation sentences are occasion sentences, not standing sentences.

But their being occasion sentences is not the only distinctive trait of observation sentences. Not only must the truth value of an observation sentence depend on the circumstances of its utterance; it must depend on intersubjectively observable circumstances. Certainly the fisherman's sentence 'I just felt a nibble' is true or false depending on the circumstances of its utterance; but the relevant circumstances are privy to the speaker rather than being out in the open for all present witnesses to share. The sentence 'I just felt a nibble' is an occasion sentence but not an observation sentence, in my sense of the term.

An observation sentence, then, is an occasion sentence whose occasion is intersubjectively observable. But this is still not enough. After all, the sentence 'There goes John's old tutor' meets these requirements; it is an occasion sentence, and all present witnesses can see the old tutor plodding by. But the sentence fails of a third requirement: the witnesses must in general be able to appreciate that the observation which they are sharing is one that verifies the sentence. They must have been in a position, equally with the speaker, to have assented to the sentence on their own in the circumstances. They are in that position in the case of 'This is red' and 'It is raining' and 'There goes an old man', but not in the case of 'There goes John's old tutor.'

Such, then, is an observation sentence: it is an occasion sentence whose occasion is not only intersubjectively observable but is generally adequate,

moreover, to elicit assent to the sentence from any present witness conversant with the language. It is not a report of private sense data; typically, rather, it contains references to physical objects.

These sentences, I say, are keyed directly to observation. But how *keyed*, now—what is the nature of the connection? It is a case of conditioned response. It is not quite the simplest kind; we do not say ‘red’ or ‘This is red’ whenever we see something red. But we do assent if asked. Mastery of the term ‘red’ is acquisition of the habit of assenting when the term is queried in the presence of red, and only in the presence of red.

At the primitive level, an observation sentence is apt to take the form of a single word, thus ‘ball’, or ‘red’. What makes it easy to learn is the intersubjective observability of the relevant circumstances at the time of utterance. The parent can verify that the child is seeing red at the time, and so can reward the child’s assent to the query. Also the child can verify that the parent is seeing red when the parent assents to such a query.

In this habit formation the child is in effect determining, by induction, the range of situations in which the adult will assent to the query ‘red’, or approve the child’s utterance of ‘red’. He is extrapolating along similarity lines; this red episode is similar to that red episode by his lights. His success depends, therefore, on substantial agreement between his similarity standards and those of the adult. Happily the agreement holds; and no wonder, since our similarity standards are a matter partly of natural selection and partly of subsequent experience in a shared environment. If substantial agreement in similarity standards were not there, this first step in language acquisition would be blocked.

We have been seeing that observation sentences are the starting points in the learning of language. Also, they are the starting points and the check points of scientific theory. They serve both purposes for one and the same reason: the intersubjective observability of the relevant circumstances at the time of utterance. It is this, intersubjective observability at the time, that enables the child to learn when to assent to the observation sentence. And it is this also, intersubjective observability at the time, that qualifies observation sentences as check points for scientific theory. Observation sentences state the evidence, to which all witnesses must accede.

I had characterized science as a linguistic structure that is keyed to observation at some points. Now we have seen how it is keyed to observation: some of the sentences, the observation sentences, are conditioned to observable events in combination with a routine of query and assent. There is

the beginning, here, of a partnership between the theory of language learning and the theory of scientific evidence. It is clear, when you think about it, that this partnership must continue. For when a child learns his language from his elders, what has he to go on? He can learn observation sentences by consideration of their observable circumstances, as we saw. But how can he learn the rest of the language, including the theoretical sentences of science? Somehow he learns to carry his observation terms over into theoretical contexts, variously embedded. Somehow he learns to connect his observation sentences with standing sentences, sentences whose truth values do not depend on the occasion of utterance. It is only by such moves, however ill understood, that anyone masters the non-observational part of his mother tongue. He can learn the observational part in firm and well-understood ways, and then he must build out somehow, imitating what he hears and linking it tenuously and conjecturally to what he knows, until by dint of trial and social correction he achieves fluent dialogue with his community. This discourse depends, for whatever empirical content it has, on its devious and tenuous connections with the observation sentences; and those are the same connections, nearly enough, through which one has achieved one's fluent part in that discourse. The channels by which, having learned observation sentences, we acquire theoretical language, are the very channels by which observation lends evidence to scientific theory. It all stands to reason; for language is man-made and the locutions of scientific theory have no meaning but what they acquired by our learning to use them.

We see, then, a strategy for investigating the relation of evidential support, between observation and scientific theory. We can adopt a genetic approach, studying how theoretical language is learned. For the evidential relation is virtually enacted, it would seem, in the learning. This genetic strategy is attractive because the learning of language goes on in the world and is open to scientific study. It is a strategy for the scientific study of scientific method and evidence. We have here a good reason to regard the theory of language as vital to the theory of knowledge.

When we try to understand the relation between scientific theory and the observation sentences, we are brought up short by the break between occasion sentences and standing sentences; for observation sentences are of the one kind while theoretical sentences are of the other. The scientific system cannot digest occasion sentences; their substance must first be converted into standing sentences. The observation sentence 'Rain' or 'It is

raining' will not do; we must put the information into a standing sentence: 'Rain at Heathrow 1600 G.M.T. 23 February 1974.' This report is ready for filing in the archives of science. It still reports an observation, but it is a standing report rather than an occasion sentence. How do we get from the passing observation of rain to the standing report?

This can be explained by a cluster of observations and observation sentences, having to do with other matters besides the rain. Thus take the term 'Heathrow'. Proper names of persons, buildings, and localities are best treated as observation terms, on a par with 'red' and 'rain'. All such terms can be learned by ostension, repeated sufficiently to suggest the intended scope and limits of application. 'Here is Heathrow', then, is an observation sentence on a par with 'It is raining'; and their conjunction, 'Raining at Heathrow', is an observation sentence as well. It is an occasion sentence still, of course, and not a standing report of observation. But now the two further needed ingredients, hour and date, can be added as pointer readings: 'The clock reads 1600' and 'The calendar reads 23 February 1974' are further observation sentences. Taking the conjunction of all four, we still have an observation sentence: 'Rain at Heathrow with clock at 1600 and calendar at 23 February 1974.' But it is an observation sentence with this curious trait: it gives lasting information, dependent no longer on the vicissitudes of tense or of indicator words like 'here' and 'now'. It is suitable for filing.

True, the clock and calendar may have been wrong. As an observation sentence our report must be viewed as stating the temporal readings and not the temporal facts. The question of the temporal facts belongs to scientific theory, somewhat above the observational level. Theoretical repercussions of this and other observations could eventually even prompt a modest scientific hypothesis to the effect that the clock or the calendar had been wrong.

I think this example serves pretty well as a paradigm case, to show how we can get from the occasion sentences of observation to the standing reports of observation that are needed for scientific theory. But this connection is by no means the only connection between observation sentences and standing sentences. Thus consider the universal categorical, 'A dog is an animal'. This is a standing sentence, but it is not, like the example of rain at Heathrow, a standing report of observation. Let us resume our genetic strategy: how might a child have mastered such a universal categorical?

I shall venture one hypothesis, hoping that it may be improved upon.

The child has learned to assent to the observation term 'a dog' when it is queried in the conspicuous presence of dogs, and he has learned to assent to 'an animal' likewise when it is queried in the conspicuous presence of dogs (though not only dogs). Because of his close association of the word 'dog' with dogs, the mere sound of the word 'dog' disposes him to respond to the subsequent query 'an animal' as he would have done if a dog had been there; so he assents when he hears 'a dog' followed by the query, 'an animal?'. Being rewarded for so doing, he ever after assents to the query 'A dog is an animal?' In the same way he learns a few other examples of the universal categorical. Next he rises to a mastery of the universal categorical construction 'An *S* is a *P*' in general: he learns to apply it to new cases on his own. This important step of abstraction can perhaps be explained in parallel fashion to the early learning of observation sentences, namely, by simple induction along similarity lines; but the similarity now is a language-dependent similarity.

Much the same account can be offered for the learning of the seemingly simpler construction, mere predication: 'Fido is a dog', 'Sugar is sweet'.

The child has now made creditable progress from observation sentences towards theoretical language, by mastering predication and the universal categorical construction. Another important step will be mastery of the relative clause; and I think I can give a convincing hypothesis of how this comes about. What is conspicuous about the relative clause is its role in predication. Thus take a relative clause, 'something that chases its tail', and predicate it of Dinah: 'Dinah is something that chases its tail.' This is equivalent to the simple sentence 'Dinah chases its tail' (or 'her tail'). When we predicate the relative clause, the effect is the same as substituting the subject of the predication for the pronoun of the relative clause. Now my suggestion regarding the learning of the relative clause is that the child learns this substitution transformation. He discovers that the adult is prepared to assent to a predication of a relative clause in just the circumstances where he is prepared to assent to the simpler sentence obtained by the substitution.

This explains how the child could learn relative clauses in one standard position: predicative position. He learns how to eliminate them, in that position, by the substitution transformation—and how to introduce them into that position by the converse transformation, superstition. But then, having learned this much, he is struck by an analogy between relative clauses and ordinary simple predicates or general terms; for these also ap-

pear in predicative position. So, pursuing the analogy, he presses relative clauses into other positions where general terms have been appearing—notably into the universal categorical construction. Or, if the child does not press this analogy on his own, he is at any rate well prepared to grasp adult usage and follow it in the light of the analogy. In this way the relative clause gets into the universal categorical construction, from which it cannot be eliminated by the substitution transformation. It is there to stay.

We can easily imagine how the child might learn the truth functions—negation, conjunction, alternation. Take conjunction: the child notices, by degrees, that the adult affirms ' p and q ' in only those circumstances where he is disposed, if queried, to assent to ' p ' and also to ' q '.

We have now seen, in outline and crude conjecture, how one might start at the observational edge of language and work one's way into the discursive interior where scientific theory can begin to be expressed. Predication is at hand, and the universal categorical, the relative clause, and the truth functions. Once this stage is reached, it is easy to see that the whole strength of logical quantification is available. I shall not pause over the details of this, except to remark that the pronouns of relative clauses take on the role of the bound variable of quantification. By further conjectures in the same spirit, some of them more convincing and some less, we can outline the learner's further progress, to where he is bandying abstract terms and quantifying over properties, numbers, functions, and hypothetical physical particles and forces. This progress is not a continuous derivation, which, followed backward, would enable us to reduce scientific theory to sheer observation. It is a progress rather by short leaps of analogy. One such was the pressing of relative clauses into universal categoricals, where they cease to be eliminable. There are further such psychological speculations that I could report, but time does not allow.

Such speculations would gain, certainly, from experimental investigation of the child's actual learning of language. Experimental findings already available in the literature could perhaps be used to sustain or correct these conjectures at points, and further empirical investigations could be devised. But a speculative approach of the present sort seems required to begin with, in order to isolate just the factual questions that bear on our purposes. For our objective here is still philosophical—a better understanding of the relations between evidence and scientific theory. Moreover, the way to this objective requires consideration of linguistics and logic along with psychology. This is why the speculative phase has to precede, for

the most part, the formulation of relevant questions to be posed to the experimental psychologist.

In any event the present speculations, however inaccurate, are presumably true to the general nature of language acquisition. And already they help us to understand how the logical links are forged that connect theoretical sentences with the reports of observation. We learn the grammatical construction ' p and q ' by learning, among other things, to assent to the compound only in circumstances where we are disposed to assent to each component. Thus it is that the logical law of inference which leads from ' p and q ' to ' p ' is built into our habits by the very learning of 'and'. Similarly for the other laws of conjunction, and correspondingly for the laws of alternation and other truth functions. Correspondingly, again, for laws of quantification. The law of inference that leads from ' $(x)Fx$ ' to ' Fa ' should be traceable back, through the derivation of quantification that I have passed over, until it is found finally to hinge upon the substitution transformation by which we learn to use the relative clause. Thus, in general, the acquisition of our basic logical habits is to be accounted for in our acquisition of grammatical constructions.

Related remarks hold true of inferential habits that exceed pure logic. We learn when to assent to 'dog', and to 'animal', only by becoming disposed to assent to 'animal' in all circumstances where we will assent to 'dog'. Connections more accidental and casual in aspect can also come about through the learning of words; thus a child may have begun to learn the term 'good' in application to chocolate.

I characterized science as a linguistic structure that is keyed to observation here and there. I said also that it is an instrument for predicting observations on the basis of earlier observations. It is keyed to observations, earlier and later, forming a labyrinthine connection between them; and it is through this labyrinth that the prediction takes place. A powerful improvement, this, over simple induction from past observations to future ones; powerful and costly. I have now sketched the nature of the connection between the observations and the labyrinthine interior of scientific theory. I have sketched it in terms of the learning of language. This seemed reasonable, since the scientist himself can make no sense of the language of scientific theory beyond what goes into his learning of it. The paths of language learning, which lead from observation sentences to theoretical sentences, are the only connection there is between observation and theory. This has been a sketch, but a fuller understanding may be sought along the same

line: by a more painstaking investigation of how we learn theoretical language.

One important point that already stands forth, regarding the relation of theory to observation, is the vast freedom that the form of the theory must enjoy, relative even to all possible observation. Theory is empirically under-determined. Surely even if we had an observational oracle, capable of assigning a truth value to every standing observational report expressible in our language, still this would not suffice to adjudicate between a host of possible physical theories, each of them completely in accord with the oracle. This seems clear in view of the tenuousness of the connections that we have noted between observation sentences and theoretical ones. At the level of observation sentences, even the general form of the eventual theoretical language remained indeterminate, to say nothing of the ontology. The observation sentences were associated, as wholes, with the stimulatory situations that warranted assent to them; but there was in this no hint of what aspects of the stimulatory situations to single out somehow as objects, if indeed any. The question of ontology simply makes no sense until we get to something recognizable as quantification, or perhaps as a relative clause, with pronouns as potential variables. At the level of observation sentences there was no foreseeing even that the superimposed theoretical language would contain anything recognizable as quantification or relative clauses. The steps by which the child was seen to progress from observational language to relative clauses and categoricals and quantification had the arbitrary character of historical accident and cultural heritage; there was no hint of inevitability.

It was a tremendous achievement, on the part of our long-term culture and our latter-day scientists, to develop a theory that leads from observation to predicted observation as successfully as ours. It is a near miracle. If our theory were in full conformity with the observational oracle that we just now imagined, which surely it is not, that would be yet a nearer miracle. But if, even granted that nearer miracle, our theory were not still just one of many equally perfect possible theories to the same observational effect, that would be too miraculous to make sense.

But it must be said that the issue of under-determination proves slippery when we try to grasp it more firmly. If two theories conform to the same totality of possible observations, in what sense are they two? Perhaps they are both stated in English, and they are alike, word for word, except that one of them calls molecules electrons and electrons molecules. Li-

terally the two theories are in contradiction, saying incompatible things about so-called molecules. But of course we would not want to count this case; we would call it terminological. Or again, following Poincaré, suppose the two theories are alike except that one of them assumes an infinite space while the other has a finite space in which bodies shrink in proportion to their distance from centre. Even here we want to say that the difference is rather terminological than real; and our reason is that we see how to bring the theories into agreement by translation: by reconstruing the English of one of the theories. At this point it may be protested that after all there can never be two complete theories agreeing on the total output of the observational oracle. It may be protested that since such theories would be empirically equivalent, would have the same empirical meaning, their difference is purely verbal. For surely there is no meaning but empirical meaning, and theories with the same meaning must be seen as translations one of the other. This argument simply rules out, by definition, the doctrine that physical theory is under-determined by all possible observation.

The best reaction at this point is to back away from terminology and sort things out on their merits. Where the significant difference comes is perhaps where we no longer see how to state rules of translation that would bring the two empirically equivalent theories together. Terminology aside, what wants recognizing is that a physical theory of radically different form from ours, with nothing even recognizably similar to our quantification or objective reference, might still be empirically equivalent to ours, in the sense of predicting the same episodes of sensory bombardment on the strength of the same past episodes. Once this is recognized, the scientific achievement of our culture becomes in a way more impressive than ever. For, in the midst of all this formless freedom for variation, our science has developed in such a way as to maintain always a manageably narrow spectrum of visible alternatives among which to choose when need arises to revise a theory. It is this narrowing of sights, or tunnel vision, that has made for the continuity of science, through the vicissitudes of refutation and correction. And it is this also that has fostered the illusion of there being only one solution to the riddle of the universe.

Facts of the Matter

1977

It was emphasized by rationalists and empiricists alike that inquiry should begin with clear ideas. I agree about the clarity, but I balk at ideas. The British empiricists themselves balked at abstract ideas. *Nihil in mente*, they declared in their orotund British measures, *quod non prius in sensu*. They echoed their nominalist ancestors, for whom abstract ideas were *flatus vocis*—words, words, words.

What then about concrete ideas? Even a strictly sensory idea is elusive unless it is reinforced by language. This point was made by Wittgenstein. Unaided by language, we might treat a great lot of sensory events as recurrences of one and the same sensation, simply because of a similarity between each and the next; and yet there can have been a serious cumulative slippage of similarity between the latest of these events and the earliest of them. But if we have learned society's word for the sensation, then social intercourse will arrest the drift and keep us in line. We will be saved by the statistical fact that the speakers have not all drifted in the same direction.

Let us therefore recognize that the whole idea idea, abstract and concrete, is a frail reed indeed. We must seek a firm footing rather in words. The point was urged by John Horne Tooke only shortly after Hume's time, in 1786. Tooke held that Locke's essay could be much improved by substituting the word 'word' everywhere for the word 'idea.' What is thereby gained in firmness is attended by no appreciable loss in scope, since ideas without words would have come to little in any event. We think mostly

W. V. Quine, "Facts of the Matter," in Robert W. Shahan and Kenneth R. Merrill, eds., *American Philosophy from Edwards to Quine*. Copyright © 1977 (Norman: University of Oklahoma Press). All rights reserved.

in words, and we report our thoughts wholly in words. Let us then take one leaf from the old-time philosophy and another from John Horne Tooke. Philosophical inquiry should begin with the clear, yes; but with clear words.

And what words are those? It will not do to say that they are the words that express clear ideas, or the words that clearly express ideas, for we are fleeing the idea idea. For a standard of clarity of language we must look rather to the social character of language and the use of language in communication. Bypassing the idea idea, we can still do something with clarity of communication. The vehicle of communication is the sentence, and one mark of clarity of communication is agreement as to the truth of the sentence. This is a very fallible criterion, but it is a beginning. Let us see what we can do to improve it.

If one party affirms a sentence and the other party assents, this gives little evidence of communication; for a purely random verdict would be affirmative half the time. However, there is some safety in numbers. Instead of relativizing our clarity criterion to two communicants, we may relativize it to increasing sectors of the speech community. We might consider what proportion of the community would be prepared to agree to the truth or falsity of a sentence, and we might take this figure as a measure of the clarity of the sentence.

This is better, but it still will not do. One difficulty is that there are cults, fads, and slogans that can sweep a community, prompting widespread agreement to the truth of sentences that a clear thinker would not rate as clear in the slightest. Another and opposite difficulty is that people can disagree regarding the truth of a sentence even when the sentence would be said to be clear. Now both of these difficulties can be met by appealing once again to numbers: by appealing to what Mill called concomitant variation.

For this purpose we direct our attention to a special sort of sentences, *occasion* sentences. These are the sentences that admit of verdicts and truth values not once for all but from one occasion of utterance to another, depending on what is going on in the neighborhood. They are sentences like 'It's raining', 'This is red', 'That's his uncle', 'He owes me money', 'There goes a rabbit'. Historical truths are not among them, nor are scientific hypotheses, nor credos, nor slogans. Now we might measure the clarity of an occasion sentence by the readiness of witnesses to agree in their verdicts on it from occasion to occasion. By this standard 'It's raining' and 'This is

red' rate high; 'There goes a rabbit' not quite so high; 'That's his uncle' rates lower; 'He owes me money' lower still. There are three possible responses—assent, dissent, and abstention—and we may also distinguish degrees of hesitation. The great value of this standard of clarity lies in its linking of language to nonlinguistic reality.

The occasion sentences that pass this clarity test with high marks are what I call observation sentences. They often take the form simply of single nouns or adjectives—'Rabbit', 'Raining'—but for our purposes they are best thought of still as sentences, admitting of assent or dissent in the light of each present local situation. They are expressions that we have learned to associate with publicly observable concurrent circumstances. Previous speakers have taught us some of these expressions by direct conditioning to the circumstances. They are circumstances which, thanks to their public character, can be appreciated jointly by us and our teachers. Some of these observational expressions also are learned indirectly by some of us, through explanations in other words; but all could be learned in the direct way, such is their observational character. They are our introduction to language, for they are the expressions that we can learn to use without learning to use others first. It is through them that language and science imbibe their empirical content. It is back to or toward them, also, that a scientist reverts when he is mustering evidence for a disputed hypothesis; for the distinctive trait of an observation sentence is that present witnesses will usually agree about it on the spot.

Earlier I made Wittgenstein's point: how public language anchors experience, arresting drift. Now we are noting the converse: how public experience anchors language. The observation sentence is the anchor line.

I remarked that the use of an observation sentence often is and always could be acquired directly by conditioning. This process is also called induction. By either name it is the learning process at its simplest. If an event resembles an earlier one, the subject tends to expect its sequel to resemble the sequel of the earlier one. The expectation hinges thus on similarity in some sense—similarity by the subject's own lights. This relation is one of subjective similarity, and no significance need be sought for it apart from the consequent inductive expectations themselves. From a behavioral point of view a subject's expectations are shown by his overt behavior, and his similarity standards are shown by the pattern of his expectations.

Expectations are in large part fulfilled, despite the subjectivity of similarity standards; ours is a fairly friendly world. Evolutionary biology ex-

plains this by the fact that those standards are largely innate and thus favored by natural selection according to their survival value.

Primitive inductive learning is evident in the acquisition of various observation sentences. To acquire an observation sentence is to learn when to expect a veteran speaker to approve one's utterance of it, or to assent to it on his own account. This can be learned from sample instances by induction: by extrapolating to further cases along lines of subjective similarity. These linguistic inductions tend to be highly successful—more so still than the general run of inductions in our fairly friendly world. The reason is that, whereas one's inductions regarding nature owe their success only to a rough congruity between one's similarity standards and the trend of events in nature, on the other hand one's inductions regarding the veteran speaker's assent to the observation sentence owe their success to a sharing of similarity standards by the speaker and oneself. Heredity, environment, and social interaction have fostered such sharing of similarity standards to a high degree.

Direct conditioning or simple induction does not suffice for the acquisition of language generally. The learning process has to be more elaborate when we move on to grammatical constructions, to past and future tenses, to conditionals and conjecturals and metaphors, and to theoretical and abstract terms. It is evident that these further linguistic structures are based, however precariously, on the observational vocabulary that was learned by direct confrontation and simple conditioning. The superstructure is cantilevered outward from that foundation by imitation and analogy, by trial and error. In the course of mastering it we may check up now and again by noting the reaction of the listener. But it is in the observational vocabulary that language makes its principal contact with experience. It is this part of language that we first learn to apply, and to which we retreat when a check point is needed.

The situations that command assent to a given observation sentence will not be quite alike. They will be similar by our lights and by the lights of other speakers. But we can count on a curious tolerance of spatial reorientation in these similarity standards. We can see why if we reflect that the language learner and his informant are not situated eye to eye. They see things from unlike angles, receiving somewhat unlike presentations. The learner is thus made to associate with his presentation a word or occasion sentence that was elicited from the informant by a somewhat different pre-

sentation. It will have to be a versatile word or sentence indifferently applicable throughout a whole group of presentations.

My talking of observation sentences rather than observation terms is a matter of first things first. We can learn to assent to and dissent from observation sentences as wholes, under appropriate stimulatory conditions, with no thought of what sentences or parts of sentences to count as terms or what objects to count them as referring to. And now what happens when at last we can be said to use some of these sentences or parts of sentences as terms denoting some sort of supposed objects?

The main thing to settle, in the way of fixing the objects, is their individuation: we have to fix standards of sameness and difference. Now it is clear that at this point little or no attention will be paid to differences of perspective; for we saw that such differences are bound to be transcended in the learning of words. What are posited as objects for the terms to refer to will be, primarily, objects that are counted identical under changes of perspective. This explains the primacy of bodies. If clarity can be ascribed to things as well as to words, then bodies are things at their clearest. If inquiry is to begin with what is clear, then let us begin as physicalists.

The move from sentences to terms is already a major step in language learning. On the one hand there is the simple observation sentence 'Rabbit', comparable to 'Red' or 'It's raining'; it commands assent in the presence of rabbits. On the other hand there is the term 'rabbit', which denotes the rabbits. A speaker may be said to have mastered this term, and to have achieved objective reference, only when he has learned to subject the term to all the grammatical apparatus of particles and constructions that go to implement objective reference: the apparatus of singular and plural, of definite and indefinite articles, of pronominal cross reference, of identity and distinctness, and of counting. When he has come this far, he has risen above the primitive base afforded by observation sentences, and has ventured somewhat out onto the cantilevered superstructure. Language learning at this stage is beyond the reach of simple induction; it proceeds by imitation and analogy in more complicated ways.¹

Various of the one-word observation sentences like 'Rabbit' and 'Apple', which were themselves learned in the simple inductive way, will now

1. See W. V. O. Quine, *The Roots of Reference* (LaSalle, Ill.: Open Court, 1973) for a speculative account of the steps involved.

spawn terms in their likeness—terms denoting bodies. The terms are already theoretical. A body is conceived as retaining its identity over time between appearances. Whether we encounter the same body the next time around, the same apple, for instance, or only another one like it, is a question not to be settled by simple induction. It is settled, if at all, by inference from a network of hypotheses that we have internalized little by little in the course of acquiring the nonobservational superstructure of our language. These hypotheses are supported only indirectly by past observation: they owe their plausibility to our having inferred other consequences from them that were borne out by observation. Such is the continuing method of science: not simple induction, but the hypothetico-deductive method.

Bodies are basic to our way of thought, as objects go. They are the paradigmatic objects, clearer and more perspicuous than others. Imitation and analogy continue their work, however, not stopping with an ontology of bodies. Grammatical analogy between general terms and singular terms encourages us to treat a general term as if it designated a single object, and thus we come to posit a realm of objects for the general terms to designate: a realm of properties, or sets. What with the nominalizing also of verbs and clauses, a vaguely varied and very untidy ontology grows up.

The common man's ontology is vague and untidy in two ways. It takes in many purported objects that are vaguely or inadequately defined. But also, what is more significant, it is vague in its scope; we cannot even tell in general which of these vague things to ascribe to a man's ontology at all, which things to count him as assuming. Should we regard grammar as decisive? Does every noun demand some array of denotata? Surely not; the nominalizing of verbs is often a mere stylistic variation. But where can we draw the line?

It is a wrong question; there is no line to draw. Bodies are assumed, yes; they are the things, first and foremost. Beyond them there is a succession of dwindling analogies. Various expressions come to be used in ways more or less parallel to the use of the terms for bodies, and it is felt that corresponding objects are more or less posited, *pari passu*; but there is no purpose in trying to mark an ontological limit to the dwindling parallelism.

It is only our somewhat regimented and sophisticated language of science that has evolved in such a way as really to raise ontological questions. It is an object-oriented idiom. Any idiom purports to tell the truth, but this idiom purports, more specifically, to tell about objects. Its referential apparatus, the apparatus for referring to objects, is explicit; there is no question

of a dwindling parallelism. Just what those objects are—what else besides bodies—is still as may be; but it becomes a significant question, and it can be variously answered in various scientific systems of the world.

The basic structure of the language of science has been isolated and schematized in a familiar form. It is the predicate calculus: the logic of quantification and truth functions. In representing it thus I do not mean to take issue with those quantum physicists who recommend a different logic of a non-truth-functional kind, but I set them aside in order not to complicate the picture. Also I do not mean to deprecate alternative formulations of standard logic, such as predicate-functor logic; but as long as these are intertranslatable with the classical predicate calculus, we lose nothing in adhering to the latter. For concreteness, then, let us adhere to it; for it is familiar.

Language thus regimented has a simple grammar. There is a lexicon of predicates. Each atomic sentence of the language consists of a predicate, say an n -place predicate, adjoined to n variables. The rest of the sentences are built up of the atomic ones by truth functions and quantification.

Thus the only singular terms are the variables, used for quantification. It would be all right to allow also names as further singular terms and to allow functors for building complex singular terms from the names and the variables. But we can pass over these further conveniences; for there are well-known ways of dispensing with them, however inconveniently, by systematic paraphrasing of contexts.

When language is thus regimented, its ontology comprises just the objects that the variables of quantification admit as values. Some of the turns of phrase in ordinary language that seem to involve novel sorts of objects will disappear under the regimentation. Still we must not expect to end up with bodies as the only values of the variables. Much of the positing of abstract objects that seems to go on in ordinary language proves to be gratuitous and eliminable, but much of it also proves valuable. How *sets* can pay their way is classically illustrated by the definition of the *closed iterate* of a two-place predicate. Ancestor, for instance, is the closed iterate of parent. Neither parenthood nor ancestry has to do with sets, but sets enable us to define ancestor in terms of parent. For every predicate in our language we can express also its closed iterate, if we allow ourselves to quantify over sets as values of our variables.

It must be emphasized that when we reckon the ontology as comprising just the values of the variables, we are assuming the strictly regimented no-

tation: just predicates, variables, quantifiers, and truth functions. Admission of additional linguistic elements can upset this ontological standard. Thus suppose someone adopts outright an operator for forming the closed iterates of predicates, instead of defining it with help of an ontology of sets. Are we to say that he has saved on ontology? I say rather that he has shelved the ontological question by switching to a language that is not explicit on ontology. His ontology is indeterminate, except relative to some agreed translation of his notation into our regimented one.

Another way in which quantification over sets or numbers or other abstract objects can sometimes be avoided is by admitting a modal operator of necessity, if we can see our way to making appropriate sense of this device.² Here again we are presented not with an ontological saving but with a question of foreign exchange.

We have just been seeing how the values of the variables may understate the ontology in the presence of some foreign notations. Other foreign notations may work oppositely. If idioms of prepositional attitude were admitted, such as ' x believes that p ', then the variables might seem to overreach the ontology; for x can believe that $(\exists y)(y$ is a unicorn) without there being any unicorns. The ontological question for such a language, as for ordinary language generally, makes sense only relative to agreed translations into ontologically regimented notation. A language is not necessarily defective in being thus ontologically indecisive; it is just not a language of the object-oriented type.

Translation of ordinary language into the regimented idiom is not determinate. For some sentences there are various acceptable regimentations not equivalent to one another in point of ontology, and for some sentences there is no acceptable regimentation at all. In general this translation venture is significant only when undertaken systematically for a substantial corpus of sentences, a branch of science, rather than for stray sentences in isolation. Many sentences that seem from their grammatical form to talk of abstract objects of various sorts will be translated into regimented sentences that are innocent of those ontic commitments, for the translator will favor ontic economy where he can. Regiment as he will, however, he cannot make do with just bodies. By quantifying over classes he increases the yield of his apparatus, as illustrated by the closed iterates. By quantify-

2. See Hilary Putnam, "Mathematics without Foundations," *Journal of Philosophy* 64 (1967): 5–22.

ing over numbers and functions he is able to make systematic use of measurement and thus to develop his scientific theory along quantitative lines.

These sets, numbers, and functions are posited, as denizens of the universe supplementary to the primordial bodies, in order to strengthen and simplify the over-all theory. To do so is not to repudiate physicalism. The physicalist does not insist on an exclusively corporeal ontology. He is content to declare bodies to be *fundamental* to nature in somewhat this sense: there is no difference *in the world* without a difference in the positions or states of bodies. I say 'in the world' so as not to include differences between abstract objects, as of mathematics.

My qualification 'in the world' may seem to deprive the statement of content, as if to say that there is no difference in the *physical* world without a difference in the positions or states of bodies. I may better phrase the matter in terms of *change*: there is no change without a change in the positions or states of bodies. This serves still to exempt mathematical objects, which are changeless.

One application of this physicalist principle is to dispositions. There is no change even in unactualized dispositions without physical change, no difference in dispositions without physical difference. But the main thrust of the doctrine, of course, is its bearing on mental life. If a man were twice in the same physical state, then, the physicalist holds, he would believe the same things both times, he would have the same thoughts, and he would have all the same unactualized dispositions to thought and action. Where positions and states of bodies do not matter, there is no fact of the matter.

It is not a reductionist doctrine of the sort sometimes imagined. It is not a utopian dream of our being able to specify all mental events in physiological or microbiological terms. It is not a claim that such correlations even exist, in general, to be discovered; the groupings of events in mentalistic terms need not stand in any systematic relation to biological groupings.³ What it does say about the life of the mind is that there is no mental difference without a physical difference. Most of us nowadays are so ready to agree to this principle that we fail to sense its magnitude. It is a way of saying that the fundamental objects are the physical objects. It accords physics its rightful place as the basic natural science without venturing any

3. See Donald Davidson, "Mental Events," ed. Lawrence Foster and J. W. Swanson, *Experience and Theory* (Amherst: University of Massachusetts Press, 1970), 79–101; "The Material Mind," ed. P. Suppes et al., *Logic, Methodology, and Philosophy of Science* 4 (Amsterdam: North-Holland Publishing Co., 1973), 709–22.

dubious hopes of reduction of other disciplines. It has further important implications that we tend not to see.

If there is no mental difference without a physical difference, then there is pointless ontological extravagance in admitting minds as entities over and above bodies; we lose nothing by applying mentalistic predicates directly to persons as bodies, much in the manner of everyday usage. We still have two species of predicates, mental and physical, but both sorts apply to bodies. Thus it is that the physicalist comes out with an ontology of just physical objects, together with the sets or other abstract objects of mathematics; no minds as additional entities.

Note that the situation is not symmetrical. The converse move of dispensing with bodies in favor of minds is not open to us, for we would not allow that there is no physical difference in the world without a mental difference—not unless we were idealists.

I have been talking easily of physical predicates, physical differences, as over against mental ones. Until this notion is better defined or delimited, my formulations of physicalism are inadequate. Thus take the dictum ‘no mental difference without a physical difference’. We must not explain ‘physical difference’ merely as any difference between bodies; this would trivialize the dictum. For, even if we were to recognize minds as entities distinct from bodies and merely associated with them, it would be trivial to say that there is no difference in states of mind without a difference in the associated bodies. The bodies differ at least to the extent of being associated with minds that are in those different states.

Thus the dictum tells us nothing until we define ‘physical difference’ more narrowly. Similarly for my preceding versions of physicalism: ‘no difference in the world without a difference in the positions or states of bodies’, ‘no change without a change in the positions or states of bodies’. We must say what to count as states of bodies.

One major motivation of physics down the centuries might be said to have been just that: to say what counts as a physical difference, a physical trait, a physical state. The question can be put more explicitly thus: what minimum catalogue of states would be sufficient to justify us in saying that there is no change without a change in positions or states?

Thus take primitive atomic theory. Atoms are posited, small analogues of the primordial bodies. Here, as in the positing of sets or other mathematical objects, one motive is simplification of the overall system of the world. But here we may recognize also the deeper motive of fixing the no-

tion of a physical difference, a physical state. According to primitive atomic theory with its uniform atoms, any physical difference is a difference in the number or arrangement or trajectories of the component atoms.

Physicalism, on these terms, would say that where there are no such atomic differences there are no differences in matters of fact—and in particular no mental differences. But it would never have held out hope of actually describing mental states or even most gross bodily states in terms of the number, arrangement, and trajectories of atoms.

Atoms have since given way to a bewildering variety of elementary particles. Latter-day physicists have been finding even that the very notion of particle is inappropriate at points; paradoxes of identification and individuation arise. There are indications that the utility of the particle model, the extrapolation of the primordial body into the very small, is now marginal at best. A field theory may be more to the point: a theory in which various states are directly ascribed in varying degrees to various regions of space-time. Thus at last bodies themselves go by the board—bodies that were the primordial posits, the paradigmatic objects most clearly and perspicuously beheld. *Sic transit gloria mundi*.

What then is the brave new ontology? There are the real numbers, needed to measure the intensity of the various states, and there are the space-time regions to which the states are ascribed. By identifying each space-time point with a quadruple of real or complex numbers according to an arbitrary system of coordinates, we can explain the space-time regions as sets of quadruples of numbers. The numbers themselves can be constructed within set theory in known ways, and indeed in pure set theory; that is, set theory with no individuals as ground elements, set theory devoid of concrete objects. The brave new ontology is, in short, the purely abstract ontology of pure set theory, pure mathematics.⁴ At first we just tolerated these abstract objects as convenient adjuncts to our central corporeal ontology because of the power and simplification that they contributed. In the end, like the camel who got his nose under the tent, they have taken over.

A lesson to be drawn from this debacle is that ontology is not what mainly matters. When bodies first came into my story I warned that they, even they, were theoretical. All theoretical entities are here strictly on suf-

4. I develop the point a little more fully in "Whither Physical Objects?," *Boston Studies in the Philosophy of Science* 39 (1976): 303–10.

ference; and all entities are theoretical. What were observational were not terms but observation sentences. Sentences, in their truth or falsity, are what run deep; ontology is by the way.

The point gains in vividness when we reflect on the multiplicity of possible interpretations of any consistent formal system. For, consider again our standard regimented notation, with a lexicon of interpreted predicates and some fixed range of values for the variables of quantification. The sentences of this language that are true remain true under countless reinterpretations of the predicates and revisions of the range of values of the variables. Indeed any range of the same size can be made to serve by a suitable reinterpretation of the predicates. If the range of values is infinite, any infinite range can be made to serve; this is the Skolem-Löwenheim theorem. The true sentences stay true under all such changes.

Perhaps then our primary concern belongs with the truth of sentences and with their truth conditions, rather than with the reference of terms. If we adopt this attitude, questions of reference and ontology become incidental. Ontological stipulations can play a role in the truth conditions of theoretical sentences, but a role that could be played as well by any number of alternative ontological stipulations. The indecisiveness of ordinary language toward questions of reference is the more readily excused.

What now of physicalism? To profess materialism, after all this, would seem grotesquely inappropriate; but physicalism, reasonably reformulated, retains its vigor and validity. Our last previous formulation came to this: there is no difference in the world without a difference in the number or arrangement or trajectories of atoms. But if we make the drastic ontological move last contemplated, all physical objects go by the board—atoms, particles, all—leaving only pure sets. The principle of physicalism must thereupon be formulated by reference not to physical objects but to physical vocabulary. Let us take stock of the vocabulary.

Our language still has the standard regimented form; there are the truth functions, the quantifiers and their variables, and a lexicon of predicates. The variables now range over the pure sets. The predicates comprise the two-place mathematical predicate '∈' of set membership and, for the rest, physical predicates. These will serve to ascribe physical states to space-time regions, each region being a set of quadruples of numbers. Presumably regions are always wanted rather than single points—sometimes because of indeterminacy at the quantum level and sometimes for more obvious reasons, as in the case of temperature or entropy. A state may be ascribed out-

right, for example leftward spin, or quantitatively, for example temperature. In the one case the form of predication is ' Fx ', combining a one-place predicate and a variable whose relevant values are sets of quadruples of numbers. In the other case the form is ' Fxy ', combining a two-place predicate and two variables. The relevant values of one of the variables are again sets of quadruples of numbers, and those of the other variable are single real numbers measuring the quantitative state. Thus this two-place predicate ' F ' might read 'the temperature in degrees Kelvin of the region . . . is. . . .' Also there may be polyadic predicates ascribing relations, absolute or quantitative, to pairs of regions, or to triples, or higher. In any event the lexicon of physical predicates will be finite, such being the way of lexica.

A nice contrast emerges, incidentally, between physical law and physical description. The laws favor no specific space-time regions as values of the variables. Thus they are independent of the parochial specificity that goes into our choice of spatio-temporal coordinates. The specificity shows itself only in more mundane pursuits such as astronomy, geography, and history, where it is welcome.

But this is by the way. What now is the claim of physicalism? Simply that there is no difference in matters of fact without a difference in the fulfillment of the physical-state predicates by space-time regions. Again this is not reductionism in any strong sense. There is no presumption that anyone be in a position to come up with the appropriate state predicates for the pertinent regions in any particular case.

This formulation, 'fulfillment of physical-state predicates by space-time regions', is decidedly unfinished. The space-time regions are sets of quadruples of numbers, determined according to some system of coordinates that I have not paused over. The physical-state predicates are the predicates of some specific lexicon, which I have only begun to imagine, and which physicists themselves are not ready to enumerate with conviction. Thus I have no choice but to leave my formulation of physicalism incomplete. I suggested before that a major purpose of physics has been to find a minimum catalogue of states—elementary states, let us call them—such that there is no change without a change in respect of them. This is true equally of physics today.

In conclusion I want to relate physicalism to my perennial criticisms of mentalistic semantics. Readers have supposed that my complaint is ontological; it is not. If in general I could make satisfactory sense of declaring two expressions to be synonymous, I would be more than pleased to recog-

nize an abstract object as their common meaning. The method is familiar: I would define the meaning of an expression as the set of its synonyms. Where the trouble lies, rather, is in the two-place predicate of synonymy itself; it is too desperately wanting in clarity and perspicuity.

Translation proceeds, presumably, by interlinguistic equivalence of synonymy of sentences. So, in order to make the problem of synonymy graphic, I developed a thought experiment in radical translation—that is, in the translation of an initially unknown language on the strength of behavioral data.⁵ I argued that the translations would be indeterminate, in the case of sentences at any considerable remove from observation sentences. They would be indeterminate in this sense: two translators might develop independent manuals of translation, both of them compatible with all speech behavior and all dispositions to speech behavior, and yet one manual would offer translations that the other translator would reject. My position was that either manual could be useful, but as to which was right and which wrong there was no fact of the matter.

My present purpose is not to defend this doctrine. My purpose is simply to make clear that I speak as a physicalist in saying there is no fact of the matter. I mean that both manuals are compatible with the fulfillment of just the same elementary physical states by space-time regions.

Radical translation proceeds in the light of observed behavior, and behavioral criteria will ordinarily decide in favor of one translation rather than another. When they do, there is emphatically a fact of the matter by microphysical standards; for clearly any difference in overt behavior, vocal or otherwise, reflects extravagant differences in the distribution of elementary physical states. On the other hand my doctrine of indeterminacy had to do with hypothetical manuals of translation both of which fitted all behavior. Since translators do not supplement their behavioral criteria with neurological criteria, much less with telepathy, what excuse could there be for supposing that the one manual conformed to any distribution of elementary physical states better than the other manual? What excuse, in short, for supposing there to be a fact of the matter?

We have here an illustration of what I consider the proper function of behaviorism. Mental states and events do not reduce to behavior, nor are they explained by behavior. They are explained by neurology, when they are explained. But their behavioral adjuncts serve to specify them objec-

5. In *Word and Object* (Cambridge, MA: MIT Press, 1960), chap. 2.

tively. When we talk of mental states or events subject to behavioral criteria, we can rest assured that we are not just bandying words; there is a physical fact of the matter, a fact ultimately of elementary physical states.

We learn mentalistic idioms, like other idioms, from elder speakers of our language, in distinctive and intersubjectively observable circumstances. Those circumstances differ from others in respect of the distribution, however inscrutable, of elementary physical states. As long as we use such an idiom in a form and in circumstances closely similar to the original ones, we communicate information; there is a fact of the matter. But our mentalistic idioms, like other idioms, go on growing and stretching by analogy. Factual content becomes meanwhile more tenuous and more elusive and can disappear altogether.

Thus consider the prepositional attitudes; consider belief. There are unproblematical attributions of belief—problematical attributions even to dumb animals. Observation of behavior would normally prompt us to agree that the dog believes his master is coming, or that he believes the ball is under the sofa. When we attribute a belief about ancient history to someone, on the other hand, we are dependent on what he says—even though we are loath to equate belief with lip service. If the believer is a foreigner, our attribution may be subject also to the vagaries of translation of his testimony into our language. In some cases factual content is lacking; in others it is sparse and ill defined.

I do not advise giving up ordinary language, not even mentalistic language. But I urge awareness of its pitfalls. There is an instructive parallel between questions of reference, on the part of ordinary language, and questions of factuality. Let me recall what I said earlier when discussing ontology. Ordinary language is only loosely referential, and any ontological accounting makes sense only relative to an appropriate regimentation of language. The regimentation is not a matter of eliciting some latent but determinate ontological content of ordinary language. It is a matter rather of freely creating an ontology-oriented language that can supplant ordinary language in serving some particular purposes that one has in mind.

Now factuality is similar. Ordinary language is only loosely factual, and needs to be variously regimented when our purpose is scientific understanding. The regimentation is again not a matter of eliciting a latent content. It again is a free creation. We withdraw to a language which, though not limited to the assigning of elementary physical states to regions, is visibly directed to factual distinctions—distinctions that are unquestion-

ably underlain by differences, however inscrutable, in elementary physical states. This demand is apt to be met by stressing the behavioral and the physiological.

Within these limits there is still much scope, of course, for better and worse. The terms that play a leading role in a good conceptual apparatus are terms that promise to play a leading role in causal explanation; and causal explanation is polarized. Causal explanations of psychology are to be sought in physiology, of physiology in biology, of biology in chemistry, and of chemistry in physics—in the elementary physical states.

Cognitive Meaning

1979

Words and phrases refer to things in either of two ways. A name or singular description *designates* its object, if any. A predicate *denotes* each of the objects of which it is true. Such are the two sorts of reference: designation and denotation. We are often told, and rightly, that neither sort is to be confused with *meaning*. The descriptions ‘the author of *Waverley*’ and ‘the author of *Ivanhoe*’ designate the same man, after all, but differ in meaning; and a predicate may denote each of many things, while having only one meaning.

There are no restraints on what can be referred to. All sorts of things are designated and denoted. On the other hand a meaning is apparently some special sort of thing. But just what?

You would think we would know. The word ‘meaning’ is a common noun, a very common noun, on the tip of everyone’s tongue. It occurs in a few frequent phrases. We ask the meaning of a word; we give the meaning of a word; we speak of knowing the meaning of an expression; we speak of an expression as having or lacking meaning; and we speak of expressions as alike in meaning. But one context in which we do not normally encounter the word ‘meaning’ is the context ‘a meaning is’; ‘a meaning *is* such and such’. The question ‘What is a meaning?’ thus qualifies as a peculiarly philosophical question.

Meanings are meanings of expressions, so I had better begin by explaining my use of the expression ‘expression’. An expression, for me, is a string of phonemes—or, if we prefer to think in terms of writing, a string of letters and spaces. Some expressions are sentences. Some are words. Thus when I speak of a sentence, or of a word, I am again referring to the sheer string of phonemes and nothing more. I must stress this because there is a

widespread usage to the contrary. The word or sentence is often thought of rather as a combination, somehow, of a string of phonemes and a meaning. Homonyms are thus treated as distinct words. This usage is often convenient in the study of language, and in its proper place I have no quarrel with it. But it cannot be allowed here, because our purpose is to isolate and clarify the notion of meaning.

A meaning, still, is something that an expression, a string of phonemes, may *have*, as something external to it in the way in which a man may have an uncle or a bank account. It has it by virtue of how the string of phonemes is used by people. The same expression, the same string, may by coincidence turn up in two languages and mean differently in them. It may mean in two ways even in a single language; and then it is called ambiguous. But it remains, in my sense, one and the same expression. On this approach ambiguity and homonymy are on a par: what is concerned in either case is a single string of phonemes.

The point is that the notion of an expression must not be allowed to presuppose the notion of meaning. One may suspect, however, that my identifying expressions with strings of phonemes is then self-defeating, because the notion of a phoneme is itself commonly so defined as to presuppose the notion of sameness and difference of meaning. Two acoustically distinguishable sounds are counted as occurrences of the same phoneme if substitution of the one sound for the other leaves meanings unaffected. Now this appeal to meaning is happily not needed. We can simply say that two sounds count as occurrences of the same phoneme if the substitution has no effect on a speaker's readiness to assent to any sentence.

Expressions, then, are strings of phonemes in this innocent sense, and it is expressions that are to have meanings. What sort of things these meanings are is the question before us. But actually we are rather rushing matters in supposing there to be such things as meanings; for one can perhaps talk of meaning without talking of meanings.

Thus let us start rather with the verb 'mean' as an intransitive verb. An expression means; meaning is what it does, or what some of them do. To say that two expressions are alike in meaning, then, is to say that they mean alike. Some expressions sound alike, some mean alike. It is significant that when we ask for the meaning of an expression we are content to be given another expression on a par with the first—like it in meaning. We do not ask for something that the two of them mean. The French idiom is more to the point: *cela veut dire*.

We could wish for further light on just what an expression is doing when it means, and how it does it, but we need not seek anything that the expression is doing it to—anything that gets meant. We are not tempted to if we take ‘mean’ as intransitive.

In a longer view, however, this bid for ontological economy is idle. For, once we understand what it is for expressions to mean alike, it is easy and convenient to invoke some special objects arbitrarily and *let* them be meant—thus reconstituting our verb as transitive. In choosing a domain of objects for this purpose and assigning them to expressions as their so-called meanings, all that matters is that the same one be assigned always and only to expressions that mean alike. If we can manage this, then we can blithely say thereafter that expressions that mean alike *have* the *same meaning*. We should merely bear in mind that ‘mean alike’ comes first, and the so-called meanings are then concocted.

How do we concoct them? I said it was easy. Just take each meaning as a set of expressions that mean alike. The meaning of an expression is the set of all expressions that mean like it. Clearly this definition, which is by no means new, meets the stated requirement; it assigns the same meaning to two expressions if and only if they mean alike. The air of artificiality is no drawback, since there was no preconception as to what sort of thing a meaning ought to be.

So we see that if we know what it is for expressions to mean alike, the rest is easy. And we would seem to know, if facile lip service were indicative; but it is not. Fluency and clarity are poorly correlated. Philosophers have the problem of taking up the slack.

Likeness of meaning, then, is the relation that needs to be clarified. We cannot clarify it by saying that the expressions so related call up the same mental images or ideas. What images or ideas are evoked in a given individual by a given expression is a vague and rather inscrutable matter at best, and apt, moreover, to vary from occasion to occasion. Something to do with social transaction would be more to the point, since language is a social art. We learn language from other people and we use it to gain information from other people and to influence the behavior of other people. Likeness of meaning should consist somehow in sameness of the stimulations that evoke the expressions, or sameness in the behavior elicited by the expressions.

But there are troubles here too. No two expressions are evoked by quite the same stimulations, since one of them can be evoked only to the exclu-

sion of the other. What favors the one over the other can be some trivial accident of association, but then we need a criterion for screening out the trivialities. Anyway, utterances cannot usually be accounted for by concurrent external stimulation: too much depends on antecedent stimulation, antecedent lines of thought, and varying motives. Causes of an utterance lie deeper than the concurrent external stimuli. There would be similar difficulties in identifying the effects of an utterance.

We might make matters more manageable by treating expressions as alike in meaning only in limited respects. If the expressions concerned are the sentences of some science, the scientist will be more concerned with what makes them true or false than with the motives of vanity or helpfulness or deception that prompt someone to utter them, or with the surprise or disbelief or frenzied activity that their utterance might induce in someone else. He will be interested, in short, in likeness of cognitive meaning. Sentences are alike in cognitive meaning—*cognitively equivalent*, let us say—if, vaguely speaking, they have the same truth conditions. This is the relation that one would like to render precise and explicit. I shall not succeed in doing so, but I shall somewhat clarify the nature of the problem.

There is an important relation that is a sufficient but not necessary condition of cognitive equivalence. It is one way in which sentences can be alike in cognitive meaning. The relation is exemplified by the pair of sentences 'Brutus killed Caesar' and 'Caesar was killed by Brutus'. I shall call it *equivalence by grammatical structure*. It does not mean that they are equivalent *in* grammatical structure; they differ structurally, but are semantically equivalent *by virtue* of their two structures. The relation is pretty clear from the example, but more needs to be said about it.

To discern the grammatical structure of a sentence, we must distinguish between those words or morphemes that are mere grammatical particles and those that belong to the lexicon; for the grammatical structure of the sentence is what remains of it when we put blanks for the lexical elements and keep just the grammatical particles. The grammatical structure of the sentence 'Brutus killed Caesar' may be depicted as 'B A'd C', where the capital letters serve as blanks for lexical elements and the suffix '-d' is the lone grammatical particle. Again the grammatical structure of the sentence 'Caesar was killed by Brutus' may be depicted as 'C was A'd by B'; here we recognize three grammatical particles, 'was', '-d', and 'by'.

How can we distinguish between grammatical particles and full-fledged lexical elements? If we make a lot of inessential revisions of our ordinary

spoken language, it becomes possible to distinguish between the particles and the lexicon by a neat quantitative criterion. It is characteristic of a grammatical particle that few if any other expressions can be substituted for it in all sentences without sometimes producing ungrammatical nonsense. For a lexical element, on the other hand, there are countless words and phrases that can be substituted without ever violating grammar. However, this contrast between particles and lexicon proves satisfactory only if sundry common turns of phrase are cleared away. Such idioms as 'lay claim to', or 'take offense at', would force the nouns 'claim' and 'offense' into the unintended status of grammatical particles; for there are few expressions that could supplant these words in those idioms without violating grammar.

Taking English as it stands, then, we can perhaps do no better than to leave the distinction between particle and lexicon to the grammarian's caprice or convenience. The grammarian lists a limited assortment of highly productive grammatical constructions, marked by special short words or affixes; and these markers are the grammatical particles. No one codification of grammar, however, is uniquely determined by our speech dispositions; codifications can even agree in their output and yet differ materially in internal ways. They may divide up and sort out the grammatical constructions along different lines, issuing in the same net results. A word may be reckoned as a particle marking a grammatical construction under one codification and not under another. In speaking of a distinction between particles and lexicon, then, we may treat it as relative to one or another particular codification of grammar. Similarly, then, when we speak of the grammatical structure of sentences, we may acknowledge the particular codification of grammar as a tacit parameter.

Let us take our bearings. I am on the way to explaining a relation called equivalence by grammatical structure. We depict the grammatical structure of a sentence by putting blanks or schematic letters in place of all the lexical elements, leaving only the grammatical particles. Conversely, we get one or another sentence from such a grammatical scheme by substituting lexicon for the various schematic letters. If we start with two such grammatical schemata and make the same substitutions in each, the two resulting sentences may be called *corresponding instances* of the two grammatical structures in question. Thus 'Brutus killed Caesar' and 'Caesar was killed by Brutus' are corresponding instances of 'B A'd C' and 'C was A'd by B'. Two grammatical structures may be called *equivalent* if, as in this example,

their corresponding instances can always be depended upon to agree with each other in truth value. Finally, sentences are *equivalent by grammatical structure* when they are corresponding instances of equivalent grammatical structures.

Logical equivalence comes under this head. The logical schemata ‘Either p and q or p and r ’ and ‘ p and either q or r ’ depict equivalent grammatical structures, and their corresponding instances are logically equivalent sentences. Many of us would likewise call ‘Brutus killed Caesar’ and ‘Caesar was killed by Brutus’ logically equivalent, though this passive construction is not part of standard logic. It would be pleasant and not unreasonable to drop the clumsy term ‘equivalent by grammatical structure’ and simply call any such sentences logically equivalent, in an extended sense. But I forbear, so as not to raise an issue extraneous to present concerns.

Equivalence by grammatical structure is a special case of cognitive equivalence. We could easily proceed to the general case, defining what it is in general for two sentences to be cognitively equivalent, if we had at our disposal a certain relation of *cognitive synonymy* applicable to simple terms. We could then say that two sentences are cognitively equivalent if and only if one can be converted into the other by a sequence of steps each of which either carries a sentence into an equivalent by grammatical structure or consists in supplanting a component term of a sentence by a cognitive synonym.

I have not defined cognitive synonymy of terms, and before I am through I shall despair of doing so. But meanwhile I shall sketch some intended traits of that relation and illustrate how one would want it to work.

In saying that it should apply to simple terms I meant that it should relate simple terms—single morphemes—to simple *or* complex terms. Thus, presumably we would want to count the simple singular term ‘Hesperus’ cognitively synonymous to the complex term ‘the Evening Star’, and the simple general term ‘bachelor’ to the complex term ‘unmarried man’. I require at least one of the two synonymous terms to be simple merely in order to skip the case where two complex terms are synonymous by grammatical structure, since I already defined grammar’s contribution to the cognitive equivalence of sentences; what is still needed, to complete the definition of cognitive equivalence of sentences, was what synonymy contributes apart from structure.

In treating of synonymy of general terms we do better to think not of nouns and noun phrases, e.g., ‘bachelor’ and ‘unmarried man’ above, but

rather of *predicates* in the sense of my *Methods of Logic*; that is, as sentence forms with numbered blanks. We are to view the term 'bachelor' as having as its basic form the predicate '① is a bachelor', as if this were a single morpheme. The term 'cousin' is basically the predicate '① is cousin of ②'. Predication of such predicates consists in putting the logical subject or subjects in place of the circled numerals: 'John is a bachelor', 'John is cousin of Maud'. We need the flexibility of predicates, with their numbered places, in order to provide, for the cognitive equivalence of such sentences as 'John is cousin of Maud' and 'Maud is cousin of John'. These qualify as cognitively equivalent because the one is convertible into the other by supplanting the simple predicate '① is cousin of ②' by a cognitive synonym, '② is cousin of ①'. What I have called the supplanting of a term by a cognitive synonym is not, thus, a mere blocklike substitution in the case of predicates. It is a matter rather of lifting the logical subjects, restoring the circled numerals, substituting the new predicate, and then putting the subjects back according to the new circled numerals.

This method of predicates requires a certain regimenting of language, such as to maneuver all general terms into predicative position. Thus 'A cousin of Maud's is a bachelor', or 'Maud has a cousin who is a bachelor', should be paraphrased as 'There is someone who is a cousin of Maud and who is a bachelor'. Thus paraphrased, the sentence exhibits direct predication of the predicates '① is cousin of ②' and '① is a bachelor'. The alternative course would be to admit all the ordinary idioms and then devise a very elaborate description of what is to count as replacement of a general term by a cognitive synonym.

If we talk thus of regimenting our language by paraphrasing our sentences, however, it may be protested that we are thereby presupposing the very relation, cognitive equivalence, that we wanted to define. The sentences are cognitively equivalent to their paraphrases. Let us then not talk of paraphrase at this point. Let us consider rather that we are merely *limiting* our study to a convenient *part* of English, the part where sentences exhibit direct predication of their predicates.

Let us now again pick up the thread. We are assuming an as yet undefined relation of terms, called cognitive synonymy, and we are seeing how the cognitive equivalence of sentences could be defined on that basis. Sentences are cognitively equivalent if one goes into the other by a sequence of steps of two kinds. The first kind is equivalence by grammatical structure. The second kind is substitution of cognitively synonymous terms.

The substitutions are to be limited, of course, to proper occurrences. We are not to recognize the occurrence of 'table' in 'stable', nor of 'Maud' in 'maudlin'. The notion of a proper occurrence, like that of grammatical structure, should perhaps be viewed as dependent on some codification of grammar, as tacit parameter; for a proper occurrence of an expression is its occurrence in a grammatical construction, either as a component of the construction or as a particle marking the construction.

But besides excluding improper occurrences from consideration, we must at this point also exclude occurrences in direct quotation. This is evident from the following example. We want to regard 'Hesperus' as synonymous with 'the Evening Star', even though not with 'the Morning Star'; also we want to regard 'glittereth' as synonymous with 'glitters'; but either of these substitutions would turn the true sentence:

'Hesperus glittereth' is dactylic

into a falsehood, and hence certainly not into a cognitive equivalent.

In substituting terms for their cognitive synonyms in order to turn sentences into their cognitive equivalents, then, we must substitute only for proper occurrences outside quotation. This sealing off of quotation could indeed be reduced to the mere requirement of proper occurrence. We could agree that quotation is not a grammatical construction and hence that an occurrence within quotation is not a proper occurrence relative to the outer context, being separated from it by a grammatical discontinuity. There is little point, however, for there is further trouble ahead. Consider this:

(1) Tom does not know that Hesperus is the Evening Star.

It may well be true. And certainly no cognitive equivalent is obtained by substituting 'the Evening Star' for its synonym 'Hesperus'; we get rather the palpable falsehood:

(2) Tom does not know that the Evening Star is the Evening Star.

The occurrence of 'Hesperus' is thus no more eligible for substitution in (1) than it would be in quotation. And certainly this occurrence cannot be accounted improper, in the sense of being grammatically unconnected with the outer context.

So our restraint on substitution of synonyms must be extended beyond improper occurrences and quoted occurrences, to cover also occurrences

like that of 'Hesperus' in (1). Now how are such occurrences to be distinguished from others?

One recalls a distinction between the use or occurrence of a term *de re* and its occurrence *de dicto*. Where a term occurs *de dicto*, the sentence containing it depends for its truth not just on the traits of the designated object or denoted objects, but on further traits of the term itself. Where a term occurs *de re*, the sentence depends for its truth only on the objects referred to and not on what they were called. Commonly ordinary usage is ambiguous on the point. Thus, for someone who calls the truths of arithmetic necessary and those of astronomy contingent, the sentence:

The number of planets necessarily exceeds 5

counts as true when the occurrence of 'the number of planets' is construed *de re* and false when that occurrence is construed *de dicto*.

Now the occurrence of 'Hesperus' in (1) is obviously *de dicto*, and we saw that it resists substitution of a synonym. We would be wrong, however, to try to safeguard our definition of cognitive equivalence of sentences by excluding substitution for *de dicto* occurrences. For, consider the example:

(3) The Sumerians did not know that Hesperus is the Morning Star.

The occurrence of 'Hesperus' here, as in (1), is *de dicto*; for though Hesperus is the Morning Star, substitution of 'the Morning Star' for 'Hesperus' turns (3) into an absurdity as utter as (2). Yet we do want to allow substitution of synonyms for that occurrence; we do want (3) to qualify as cognitively equivalent to;

(4) The Sumerians did not know that the Evening Star is the Morning Star,

but we do not want (1) to qualify as equivalent to (2). Intuitively, the difference between (1) and (3) is that Tom is ignorant about a word and the Sumerians were ignorant about the stars. But both are *de dicto*.

Clearly then the occurrences of 'Hesperus' are *de dicto* in (1) and (3) in different ways. The occurrence in (3) might be called, more particularly, *de sensu*: the truth of (3) depends only on the meaning of 'Hesperus' and is not disturbed by substitution of a synonym, 'the Evening Star'. The occurrence in (1) is more sensitive, and may be called *de voce*: the truth of (1) depends not just on the reference of 'Hesperus' nor even on its meaning, but on the specific choice of word. Occurrence *de voce* is thus comparable

to occurrence in quotation, which might, indeed, be included under that head. (1) could itself be paraphrased into the quotational idiom without loss, thus:

Tom does not know that 'Hesperus' designates the Evening Star.

The corresponding paraphrase of our *de sensu* example, (3), would be wrong; it would be trivially true, because merely of the Sumerians' ignorance of our Latinized Greek vocabulary, and would thus lack the force of (3).

Very well: we have drawn the line on substitution for *de voce* occurrences. The definition of cognitive equivalence now runs as follows: the one sentence can be converted into the other by a series of steps each of which either carries a sentence into an equivalent by grammatical structure or consists in supplanting a component term of a sentence by an occurrence *de sensu* or *de re* of a cognitive synonym.

It will be recalled that we are leaving the cognitive synonymy of terms undefined for the present, and are defining the cognitive equivalence of sentences with help of that notion. But now we have presupposed also a second undefined notion: that of *de sensu* occurrence. Actually the two presuppositions—cognitive synonymy and *de sensu* occurrence—are reducible to one, either one; each is definable in terms of the other, as follows. Terms are cognitively synonymous if the one can supplant any *de sensu* occurrence of the other without disturbing the truth value of the context. Conversely, an occurrence of a term is definable as *de sensu* if it is not *de re* and if, still, it is replaceable by any cognitive synonym without disturbing the truth value of the context.

When we note this last definition, and also the earlier definition of equivalence by grammatical structure, we find that the definition of cognitive equivalence of sentences harbors a certain symmetry. The two kinds of steps that lead from a sentence to a cognitive equivalent are symmetrical in this way; the one is a change of structure and the other a change of a term, and both are changes that preserve truth value wherever applied. The one kind of change required, for its definition, the prior adoption of some one codification of grammar. The other required a prior notion of cognitive synonymy of terms. This latter notion, of course, is the aching void.

This synonymy relation was limited, for economy, to pairs of terms one of which is simple. Even so, the relation is not obviously more elementary or more acceptable than the one I have been using it to define, namely cog-

nitive equivalence of sentences. Might we just as well proceed oppositely, assuming the relation of cognitive equivalence of sentences and deriving that of cognitive synonymy of terms? It seems not. Starting with cognitive equivalence of sentences, we can indeed define cognitive synonymy of terms thus: terms are cognitively synonymous if and only if putting the one for a *de sensu* occurrence of the other in a sentence always yields a cognitively equivalent sentence. However, this definition uses also the notion of *de sensu* occurrence, and in fact it gains nothing from the use of the notion of cognitive equivalence; for we already saw cognitive synonymy defined in terms merely of *de sensu* occurrence and preservation of truth values.

The upshot is that cognitive synonymy and *de sensu* occurrence are interdefinable, and cognitive equivalence is definable on the basis of either, but evidently not conversely.

Thus if we could satisfactorily define *de sensu* occurrence, all would be in order: cognitive synonymy and cognitive equivalence as well. As things already stand the distinction between *de sensu* and *de re* offers no difficulty, granted the concepts of designation, denotation, and truth. A position is *de re* if changes of terms in that position are always indifferent to the truth value of the context so long as the terms are not changed in respect of their reference. The problem of definition lies rather in the distinction between *de sensu* and *de voce*. It is the difference between the examples (1) and (3), and those examples offer nothing to get hold of. Their only contrast lies in our reckoning ‘Hesperus’ and ‘the Evening Star’ alike in meaning, and unlike ‘the Morning Star’ in meaning—whatever *that* means.

The residual myth of meaning is this myth that there is a certain critical degree of affinity of terms. Synonymy is that degree of affinity. *De sensu* occurrences are occurrences where that degree of affinity suffices for interchangeability. When we try to find a criterion for that cut-off point, we sense the quality of myth that invests it.

If we give up the thought of such a boundary, we give up the thought of sameness of meaning. In so doing we give up the thought of anything that might be called a meaning; for there is no entity without identity.

These renunciations do not require us to forswear some gradualistic notion of synonymy—synonymy as a matter of degree. There is a synonymy relation of this modest sort that can be defined without regard to distinctive sorts of occurrences, *de sensu* or otherwise. We can base it simply on substitution *salva veritate*. We can say that one term is cognitively *more*

synonymous to a second term than to a third if the one term can supplant the second term *salva veritate* in all places where it can supplant the third *salva veritate*, and in some further places besides. But this relation is only a partial ordering; commonly one term will be neither more nor less synonymous to a second than to a third, nor equally synonymous either.

This partial ordering of pairs of terms, in respect of synonymy, can be made to induce also a partial ordering of pairs of sentences, in respect of equivalence; for the sentences can be got one from another by substitution of the terms. However, this equivalence relation of sentences is too fragmentary to seem worth pausing over.

At this point any prospect of a coherent theory of meaning has vanished, and the relation of partial ordering in respect of synonymy that remains to us is a pretty meager thing. But then what are we to say of all the strenuous and useful work of lexicographers? Dictionaries are reputedly occupied with explaining the meanings of words, and the work is neither myth-bound nor capricious. How does it proceed? I hold that it is not directed at cognitive equivalence of sentences, nor at synonymy of terms, and that it presupposes no notion of meaning at all. Let us consider then what the business of dictionaries really is.

Sometimes a word is explained by supplying another expression that can replace it *salva veritate* at least in *de re* positions, if not also in *de sensu* positions (whatever those may be). Sometimes, instead, a brief selection of information is set down regarding the object or objects to which the word refers. There is no pretense here of a distinction between essential and accidental traits. It is a matter purely of pedagogy: the lexicographer wants to improve his reader's chances of successful communication as best he can in a small compass. Often, moreover, a dictionary entry neither paraphrases the word nor describes its objects, but describes, rather, the use of the word in sentences. This is usually the way with grammatical particles, and it is often the way also with terms. It is bound to be the way with a term that neither refers to concrete objects nor admits of a separable, self-contained paraphrase.

Behind this seeming disorder there is a unifying principle. The goal may be seen always as the sentence. The lexicographer is out to help his reader profit by the sentences that he sees or hears, and to help him react to them in expected ways, and to help him emit them usefully. But sentences are unlimited in their variety, while words are not. So the lexicographer organizes his teaching of sentences word by word, teaching how to use each

word in making sentences. One way of teaching this, which is convenient when available, is by citing a substitute expression; for the lexicographer thus exploits the reader's presumed knowledge of how to use that substitute expression in making sentences. And the other sorts of dictionary entry likewise aim, in their different ways, at the same end: teaching the use of sentences.

Looking at semantics through the lexicographer's eyes thus brings about an abrupt change of focus. Our focus had been likeness of meaning. We did notice an assortment of other phrases in which the problematic word 'meaning' likewise commonly occurs: we ask the meaning, we know the meaning, an expression has or lacks meaning. But we focused on likeness of meaning. It is the natural focus for a philosopher who begins by being concerned over the ontology of meaning. I did urge that 'means' be viewed as an intransitive verb, so as not to assume meanings as entities at the outset. Likeness of meaning, however, still held the focus. As long as likeness of meaning is our concern, we are in effect studying meanings as objects after all; for, as I remarked, the meaning of an expression can simply be identified with the class of its synonyms.

What divides lexicography from semantics as pursued by philosophers is thus not the reification of meanings. It is rather the shift of focus from likeness of meaning to knowledge of meaning, so to speak; from synonymy of expressions, anyway, to the understanding of expressions. The lexicographer's job is to inculcate understanding of expressions; that is, to teach how to use them. He can be wholly successful in teaching the use of sentences without considering in what sense they might be said to be equivalent. Nothing, apparently, could be more remote than meanings from the lexicographer's concern. Why should they be less remote from ours?

So we might try looking to the understanding of expressions, rather than to synonymy, as the operationally basic notion of semantics. What sense can we make of it? In practice we credit someone with understanding a sentence if we are not surprised by the circumstances of his uttering it or by his reaction to hearing it—provided further that his reaction is not one of visible bewilderment. We suspect that he does not understand it if the event is drastically at variance with those conditions. But in the case of a single isolated sentence no boundary is evident, no general criterion for deciding whether he actually misunderstands the sentence or merely holds some unusual theory regarding its subject matter.

We can be more confident in imputing misunderstanding of a word

than of a sentence, for we can then observe someone's use of numerous sentences, or his response to them, all of which contain the word. We can control our experiment, choosing and querying sentences ourselves. We may find that he responds otherwise than would generally be expected when the sentences contain the word in question, and that he responds in the more usual ways to many sentences that lack that word but are much the same in other respects.

In this matter of understanding language there is thus a subtle interplay between word and sentence. In one way the sentence is fundamental: understanding a word consists in knowing how to use it in sentences and how to react to such sentences. Yet if we would test someone's understanding of a sentence, we do best to focus on a word, ringing changes on its sentential contexts. Once we have thus satisfied ourselves, through a multiplicity of such sentences, that he misunderstands the word, we are justified at last in concluding that his odd response to the original sentence containing it was due to a misunderstanding of the word and not to some odd opinion regarding matters of fact.

Understanding, behaviorally viewed, is thus a statistical effect: it resides in multiplicities. The nucleus is the word, and the mass is made up of the countless sentences in which the word occurs. A predominantly healthy or unhealthy coloring of this mass is what counts as understanding or misunderstanding of the word and the sentences; and a sharp boundary need not be sought. A forest presents a sharp boundary to the airborne observer but not to the man on the ground; an ink blot presents a sharp outline to the unaided eye and not under a magnifying glass; and we can acquiesce in a similar attitude toward the distinction between understanding and misunderstanding an expression. Lexicography has no need of synonymy, we saw, and it has no need of a sharp distinction between understanding and misunderstanding either. The lexicographer's job is to improve his reader's understanding of expressions, but he can get on with that without drawing a boundary. He does what he can, within a limited compass, to adjust the reader's verbal behavior to that of the community as a whole, or of some preferred quarter of it. The adjustment is a matter of degree, and a vague one: a matter of fluency and effectiveness of dialogue.

He is concerned to explain the less familiar word by means of more familiar ones. The relative familiarity of words varies with the individual speaker, but it can be measured over society by relative frequency. Ideally,

then, a dictionary takes on a hierarchical structure: words are explained by means of ever more frequent words, and words of very high frequency stand unexplained.

Actual dictionaries depart from this hierarchical plan in several ways. The usual dictionary does not simply pass over the high-frequency words, though from a practical point of view it might as well. Also in the lower frequencies the actual dictionary departs often from the strict frequency hierarchy. It can be useful to do so, because relative familiarity for individual speakers does not strictly follow the social frequencies. Circularity of definition can thus occur, but it is harmless as long as the words concerned are also accorded other explanations that adhere to the hierarchical plan.

Furthermore there is an urgent reason for supplementing the hierarchical plan when we come to some of the theoretical terms of physics and chemistry, notably the terms for hypothetical particles. In the main these are irreducible: that is to say, sentences that contain them cannot be serviceably paraphrased into sentences that do not. In this respect the theoretical terms in question are like the highly frequent words. Yet they are not highly frequent; they need explaining. So the lexicographer provides useful information about them, inadequate though it is for purposes of full paraphrase.

Thus the hierarchical plan should be viewed rather as only the essential core of the actual practical dictionary. But it does much to illuminate the nature of the enterprise. Let me sum it up.

What are taken up, one by one, are the words. What is explained about them, however, is their use in sentences. In general what is wanted by way of explanation of a word is a set of instructions which, in its finite compass, explains the infinitely many sentential contexts of the word. Ideally these explanations fit into the frequency hierarchy in the following way: they explain how each sentence containing the word in question might be paraphrased into a sentence containing only words of the original sentence other than that one, plus words more frequent than that one.

What is central is thus the paraphrasing of sentences, and we seem thus to be brought back to the problem of cognitive equivalence of sentences. However, we have noted the obstacles in that path, and we are now recognizing the more vaguely educational, the more broadly statistical character of the lexicographer's enterprise. The relation between a sentence and its paraphrase, suitable for lexicographic purposes, is nothing deserving of the

name of sameness of meaning. It is nothing more substantial than the lexicographer's shrewd guess that the one sentence might serve a speaker's likely purposes about as well as the other in most of the situations that are apt to arise.

I am indebted to Burton Dreben and David Lewis for helpful comments.

Grammar, Truth, and Logic

1980

It is a general practice, in intellectual pursuits, to argue from the truth of one sentence to the truth of another. Some such arguments are the business of logic, others not. They belong to logic if they hinge purely on the structure of the sentences concerned, rather than depending on content. But the structure of sentences consists in grammatical constructions. Here, then, is the intimate connection between grammar, truth, and logic. Logic studies the truth conditions that hinge solely on grammatical constructions.

By this account, to say that one sentence logically implies another is to say that the two sentences are so related, in respect of grammatical form, that no two sentences so related are respectively true and false. The definition becomes more manageable if we break it down into a sequence of three definitions. First we define a grammatical form as *logically valid* if all sentences of that form are true. Next we define a sentence as *logically true* if it has a logically valid grammatical form. Finally we say that one sentence *logically implies* another if the conditional sentence, formed of these sentences in that order by applying ‘if’ to the one and ‘then’ to the other, is logically true.

Logical implication is the central business of logic. Logical truth would be of little concern to us on its own account, but it is important as an avenue to implication. It is simpler to theorize about truth than implication because it is attributable to single sentences whereas implication relates sentences in pairs.

I have said that a logically true sentence is a sentence having *a* logically valid grammatical form. For it will commonly have several other grammatical forms as well. The sentence ‘If it rained then it rained or it snowed’

is logically true by virtue of being of the logically valid form 'If p then p or q ', but it is also of the grammatical forms 'If p then r or q ' and 'If p then s ', neither of which is logically valid.

But we have still to say what counts as a grammatical form. The sentence:

- (1) If sugar is sweet then something is sweet

is one that we should like to call logically true, on the grounds that the form:

- (2) If x is F then something is F

is logically valid. But is this a *grammatical* form? Certainly 'If x is F then salt is F ' would not be a purely grammatical form; the word 'salt' does not qualify as a mere grammatical particle. Why does the word 'something' so qualify? How can we tell what morphemes to count as full-fledged lexical elements and what ones to regard as mere particles marking grammatical constructions?

There is a quite natural criterion of a quantitative kind. Thus consider the particle 'and'. What other morphemes could be substituted for 'and' in all sentences without producing ungrammatical strings? Very few: 'or', 'but', and little else. Or again consider the suffix '-ed'; surely there is no other that could be substituted for it in sentences without usually engendering nonsense. Such is the cramped nature of grammatical particles. For a lexical element, on the other hand, there are no end of substitutes that preserve grammaticality. If for almost any common noun in a sentence you substitute almost any other common noun or noun phrase, the result may be false and may be bizarre, but it will be grammatical. Similar fluidity prevails in the various sorts of verbs and adjectives and in the singular terms. The morphemes in all these categories belong to the lexicon. Such, then, is the quantitative criterion: a morpheme is a particle or a lexical element according as there are few or many expressions in its grammatical category.

The word 'something' does seem more substantial than a grammatical particle. It is substantial and even substantival; it occurs as subject and object of verbs. By our quantitative criterion, however, it is still a grammatical particle. It does not belong to a large grammatical category of substantives. 'Something' can be supplanted by 'salt' in some sentences without producing ungrammatical results, but there are other sentences in which such substitution produces incoherence—such a one as 'There was something

new on the Rialto'. Substitution gives 'There was salt new on the Rialto', and this must presumably be disqualified, not just as an absurd sentiment, but as flatly ungrammatical. Such cases suffice to dissociate 'something' from 'salt' in respect of grammatical category.

The word 'something' actually has little company in its grammatical category. It has the word 'nothing' for company, perhaps, and a few others. Thus it is that 'something' qualifies as a grammatical particle by our criterion, and accordingly (2) qualifies as a grammatical form by our criterion, and so (1) qualifies as logically true by our criterion.

Actually this example can still be contested in another way. It may be objected that 'something' is not a morpheme but a noun phrase consisting of two morphemes. This is reasonable, but it does not affect our conclusion; for the morpheme 'some' still qualifies as a logical particle, and so does 'thing'. This is established still by the example 'There was something new on the Rialto'; neither 'some' nor 'thing' admits of many substitutes in this context without incoherence. So (2) still qualifies as a logical form, and (1) as logically true.

I suppose the first example one thinks of in the way of a logical particle is 'not'; so we would surely want this to turn out to be a grammatical particle by our criterion. We may start to worry, then, when we reflect that 'not' can be supplanted in most contexts by countless adverbs without incoherence. However, the situation is again saved by a few quirks of idiom. It is saved by the context 'not only'.

Granted, I am treading rather rocky terrain. Cases of what we would like to regard as logical truth might still turn up that are not saved by any quirk of idiom, and cases of what we would not like to regard as logical truth may issue from quirks of idiom. The firm boundary will depend rather on some actual tampering with grammar on the logician's part. I shall come to that presently.

Meanwhile let us remind ourselves of another point: in drawing lines between grammatical categories we are dependent still on a distinction between the merely absurd sentiment and the flatly ungrammatical; for we had to rule that 'There was salt new on the Rialto' is ungrammatical. Carnap's example, 'This stone is thinking about Vienna', would be viewed as absurd but grammatical; so also, probably, would Russell's example 'Quadruplicity drinks procrastination'. What is required as a wedge between grammatical categories is not just the absurd, but the ungrammatical. How do we draw that line? William Haas has offered an interesting

answer.¹ An appropriate occasion for the merely absurd can be built up in a fanciful fairy story. The more absurd the example, the more difficult the build-up; it is a matter of degree. But the ungrammatical cannot be accommodated at all in stories, however fanciful; it can be accommodated only by concocting deviant dialects.

Thus far we have laid up two logical truths: 'If it rained then it rained or it snowed' and 'If sugar is sweet then something is sweet'. We are apt to run thus to conditionals, since each logically true conditional presents an implication, and implication is the object of the game. Now here, according to the definition, are two more for our growing collection:

If Tom is taller than Dick then Dick is not taller than Tom.

If Tom is taller than Dick and Dick is taller than Harry then Tom is taller than Harry.

The grammatical particles concerned here are 'is', 'not', 'then', 'and', and the suffix '-er'. There can be no quarrel with these logical truths or the corresponding implications, but they are not in the logic books. Let us see why. What they reflect is the asymmetry and transitivity of two-place predicates formed with the suffix '-er'; and the point is theoretically insignificant, because many other predicates share those same properties without flaunting them with any marker such as 'er'. This is as it should be, since the asymmetry or transitivity of some predicates may even be an as yet undiscovered fact of nature. Moreover, asymmetry and transitivity are merely two among endlessly many properties of various predicates; why single them out? The natural course is just to state them as premises where needed.

Are we to say then that our proposed definition of logical truth and implications is faulty after all, and that 'Tom is taller than Dick' does not really imply 'Dick is not taller than Tom' logically? And correspondingly for the transitivity example? No, I would adhere to the definitions. I think most people would agree that these examples are good logical implications, even though they are not in the books. I say that they are not in the logician's book because the logician is revising grammar itself, and that the definitions of logical truth and implication in terms of grammar continue to be good. Revision of grammar is an important part of the logician's activity, and I want to discuss its nature and purposes.

1. "Meanings and Rules," *Proceedings of the Aristotelian Society* (1973): 135–155.

His revision of grammar in the present instance does not change the boundary between grammatical and ungrammatical expressions. Some of his other revisions do, as we shall see; but this one is only a change in the way we analyze the expressions. When he skips the asymmetry and transitivity of 'taller than' as none of his business, we may think of him as repudiating the grammatical particles '-er' and 'than' and treating 'taller than' as a single morpheme. It is an attractive move semantically, by the way, since 'tall' makes no clear sense anyway except by comparison. A tall man is tall for a man but not for a building; and 'tall for a man' simply means '*taller than* most men'.

Relative to grammar as thus revised, the asymmetry and transitivity of 'taller than' cease to count as logical implication; for they are not reflected in the new grammatical structure. But they may continue to be recognized as implication in a broader, extra-logical sense. Implication in this broader sense holds in many cases where there is no suggestion of relevant grammatical structure. An example is the symmetry of 'cousin': 'Tom is Dick's cousin' implies, in some sense, 'Dick is Tom's cousin'. Such implication rests on meanings of lexical elements, not on grammatical particles; and what is vague about it is the notion of meaning. Now the asymmetry and transitivity of 'taller than' continue to qualify as implications in this vaguer semantic sense, even if with the revisionary logician we view 'taller than' as a monolithic morpheme.

For the latter-day logician, logical regimentation of grammar is standard procedure. We have noted the motivation of one such reform, and it is characteristic of many. Others of his reforms serve to resolve structural ambiguities; others serve to economize on constructions. His interest in grammatical structure is one-sided: he is interested in how it channels truth conditions. If a grammatical reform makes for a more copious channeling of truth conditions and causes no complications in other quarters, he is happy to adopt it. He adopts it not as a reform to be imposed on society, but as a technical by-language to expedite scientific inference. The shift is the same in principle as programming a computer, and the same, for that matter, as the mathematician's habitual recourse to planned notations.

Would it be trivial to argue merely that the logician *can find* a revision of grammar that would make the grammatical forms coincide with the logical forms intuitively so-called? No, this would not be trivial; not if the revised grammar was still adequate to an all-purpose scientific language. But anyway what I mean to say goes beyond this: I am talking of the logician's

motive. He is not revising grammar merely to fit grammatical form to what he already happened to call logical form. My thought is rather this: what we call logical form is what grammatical form becomes when grammar is revised so as to make for efficient general methods of exploring the interdependence of sentences in respect of their truth values.

Let us therefore consider, from a practical point of view, what line the logician's grammatical regimentations might have been expected to take. Let us proceed as if we did not know the line actually taken. We may caricature him as a consulting technician whose job it is, given two sentences of some science, to determine whether the one implies the other. He will unite them in a conditional and regiment its outermost structure first. He will then work inward as needed, stopping when he has established logical truth. His first concern will thus be with grammatical constructions that are capable of being outermost—hence with sentence-forming constructions. His is a strategy of shallow analysis: expose no more structure than is needed for the purpose at hand. In paraphrasing the constructions of ordinary language into his regimented grammar, moreover, he will cheerfully relinquish various connotations of the old idioms, unless they seem germane to the scientific business in hand. He will be intent on truth conditions, since he is seeking logical truth. Taking these considerations together, we may well expect the logician's first concern to be with sentence-forming constructions, and of these we may expect him to favor truth functions where he can.

The necessity operator, like the truth functions, builds sentences from sentences. Some logicians admit this construction to their regimented grammar. Others, however, exclude the notion—not just from logic, but altogether—as obscure and ill-conceived. There are further constructions, of course, that likewise build sentences from sentences. One is 'because'. This construction cannot be so lightly dismissed from general discourse as the necessity idiom can, but it is equally problematic. The causal relation is notoriously obscure, obscure out of all proportion to what it would contribute to logical truth or implication. Further constructions that build sentences from sentences are the idioms of prepositional attitude—thus '*x* believes that *p*', '*x* hopes that *p*'. These again are not to be lightly dismissed from general discourse, but they are logically inert: patterns of implication are not in evidence at all, except that '*x* knows that *p*' and '*x* regrets that *p*' and a few others may be said to imply '*p*'.

The truth functions, at any rate, emerge as the logician's first concern.

But he cannot rest thus with the resolving of sentences into component sentences. There are implications that depend on more fragmentary components. Here again, heeding the maxim of shallow analysis, the logician will shun undue fragmentation.

What a sentence says about anything can be assembled in a self-contained expression by means of a grammatical construction that I call *predicate abstraction*. It is the ‘such that’ construction, and is a logical regimentation of the relative clause. The role of the relative pronoun is played by a bound variable, and the words ‘such that’ are rendered by an inverted epsilon. Thus consider the sentence ‘Paul lost his passport’. What it says about Paul is ‘ $x \ni (x \text{ lost } x\text{'s passport})$ ’. What it says about his passport is ‘ $y \ni (\text{Paul lost } y)$ ’. These ‘such that’ clauses are not names of properties or classes; they are mere predicates, tantamount to relative clauses or to verb phrases. The clause ‘ $x \ni (x \text{ lost } x\text{'s passport})$ ’ is equivalent to the predicate ‘lost one’s passport’, and the clause ‘ $y \ni (\text{Paul lost } y)$ ’ is equivalent to ‘lost by Paul’. The point of converting to the ‘such that’ style, with its variable, is that it extracts or abstracts predicates uniformly from all manner of sentences, without requiring various structural transformations within the sentence. To get the predicate ‘lost one’s passport’ in the vernacular we had to change an original ‘his’ to ‘one’s’; to get the predicate ‘lost by Paul’ we had to switch to passive voice; and other examples would call for other maneuvers, often quite devious. The ‘such that’ construction well illustrates what the logician is up to in his grammatical reforms. He seeks structural uniformity so that his tests or proofs of implication may proceed smoothly according to general rules.

Observe, moreover, the contrast between this modification of grammar and the earlier reform that consisted in treating comparative adjectives as single morphemes. That was only a change in the mode of analysis, and did not affect the linguistic output. On the other hand the ‘such that’ construction comes as an innovation even in outward form.

Predication is the grammatical construction that consists in adjoining a predicate to a singular term to recover a sentence, thus:

$[x \ni (x \text{ lost } x\text{'s passport})]$ Paul,
 $[y \ni (\text{Paul lost } y)]$ (Paul’s passport).

Each of these sentences collapses to ‘Paul lost Paul’s passport’ by the logical law of *concretion*.

Predicates are not only for predicating. If that were their only use, there

would be no point in predicate abstraction. Instead of predicating ' $x \ni (x \text{ lost } x\text{'s passport})$ ' of Paul or other persons, we could as well write Paul's name or others in place of ' x ' in ' $x \text{ lost } x\text{'s passport}$ '. Another conspicuous use of predicates, however, is in existence statements. The existence functor ' $\exists$ ' attaches to a predicate to form a sentence, true or false according as there is or is not anything that satisfies the predicate. Thus the combination of symbols ' $\exists x \ni$ ' amounts in the more usual notation to the quantifier ' $(\exists x)$ '. Actually the line of evolution that I am describing is not fanciful. The functor ' $\exists$ ' and the prefix ' $x \ni$ ' are lifted verbatim, or literatim, from the pioneer work of Peano, where existential quantification emerged explicitly as that combination.

In practice I do use the notation ' $(\exists x)$ ', but there is some philosophical point in analyzing it as ' $\exists x \ni$ '. The predicate abstract, or 'such that' clause, is a direct regimentation of the relative clause, and the use of variables in predicate abstraction is the use that most clearly reflects what bound variables are wanted for. In all their familiar uses—in quantification, in singular description, in class abstraction—their essential contribution is the same: by their recurrences they keep track of permutations and repetitions of references. There is in this nothing distinctive to do with classes or description or the universal and particular categoricals. Where we see the variable clear of foreign entanglements is in the 'such that' clause. Here the variable emerges as an instrument purely of extrication: of rearranging a sentence around some chosen component, so as to segregate it from something it was about. As for quantification, description, and class abstraction, these are most naturally rendered directly as functors upon predicates; any variables present are for abstracting those predicates themselves. We already saw ' $\exists$ ' as a sentence-forming functor on predicates: ' $\exists F$ ', 'there are F '. The ' x ' of ' $(\exists x)$ ' gives way to the ' x ' of abstraction of the predicate to which ' $\exists$ ' is applied. The ' $\forall$ ' of universal quantification becomes another sentence-forming functor on predicates: ' $\forall F$ ', 'everything is F '. The ' ι ' of description becomes a term-forming functor on predicates: ' ιF ', 'the F '. Class abstraction calls for an unaccustomed term-forming functor ' ς ': ςF is the class of all F s. The variable, properly considered, is neutral amid all this. It is simply the heir to the relative pronoun.

It is well known that names and other singular terms can be dispensed with in favor of predicates and bound variables. Instead of inventing a name ' n ' to pick out some object uniquely, we could as well invent a predicate ' P ' which is to be true of that object uniquely. Then, instead of embed-

ding the name '*n*' in a sentence, '*... n ...*,' in order to say something about the object, we could as well embed an existentially quantified variable in the sentence, subject to the identifying predicate '*P*'; thus,

$$(\exists x)(Px \cdot \dots x \dots), \quad \text{or} \quad \exists x \exists (Px \cdot \dots x \dots).$$

In practice even in formal logic it is more convenient to keep the names and other singular terms, but their eliminability is instructive in a number of theoretical connections, including the present one.

I characterized predicate abstraction as a way of segregating what a sentence says about an object. Commonly the predicate thus abstracted does not exhaust what the sentence says about the object, since the term that designates it will also say something about it in order to identify it. Now that we have eliminated singular terms other than variables, however, designation goes by the board; there is only the variable. All that an open sentence says about the value of the variable is said by the rest of the sentence; the variable says none of it. The variable is the legitimate latter-day embodiment of the incoherent old idea of a bare particular.

The logician's regimented grammar has come down to something slight. Conjunction and negation suffice as truth functions. In addition there is predicate abstraction, with its variables, and there is ' $\exists$ '. Or we can lump these latter two together and call it quantification, ' $\exists x \exists$ '; for the logician needs the two only thus combined. Predication of predicate abstracts was seen to be pointless, so predication is not called for under the logician's regimentation. As for universal quantification, it can be got from existential quantification and negation in familiar fashion; thus ' $\sim \exists x \sim$ '. We have got down to the grammar of standard or neo-classical logic; simply quantification and the truth functions.

Some logicians opt for more. Commonly, of course, they provide a grammatical niche for singular terms other than variables; but this can be seen as a mere matter of style, since it can be avoided without theoretical loss. Some logicians allow also for tense, but this is avoidable by admitting times as values of variables and exploiting temporal predicates such as 'earlier than'. Some logicians admit, as remarked, a modal necessity operator.

I said that the logician's central concern is logical implication, and that the avenue to it is logical truth. For establishing logical truth he needs a formal technique or calculus. It is directed to the recognizing or generating of logically valid grammatical forms; for the forms, not the sentences, are

what present the bare logical essentials. The sentences differ from the forms precisely in containing also the filler that does *not* contribute to validity or logical truth, but only gets in the logician's way.

These grammatical forms are depicted notationally in schemata, which depict the grammatical forms of sentences that have been regimented from the outside inward to varying depths. Unanalyzed interior portions of sentences, however complex, are schematized simply by single letters '*p*', '*q*', etc., or by '*Fx*', '*Fy*', '*Gxy*', etc. if we want to signal the presence of variables buried inside them.

Was I wrong, then, in saying that the logician's way of regimenting grammar does not assume predication? Is '*Fx*' not a case of predication? Confusion over this point is almost unavoidable. Quantification logic is even called the predicate calculus, and the schematic letter '*F*' is called a predicate letter. It is quite reasonably called a predicate letter, occupying as it does a position appropriate to a predicate; and the juxtaposition '*Fx*' then does indeed represent predication. But the point to appreciate is that '*Fx*' is no part of the logician's regimentation of grammar. It does not enter the sentences, but appears only on the logician's work sheet. Think of the logician again as the consultant, auditing the scientist's sentences for implication. He paraphrases the outward structure of sentences, imposing his quantifiers and truth function signs. He presses inward, imposing still only quantification and truth functions. When he stops doing that, he just abbreviates the unanalyzed internal residues as '*p*' or '*q*' or '*Fx*' or the like; the '*Fx*' says nothing about grammatical structure, but merely reminds him that there is a free occurrence of '*x*' somewhere inside.

I have been eliciting the logician's regimentation of grammar by reflecting on his concern with truth links and his strategy of paraphrasing inward. Because of his maxim of shallow analysis, he is not called upon to inquire into the atomic structure of sentences. Once he has paraphrased deep enough to clinch a desired implication, he will render any as yet unanalyzed segment of text by a schematic letter or so—no matter how complex that segment may be. The structure deeper in the interior is not his concern.

An easy further step, however, produces a self-sufficient grammar; we may take simple predicates, with one or more variables attached, as the atomic sentences. The logical grammar then suffices to generate the superstructure. Such is the blueprint for a general scientific language. In such a language the grammatical form of any sentence is precisely its logical form,

plus predication at the atomic level: predication of one-place simple predicates and of many-place ones as well. Such predication is an additional grammatical construction, for we saw that predication had no place in the grammatical structure that the logician imposes in paraphrasing inward from outside. In recognizing predication thus at the atomic level we round out a self-contained grammar, but we add nothing to the logical truths or implications. The grammatical structure that we have projected into the atomic sentences proves to be logically inert.

It will be recalled that the predicate abstraction that was the mainstay of our logical grammar produced only one-place predicates. Only in our present little excursion into the atomic level are we recognizing many-place predicates, and we are finding them logically inert. This is puzzling, for we were brought up to believe rather that a major strength of modern logic lies precisely in its mastery of many-place predicates. We have been told that the formal logic of past centuries was inadequate to the logic of relations and that it was the glory of modern logic to have overcome this limitation.

We can clear up this puzzle by considering what that old limitation was, more exactly, and how it was overcome. Described anachronistically in terms of variables, the limitation was that no two variables could occupy an unanalyzed part of a sentence. Thus picture the old-fashioned logician paraphrasing the outward structure of a sentence and working inward. His analysis will never reach the point of exposing two references to objects, unless it exposes them as occupying opposite components of some already exposed grammatical construction; this was the old limitation. When we recognize the 'such that' construction as a part of logical grammar, we overcome that limitation. Analysis can proceed to the point of exposing a predicate within a predicate, e.g., in the manner:

$$x \ni \exists y \ni (\dots x \dots y \dots),$$

and there it can quite well stop, leaving the unanalyzed interior text ' $\dots x \dots y \dots$ ' with the two variables associated no matter how. Thus it is that the abstraction of one-place predicates, when nested, can do many-place business.

We saw how to extend the logician's regimented grammar into atomic sentences to form an overall grammar for a language. It is perhaps rash to say that a language with that grammar can be generally adequate. We may well dismiss the necessity operator as useless or unacceptable, but one hesi-

tates to be so cavalier with the idioms of prepositional attitude. Still I do see this language pattern as a valid ideal of clarity. I find a scientific theory finally satisfactory only if I can see how to cast it in such a form.

In closing I should say something of the logic of identity, which I have persisted in skirting. I have been characterizing logical truth and implication by grammatical structure. This account excludes the logic of identity, since the identity sign is no mere grammatical particle but a predicate in the lexicon. The exclusion seems odd, since we always think of identity theory as part of logic.

We can recognize logic in a narrow sense as hinging wholly on grammatical structure, and so as excluding the theory of identity, and logic in a somewhat wider sense as hinging on grammatical structure and the identity predicate. This latter concession is not altogether arbitrary. A notable trait of logic is its central position in our conceptual scheme; it is an integral part, as Tarski has remarked, of every branch of science. Now identity is similarly central. It is definable, nearly enough, in each particular language, so long at any rate as the lexicon of simple predicates is closed off at some finite point. For, as is well known, we can then define ' $x = y$ ' by exhaustion of atomic contexts:

$$Fx \equiv Fy \cdot \dots \cdot (z)(Gxz \equiv Gyz) \cdot (z)(Gzx \equiv Gzy) \cdot \dots$$

If we view identity as so defined, there is a way after all of admitting identity to logic without giving up our grammatical theory of logical truth. We can view the identity sign as a schematic predicate symbol, subject to a different interpretation in each particular language. The languages concerned are all to share the regimented grammar of truth functions, quantification, and atomic predication that we have been contemplating. They differ from one another only in their lexica of simple predicates. The identity sign, then, becomes a schematic symbol which is to be interpreted in each of these particular languages by exhaustion of the atomic contexts in that language. True sentences of the theory of identity then—

$$'(x)(x = x)', '(x)(y)(x = y \cdot \supset \cdot y = x)', \text{ and the like—}$$

become logical truths after all, in the strict sense of being logically valid grammatical forms, when interpreted according to that plan in each particular language. On this plan the identity sign does not qualify as a grammatical particle, but its laws still belong to logic.

Ontology and Ideology Revisited

1983

To be, I have persistently held, is to be the value of a variable. More precisely, what one takes there to be are what one admits as values of one's bound variables. The point has been recognized as obvious and trivial, but it has also been deemed unacceptable, even by readers who share my general philosophical outlook. I shall begin by sorting out the pertinent considerations. Some of them hinge on confusion, but the more serious objections reflect a genuine awareness that a conspicuous dimension of existence ordinarily so called has been passed over. In conclusion I shall suggest how to meet those just demands.

The artificial notation ' $\exists x$ ' of existential quantification is explained merely as a symbolic rendering of the words 'there is something x such that'. So, whatever more one may care to say about being or existence, what there are taken to be are assuredly just what are taken to qualify as values of ' x ' in quantifications. The point is thus trivial and obvious.

It has been objected that what there is, is a question of fact and not of language. True enough. Saying or implying what there is, however, is a matter of language; and this is the place of the bound variables.

It has been objected that the logical notation of quantification is an arbitrary and parochial standard to adopt for ontological commitment. The answer is that the standard is transferable to any alternative language, insofar as we are agreed on how to translate quantification into it. For predicate-functor logic, thus, the equivalent principle is that what one takes to be are what one takes one's monadic predicates (complements included) to be true of. For ordinary English it works out that what one takes there to be are what one takes one's relative pronouns to refer to.¹ Ordinary

1. See my *Theories and Things* (Cambridge, MA: Harvard, 1981), 5/6.

discourse is indeed seldom meticulous about ontology, and consequently an assessment based on the relative pronouns of ordinary discourse is apt to bespeak a pretty untidy world; but ontological clarity and economy can be promoted by paraphrase, if one so desires, in terms still of relative clauses and pronouns rather than quantifiers and bound variables. The notation of quantification is what is most usual and familiar, currently, where one is expressly concerned with ontological niceties; hence my choice of it as paradigm.

If in some language we are at a loss to arrive at a satisfactory contextual translation of 'there is', and hence of existential quantification, then we are at a loss to assess the ontology of the speakers of that language. Some languages are perhaps so unlike ours that any translation of 'there is' or ' $\exists x$ ', however cunningly contextual, would be too far-fetched and Procrustean to rest with. To entertain the notion of an ontology at all, known or unknown, for the speakers of such a language would be an unwarranted projection on our own part of a parochial category appropriate only to our own linguistic circle.

Thus I do recognize that the question of ontological commitment is parochial, though within a much broader parish than that of the speakers and writers of symbolic logic. Some of my readers have objected even to this limitation, arguing that even a dumb animal reacts distinctively to particular things and kinds of things, thereby reifying them, recognizing their existence, incorporating them into his inarticulate ontology. Here at last is a valid point that I want to accommodate in my subsequent constructions.

Another such point has to do with what I call proxy functions. I have noted elsewhere² that if we transform our ontology by any arbitrary one-to-one transformation, and then reinterpret our predicates in a compensatory way, our entire theory of the world will persist verbatim and all its evidential links with sensory stimulation will likewise continue undisturbed. I have pointed the moral that what matters is structure; the objects, concrete and abstract, familiar and recondite, matter only as neutral nodes in the structure. It is not to be wondered that this conclusion has been found disturbing, since it seems to belie, after all, the factuality of the question what there is. I hold that actually it does not, since the question what there is, is a substantive scientific question within our system of the world, whereas the

2. Ibid., 19–21.

point about proxy functions is a point only about evidence, a point not of ontology but of the epistemology of ontology. Still I agree with my disturbed readers that ontology has undergone a humiliating demotion. This again is a point which my ensuing constructions may somewhat mitigate.

I long ago distinguished between ontology and what I called ideology,³ this latter being one's stock of simple and complex terms or predicates. I now propose to extend ideology beyond the subject's own verbal limits, to cover inarticulate abilities to recognize and discriminate. If the subject is disposed to react to the presence of any and every cat in some manner in which he is not disposed to react to anything but cats, then I shall reckon the term 'cat' to his ideology, though it is strictly my term, with which he may or may not be conversant. He need have no corresponding term; he may be dumb; his reaction to cats may even be allergic, and independent of the five senses. Again if he is disposed to react to Geach's Tabitha as to no other cat or noncat, I shall reckon 'Tabitha' to his ideology, though he be ignorant of the name.

For further generality let me switch from terms or predicates to occasion sentences. If the subject reacts distinctively in the rain, or at sunset, or when the chill factor drops below zero Fahrenheit, then the occasion sentence, 'It's raining', 'The sun is setting', or 'The chill factor is below zero Fahrenheit' will be reckoned to his ideology, irrespective of his powers of verbal expression. If he has the gift of language, his ability to call a spade a spade is just one among his dispositions to react in distinguishing ways.

Ideology so construed may be called *perceptual* ideology, to mark both its breadth and its limits. It is broad in transcending the subject's lexicon, if any, and narrow in treating only of his direct responses to present stimulation. It accommodates dumb animals and remote aliens, thus supplying what was found wanting in my appeal to values of variables. Also, since it shuns the subject's semantics, it is untouched by the consideration of proxy functions except as this bears on the theory of evidence of all science.

From perceptual ideology we could even derive, as a special case, a notion of perceptual ontology. Thus we might reckon all cats to a subject's perceptual ontology if 'It's a cat' belongs to his ideology. We might reckon Tabitha to his perceptual ontology so long as 'That's Tabitha' belongs to his

3. "Ontology and Ideology," reprinted in *Ways of Paradox* (New York: Random House, 1966, and Cambridge, MA: Harvard, 1976.)

ideology, even if 'It's a cat' does not. Should we reckon further elements to his perceptual ontology on the strength of such further elements of his ideology as 'It's raining' and 'The sun is setting' and the one about the chill factor? The choice is ours, and depends on our own ontology in my serious sense, the values of our bound variables. If we countenance such things as rain, sunsets, and chill factors as values of our variables, we may choose to reckon them to the subject's perceptual ontology in the light of his ideology.

Such a notion of perceptual ontology would be straightforward but idle, given that of perceptual ideology. The latter already covers the ground, and it dispenses with any arbitrary subsequent decision as to what to reckon as objects. For behavioral psychology and ethology we should look merely to perceptual ideology as the key idea, and for a properly philosophical theory of reference and ontology we should continue to look to the values of the variables. It is here that we can take account of the reification not only of perceptible things and situations but also of microphysical particulars and abstract objects. It is here that we can make sense of ventures in ontological economy, and of the place of ontology in Tarski's recursive definition of satisfaction and truth. It is here, finally, that we come to appreciate the role of reification in the theory of scientific evidence: that it is a mere positing of neutral nodes in the structure of scientific theory, as shown by the proxy functions.

Relativism and Absolutism

1984

My view of science involves both relativistic and absolutistic strains. I shall try to sort them out.

Scientific theory is supported in part by systematic considerations of coherence, notably simplicity in some sense, and in part by external information in the form of sensory stimulation. This external information finds its way into the formulation of the theory through the medium of certain sentences which are keyed exclusively, for our verdicts, to concurrent stimulation. Examples are 'It's raining', 'It's cold', 'That's a dog'; present stimulation decides them for us irrespective of what else may have been going on lately in our minds or environment. I call them observation sentences, but, as these examples show, they are not *about* observations or sensations. They are not even incorrigible; they may be recorded and subsequently repudiated for systematic reasons. But they command unwavering verdicts at the time.

Relativism already intrudes. 'That's an X-ray machine' will qualify as observational, in the present sense, for an initiate and not for a novice. This relativism, however, is readily transcended. The empirical evidence for a theory is reducible ultimately to what can be conveyed in observation sentences at the novice's level. It just takes a little stubbornness, a Missourian insistence on being shown.

Relativism of a higher order remains: relativity to one's language. The sentences, observational or otherwise, of one language are the gibberish of another. But this relativism, again, can be transcended. Observationality of a sentence consists in sheer concomitance between verdicts and concurrent external stimulatory situations; so we can teach our observation sentences

to a foreigner by simple conditioning. Perhaps we can sum this up by saying that observation, properly so called, is independent of language.

There remains, even so, a relativity to man's implicit standards of similarity. A conditioned response is conditioned not to a unique stimulation but to a class of sufficiently similar stimulations. People have to be in substantial agreement, however unconscious, as to what counts as similar if they are to succeed in learning, one person from another, when next to assent to a given observation sentence. Here, then, is an irreducible kernel of relativism: all sensory evidence as reflected in observation sentences is relative to the neural organization that determines what different triggerings of nerve endings will favor the same response. Subjects radically at odds in this neural way could never learn observation sentences or anything else from one another.¹ Our training even of a dog, horse, bear, seal, or elephant hinges on a conformity of his inarticulate similarity standards to our own. Such, then, is the residual relativism of empirical evidence.

Next let us consider in what further way science may be said to be relative, supposing its empirical evidence fixed. It has been urged by many, including me, that scientific theory is under-determined by all possible data; in other words, that different theories can be empirically equivalent.² But this depends on what verbal formulations to count as formulations of different theories and what ones to count rather as different formulations of one and the same theory; and surely this question is philosophically uninteresting. What we are given to compare are the verbal formulations, variously unlike but empirically equivalent. Whether and when to rate them as formulations of the same theory is an inconsequential question of words; we can take the theory formulations and let the theories go. At this point the question of relativism versus absolutism, as applied to empirically equivalent theory formulations, loses its substance.

I still maintain, in the way or in lieu of the old thesis of under-determination, that empirically equivalent theory formulations can be drastically unlike; but on this thesis there can scarcely be disagreement except as to the possible degree of unlikeness. I have pointed out elsewhere that even the ontology of a theory formulation—that is, the values of its

1. This sentence appeared in the original published version but is marked for deletion in Quine's personal copy of the printed paper.

2. Reasonably clear sense can be made of this appeal to all possible data, or empirical equivalence. See my *Theories and Things* (Cambridge, MA: Harvard University Press, 1981), 24–28.

bound variables—can be subjected to any arbitrary “proxy function” or one-to-one transformation, and the resulting theory formulation will still be empirically equivalent to the old if we systematically revise its terms in a compensatory way.³ This perhaps smells of relativism, but the odor is vague.

Or is the odor sharp and unmistakable when the two empirically equivalent theory formulations are not merely unlike but logically incompatible, as they are apt to be in the case of the proxy function? No; on closer inspection, logical incompatibility on the part of empirically equivalent theory formulations is seen to be a red herring. For, imagine two such formulations. There must be a sentence that is implied by one of the formulations and whose negation is implied by the other. It must be a sentence whose truth value is undetermined by all possible observations, since the two theory formulations are empirically equivalent. At least one term in the sentence, then, must be a theoretical term whose conditions of application are not wholly fixed by empirical criteria. We can exploit its slack by treating the term rather as two inequivalent terms, one in the one theory formulation and the other in the other, and changing its spelling in one of the theory formulations. The two sentences cease to be contradictories. Similar adjustments elsewhere, as needed, render the two theory formulations logically compatible. I owe this expedient to Donald Davidson.

We are thus left only with empirically equivalent theory formulations that are logically reconcilable. If we subscribe to one of them as true, we can call them all true and view them as different descriptions of one and the same world. We are no strangers, after all, to strange languages. If this be relativism, make the most of it.

There is indeed a clear and trivial sense in which truth, Tarski’s, is relative to language. What are true are sentences, and a sentence of one language is—I say it again—the gibberish of another. Relativism in this sense I can unabashedly own.

At the same time there is an absolutism, a robust realism, that is part and parcel of my naturalism. Science itself, in a broad sense, and not some ulterior philosophy, is where judgment is properly passed, however fallibly, on questions of truth and reality. What is affirmed there, on the best available evidence, is affirmed as absolutely true.

This is not to say that it is affirmed as certain. Part of scientific method

3. *Ibid.*, 19–23.

is readiness to change one's mind. When one does so, one recants the ascription of truth; but Truth with a capital 'T' is what it is a question of, first and last.

Belief and evidence, unlike truth, remain a matter of degree. What then of knowledge? Knowledge, nearly enough, is true belief on strong evidence. How strong? There is no significant cut off point. 'Know' is like 'big': useful and unobjectionable in the vernacular where we acquiesce in vagueness, but unsuited to technical use because of lacking a precise boundary. Epistemology, or the theory of knowledge, blushes for its name.

The truth of physical theory and the reality of microphysical particles, gross bodies, numbers, sets, are not impugned by what I have said of proxy functions and of wildly deviant but empirically equivalent theory formulations. Those remarks had to do not with what there is and what is true about the world, but only with the evidence for what there is and what is true about the world. I was showing that scientific discourse radically unlike our own, structurally and ontologically, could claim equal evidence and that we are free to switch. Still we can treat of the world and its objects only within some scientific idiom, this or another; there are others, but none higher. Such, then, is my absolutism. Or does it ring relativistic after all?⁴

4. I thank Burton Dreben for helpful comments on an earlier draft.

States of Mind

1985

Unless a case is to be made for disembodied spirits, we can argue that a dualism of mind and body is an idle redundancy.¹ Corresponding to every mental state, however fleeting or however remotely intellectual, the dualist is bound to admit the existence of a bodily state that obtains when and only when the mental one obtains. The bodily state is trivially specifiable in the dualist's own terms, simply as the state of accompanying a mind that is in that mental state. Instead of ascribing the one state to the mind, then, we may equivalently ascribe the other to the body. The mind goes by the board, and will not be missed.

Not that we would continue to refer to the bodily state as the state of accompanying a mind in the mental state. That formulation was for the dualist's benefit, to show him that the bodily state was undeniable from his own point of view and specifiable in his own terms. For our part, we just appropriate the mentalistic terms themselves and construe them as referring to those bodily states. We even continue to speak of the states as mental. The only change is that we reckon mental states as states of another substance, the mind.

If this effortless physicalism smacks of trickery, we may do well to reflect on how we learn the mentalistic terms in the first place. All talk about one's mental life presupposes external reference. Introspect our mental states as we will, how do we know what to call them? How did we learn to call our anxieties anxieties, our dull aches dull aches, our joys joys, and our aware-

1. This is an outgrowth of my piece in a colloquium of the American Philosophical Association, Detroit, 1980. Burton Dreben lately suggested some changes and urged publication. The key idea of the first half was briefly noted meanwhile in my *Theories and Things* (Cambridge, MA: Harvard, 1981), 18.

ness awareness? Why do we suppose that what we call joys and anxieties are what other people call by those names? How do we know what we are talking about? Clearly the answer is that such terms are applied in the light of publicly observable symptoms; bodily symptoms strictly of bodily states, and the mind is as may be. Someone observes my joyful or anxious expression, or perhaps observes my gratifying or threatening situation itself, or hears me tell about it. She then applies the word 'joy' or 'anxiety'. After another such lesson or two I find myself applying these words to some of my subsequent states in cases where no outward signs are to be observed beyond my report itself. Without the outward signs to begin with, mentalistic terms could not be learned at all. To apply these terms to the state of the body is just to put them back where they belonged to begin with.

I am not applying the terms to behavior. A mental state is not always manifested in behavior. Physically construed, it is a state of nerves. We can say which state it is, however, and tell one from another, without knowing the neural mechanism. We specify it with help of the mentalistic term, which in turn was learned on the strength of behavioral signs. Thus the net result, if we shortcut the mental bit, is that the behavior provides incomplete and sporadic symptoms whereby to identify and distinguish various complex states of nerves whereof the neural detail may still be a matter of conjecture; and it is these states that the mental terms may be seen as denoting.

Mental states, construed as states of nerves, are like diseases. A disease may be diagnosed in the light of observable signs though the guilty germ be still unknown to science. Incidentally, diagnosis depends heavily on symptoms reported by the patient; and such is the way, overwhelmingly, with the detection of mental states.

What now can we make of the difference between identifying the mental states with the states of nerves, as I just did, and repudiating them rather in favor of states of nerves? I see no difference. In either case the states of nerves are retained, mental states in any other sense are repudiated, and the mental terms are thereupon appropriated to states of nerves. So I may as well persist in calling my proposed reduction of mind to body an identification of mental states with bodily ones, neural ones; a construing of the mental as neural.

There is no presumption that the mentalistic idioms would in general be translatable into the anatomical and biochemical terminology of neuro-

ogy, even if all details of the neurological mechanisms were understood. Thus take belief. Assessed on its objective manifestations, belief is a very mixed bag. Lip service is our most convenient clue to belief, but is neither a necessary nor a sufficient condition. Acceptance of wagers is a firmer sign, and the accepted odds even afford a measure of the strength of the belief; but this test is available only if there is a prospect of subsequently finding the answer, acceptably to both parties, and settling the bet. Other behavior, such as searching or fleeing or standing expectantly, can serve tentatively as manifestations of one or another belief, but these manifestations vary drastically and unsystematically with the content of the belief to be ascribed. Other grounds for ascribing beliefs may be sought unsystematically by probing the subject's past for probable causes of his present state of mind, or by seeing how he will defend his purported belief when challenged.

The empirical content of ascriptions of belief is thus heterogeneous in the extreme, and the physiological mechanisms involved are no less so. The heterogeneity is cloaked under a linguistic uniformity: the connective 'believes that' followed by a subordinate sentence. The other idioms of prepositional attitude have the same disarmingly uniform structure as the belief idiom, and it cloaks much the same heterogeneity in respect of empirical evidence and neural mechanism. Hence the insistence, from Brentano onward, on the need for an independent science of intension.² Even those of us who do not acquiesce in a metaphysical dualism of mind and body must make the best of what Davidson has called anomalous monism.³

The stubborn idioms of prepositional attitude are as deeply rooted as the overtly physicalistic ones. One of them is almost coeval, it would seem, with observation sentences. Thus take the observation sentence 'It's raining'. Tom is learning it from Martha by ostension. Martha's business is to encourage Tom in uttering the sentence, or in assenting to it, when she sees that he is noticing appropriate phenomena, and to discourage him otherwise. Thus Tom's mastery of the physicalistic sentence 'It's raining' hinges on Martha's mastery, virtual if not literal, of the mentalistic sentence 'Tom perceives that it's raining'.

Observation sentences, learned ostensibly, are where our command of language begins, and our learning them from our elders depends heavily

2. See Roderick Chisholm, *Perceiving* (Ithaca, N.Y.: Cornell, 1957), chapter 11.

3. Donald Davidson, *Action and Events* (New York: Oxford, 1980), 214–225.

on the ability of our elders to guess that we are getting the appropriate perception. The handing down of language is thus implemented by a continuing command, tacit at least, of the idiom ‘ x perceives that p ’ where ‘ p ’ ranges over observation sentences. This degree of implicit mastery of one idiom of prepositional attitude would seem therefore to be nearly as old as language. The structure of this idiom, moreover—its embedding of a subordinate sentence—would have been clearly dictated by its primitive use in assessing children’s acquisition of observation sentences. Analogical extension of the idiom to other than observation sentences would follow inevitably, and the development of parallel idioms for other prepositional attitudes would then come naturally too, notwithstanding their opacity from a logical point of view. Naturalness is one thing, transparency another; familiarity one, clarity another.

The Sensory Support of Science

1986

Epistemology, or the theory of knowledge, was primarily concerned in recent centuries with determining and justifying the steps by which we derive our knowledge of the external world from our immediate sensory experience. Immediate sensory experience was regarded as consisting of bits that Hume called impressions and others have called *sensa*, *sensabilia*, or *sense data*.

Meanwhile there have been hard-headed types, or perhaps I should say blithe spirits, who looked upon *sense data* as a pointless intrusion of trouble-makers: why not just recognize that we perceive external objects outright, sticks and stones as sticks and stones?

One obvious difficulty with this position is the difficulty of drawing a line between what to count as direct perception of an external object and what to count as indirect or inferred: for surely we are constantly inferring such things as a half-hidden stick or the unseen back of a stone, let alone an unseen ninth planet or a host of elementary particles. What is sensed directly and what is not? The positing of *sense data* was an effort to segregate the immediate.

The Gestalt psychologists took an intermediate line. They opposed the atomistic notion of *sense data* on experimental grounds. They established that what people and other respectable animals are aware of, from early infancy, are not the simple elements of sensation, but various significant complexes of them, taken as wholes.

Such have been the disagreements over what to take as the epistemological given: how to reckon epistemic priority. Should awareness matter, as the Gestalt psychologists held? What are we trying to do?

The question is what to count as immediate sensory experience, and

why. "Why" is the operative part of the question. The champions of atomic sense data were seeking the unscientific raw materials from which natural science is made, but in so doing they were being guided, all unawares, by an old discovery that was the work of natural science itself. That discovery, unnamed and undated, had dawned gradually, but it is a familiar discovery once we state it. It is the discovery that all our information about the external world reaches us through the impact of external forces on our sensory surfaces. There is no telepathy, clairvoyance, revelation, or extrasensory perception. This is a scientific finding, open, as usual, to reconsideration in the light of new evidence.

This scientific finding was the reason, all unrecognized, for conceiving of the epistemic given as comprised of atomic sense data. These were the subjective elements closest causally to the impacts of external forces on our sensory surfaces. The subject was perhaps unaware of the separate elements, as the Gestalt faction complained, but that was beside the point.

The sensory atomist was motivated, I say, by his appreciation that any information about the world is channeled to us through the sensory surfaces of our bodies; but this motivation remained obscure to him. It was obscured by his concern to justify our knowledge of the external world. The justification would be vitiated by circularity if sensory surfaces and external impacts on nerve endings had to be appealed to at the outset of the justification.

There is much clarity to be gained by dropping the project of justifying our knowledge of the external world but continuing to investigate the relation of that knowledge to its sensory evidence. Obscurity about the nature of the given, or epistemic priority, is then dissipated by talking frankly of the triggering of nerve endings. We then find ourselves engaged in an internal question within the framework of natural science. There are these impacts of molecules and light rays upon our sensory receptors, and there is all this output on our part of scientific discourse about sticks, stones, planets, numbers, molecules, light rays, and, indeed, sensory receptors; and then we pose the problem of linking that input causally and logically to that output. It smacks of psychology, yes; but it embraces the logic of science too, the whole problem of scientific evidence, the relation of observation to theory. In pursuing it we might consider positing sense data as auxiliary theoretical objects, but I think they would only be in the way.

So what we now face, in what I call naturalized epistemology, is the

problem of relating certain natural events, namely the triggering of our nerve endings, with other natural events, namely our word-mongering *about* nature. Neural events on the one hand, vocal events on the other.

Impacts of molecules and light rays upon our nerve endings set off neural impulses that travel to the brain, where they get processed in extravagantly complex and unexpected ways that the neurologists David Hubel, Torsten Wiesel, and their successors will be continuing to investigate for the foreseeable future. Stimulations and neural impulses build up speech habits and other habits in us that continue to be brought into play by subsequent sensory stimulation, through elaborate processes of association or reinforcement.

Among the myriad manifestations of these largely unfathomed processes, one outcome is a neat correlation between certain ranges of sensory stimulation and certain sentences. Exploiting these correlations, we can happily leave all the bewildering intervening processes to the neurologists, interested though we are bound to be in what they find out about them. We can limit our consideration of neural events to the superficial ones that answer directly to the impacts and input from the external world. Enabled as we are to relate them directly to certain vocal events, we then proceed to relate these in turn to the rest of the vocal or literary events that make up cognitive discourse. These latter connections can be drawn in broadly logical rather than neurological terms.

Of the sentences that correspond thus directly to certain ranges of sensory stimulation, one example is "It's raining". There are various concurrent triggerings of sensory receptors that warrant an affirmative or negative verdict on this sentence outright, at the time, and regardless of what we may have been thinking about or experiencing beforehand.

Thus consider, in contrast, the sentence "Inflation continues". An affirmative verdict on this sentence could be prompted by a current stimulation, perhaps a glimpse at the price of razor blades a few days before. Again an affirmative verdict on this sentence might be elicited by someone's query independently of other present sensory stimulation. The assent could be based on memory of a recent rise in interest rates, or something else that one remembered from the newspaper.

Sentences like "It's raining", for which the concurrent sensory stimulation suffices, I call observation sentences. They are not sentences about sense data, nor about stimulation, any more than "Inflation continues" is.

An observation sentence can be about anything, so long as our concurrent sensory stimulation regularly suffices to settle our verdict upon it.

Another typical observation sentence is “There goes a hyperthyroid”, for the experienced physician who has learned to recognize the signs of hyperthyroidism at a glance. One may have learned the words “It’s raining” by direct conditioning to concurrent sensory stimulation of appropriate sorts, whereas the physician may have learned the term “hyperthyroid” from descriptions in the books; but they qualify equally as observation sentences, as I use the term, while “Inflation continues” does not qualify. What matters is just the dependable association with purely concurrent stimulation, however that association may have been acquired.

Such are an individual speaker’s observation sentences. What qualify as such will vary with the individual’s experience and training, and also with his readiness to jump to conclusions. What finally count as evidence for scientific theory, however, are observation sentences in a socialized sense, relativized to the whole pertinent scientific community; for it is essential to the objectivity of science that all parties be able to agree on the data. A sentence is observational for the community if it is observational for every member.

Our project was to examine the relation of sensory stimulation to the sentences of scientific theory. The relation divides into two steps, and we have now dealt with one of them as such as we need to: the relation of sensory stimulation to an associated observation sentence. The next step is to relate observation sentences to the sentences of scientific theory. We are faced here with two radically different kinds of sentences. The observation sentences are *occasion* sentences: true on some occasions, false on others. Sometimes it is raining, sometimes not. On the other hand scientific theory is couched in *standing* sentences, true once for all if true at all.

Consequently the second step that we are faced with, that of relating observation sentences to theoretical ones, divides into two steps in turn. The first of these steps relates the observation sentences to a first crude species of standing sentences, which I call *observation categoricals*: and the remaining step relates these in turn to the sentences of scientific theory generally.

An observation categorical is a generalization built onto observation sentences, to say that fulfillment of the one observation sentence is invariably attended with fulfillment of the other. Examples: “Whenever it rains, it pours”. “Wherever there’s smoke, there’s flame”. “Whenever the cat’s away, the mice play”. “Whenever the sun rises, the birds sing”. Some of

them are true and some false. Most of the simple ones, such as these, are false, but they can still serve as illustrations.

The observation sentences, we saw, were occasion sentences directly linked to sensory stimulation. The observation categoricals, now, are standing sentences directly linked to observation sentences. In the observation categoricals we are already embarked upon science, in its most rudimentary form. They remind one of what used to be called natural history, in which observable traits of plants and animals were reported and correlate. But the observation categoricals are much more limited in scope even than what goes into natural history, let alone the more theoretical reaches of science: for the observations that are paired in an observation categorical have to occur together, on the same occasion. For the spanning of any appreciable distance or lapse of time between observations, we need scientific theory a cut above the observation categoricals.

Still, the observation categorical already embodies the rudiments of scientific method. Belief in an observational categorical, such as that when the sun rises the birds sing, is engendered by *induction*: we expect recurrent events to have recurrent concomitants: the first of the two observation sentences in the categorical states the experimental condition, and the other states the predicted outcome. The condition of the experiment is that we be among birds at sunrise; the predicted outcome of the experiment is that they sing. The observation categorical, "When the sun rises the birds sing", is our hypothetical law of nature. If they fail to sing, the law is refuted. If they sing, the law remains in good standing pending repetitions of the experiment.

Commonly, as in this example, two or more conditions are wanted in the first part of an observation categorical, rather than just a single condition. Instead of "When the sun rises the birds sing" we want, more strictly, "When the sun rises and birds are about, they sing". But this is not really a departure, since a string of observation sentences constitutes a single conjunctive observation sentence.

The observation categorical is the prototype not only of the experimental situation, but also of our daily pattern of expectation and fulfillment or disappointment. We get into some situation where we expect some particular event, and we check for it.

We have now looked at two of the three steps into which I have segmented the relation of evidence to theory. The first step was the relation of sensory stimulations to observation sentences, and the second was the re-

lation of observation sentences to observation categoricals¹. I turn now to the third, the relation of observation categoricals to other standing sentences at more sophisticated levels. The detail here is enormously complex, but it can be encompassed in a word: implication. The statements of science and daily cognitive discourse owe their evidence, their empirical content, to the observation categoricals that they imply. Our web of belief, scientific and otherwise, is a growing and ever changing structure of hypotheses, some deliberately framed and some a primordial heritage, which together imply a multitude of unstated observation categoricals.

The relation of theory to evidence is notably asymmetrical. An observation categorical can be refuted, we saw, by a pair of observations, affirmative and negative: the sun rises but the birds don't sing. On the other hand an observation categorical is not conclusively established by pairs of affirmative observation sentences; it can only be retained on sufferance, in the hope that it will withstand refutation. Now the same asymmetry carries over to science generally. A theory can be refuted by refuting a observation categorical that it implies, but there is no similarly conclusive proof of a scientific theory. The scientist is well portrayed by Sir Karl Popper as painstakingly inventing a hypothesis and then doing his best to refute it by cunningly contrived experiments. The more he is frustrated in these efforts at destruction, the more confident he becomes of his hypothesis.

Jumping to conclusions is our way of daily life. Jumping to reasonable conclusions is the busy scientist's way, and more power to him. A consequence is that a strict observation categorical is seldom detectable in the scientist's reasoning. Certainly there is no thought of the requirement that the paired observations occur together. But if scientists were to challenge one another compulsively down to the bitter end, rather than merely seeking a shared understanding, then genuine observation categoricals would emerge as the final arbiters. It is a matter of the theory, and not the practice, of theorizing.

Given then that scientific theory implies observation categoricals, what do we make of implication? Primarily it is a matter of logical laws which we internalize in the course of learning to use the connectives "and" and "or" and other logical particles. Thus we learn "or" mainly by learning to assent to a compound " p or q " whenever we feel warranted in assenting to

1. The published paper inadvertently included an additional sentence, which has been deleted here.

one of its components, and to dissent from the compound whenever prepared to dissent from both components. In learning to use “or”, therefore, we have learned four implications: that “ p ” implies “ p or q ”, that “neither p nor q ” implies “not p ”, and two more. In learning to use “and”, similarly, we learn that “ p and q ” implies “ p ”, that “not p ” implies “not both p and q ”, and so on. In learning the idiom of existential quantification, “Something is thus and so”, we learn that it is implied by any corresponding singular sentence, “Such and such is thus and so”. Correspondingly for universal quantification, “Everything is thus and so”. Besides these implications of a purely logical character, there will be countless interconnections of sentences resulting from truths of mathematics or merely from the subsumption of terms one under another, as ‘dog’ is subsumed under ‘animal’ or ‘gravity’ under ‘force’. These further connections can be brought somewhat under control by just adhering to the purely logical relation of implication and then reckoning the truths of mathematics and the various subsumptions rather as theoretical sentences on a par with all the others which jointly do the implying.

In any event the network of implication relating the theoretical sentences to observation categoricals is overwhelmingly complex and difficult to sort out. Part of the trouble is a fluidity in the rubric of observation sentences: highly theoretical matters come to qualify every now and then as observation sentences, in my sense of the term, through the progress of science.

I have been relating observation sentences to observation categoricals, and observation categoricals to more highly theoretical sentences. It has been a question of sentences, not of terms of objects. The observation sentence as a whole is keyed to some range of neural stimulations for assent or dissent. Likewise at the other extreme, the theoretical sentence as a whole awaits evidence of its truth or falsity. Terms of any and every sort do lurk inside the observation sentences and the theoretical sentences, but I have made no capital of them. Not, that is, until just now in discussing the implication relation. The idioms of quantification appeal expressly to objects, and the associated singular sentences hinge on singular terms that designate objects.

Evidently then, reference to objects, which has always seemed so central to science and knowledge, plays only a subordinate and auxiliary role in the structure of scientific theory. Just what *is* its contribution? Why posit objects at all?

I think its function is a tightening of connections in the building of compound sentences. I shall develop an example.

Our paradigmatic observation sentence has been “It’s raining”, which does not contain references to anything. Let us likewise rid the observation sentence “There’s a cat” of any hint of objective reference, by rendering it as “It’s catting”. But now consider the complex observation sentence:

(1) A white cat is facing a dog and bristling.

We cannot render this as mere conjunction of four, to the effect that it’s catting and there’s white and there’s bristling and there’s dog-facing. This conjunction is too loose; it tells us only that the four things are going on in the same scene. We want them all in the same part of the scene, superimposed. It is this tightening that is achieved by subjecting the fourfold conjunction to existential quantification. thus:

Something is catting and is white and is dog-facing and is bristling,
which is to say (1). An object has been posited, a cat.

For purposes of this context, a cat of the moment would suffice; no need of an enduring cat. To illustrate the need of an enduring cat I must go beyond observation sentences and suppose that we have somehow worked our way far enough up into scientific theory to treat of time: earlier and later. Suppose then we want to convey this thought:

If a cat eats a spoiled fish and sickens, then she will thereafter avoid fish.

We cannot treat this as a simple “if-then” compound of two self-sufficient component sentences. Like the “and” of the preceding example, the “if-then” connection is too weak. It has to be the same cat in both sentences, and hence an enduring cat. Our sentence is really a universally quantified conditional:

Everything is such that if it is a cat and it eats a spoiled fish and it sickens then it will thereafter avoid fish.

I think these two crude examples may be illustrative of the logical utility of reification across the board. The point is in keeping with what I have elsewhere called ontological relativity, according to which the objects of a scientific theory can be supplanted one-to-one by any arbitrary others without disturbing the support of the theory in sensory evidence. Objects

contribute the structure only as neutral nodes, and it does not matter what they are.

So much, then, for terms and their objects. Let us get back to sentences and their evidence.

It is unusual for a sentence either of science or of common sense to imply an observation categorical single-handed. Unusual but not impossible. Thomas Tymoczko pointed out to me that the observation categorical “Where there are five apples there are three” is implied by the theoretical sentence “ $5 > 3$ ” of pure arithmetic. For the most part, however, the implying of an observation categorical is a joint enterprise. Thus it is that the failure of a predicted observation typically refutes no single hypothesis, but only a substantial body of theory. It tells us that one or another statement in that body of theory must be revoked, and it leaves us to choose which. The point was stressed long ago by Pierre Duhem, and is called Duhemian holism.

Scientists do persistently speak of testing and refuting single hypotheses, and we can see why. They target some one dubious hypothesis before the experiment. It, and not another, shall be the one to fall if the implied observation categorical turns out false.

Typically, the body of theoretical sentences that together implied some failed prediction will have included not only sentences of physics, biology, or whatever, but also some truths of arithmetic. Free as we are to choose which of the implying body of sentences to revoke, we are free to *spare* arithmetic; and spare it we will, for an upset of arithmetic would reverberate through all branches of science. Scientists sensibly observe a maxim of minimum mutilation. This, I say, explains the quality of necessity that is felt to attach to arithmetical truth; and the same holds for other mathematical truths and for logical truths. The necessity of these truths lies simply in our freedom of choice; we are free to safeguard those tenets and to revise others instead, when revision is called for.

The concern to account for the necessity of mathematical and logical truth drove Carnap to a tenacious defense of the slippery notion of meaning; for he wanted to explain those truths as true by virtue of the meaning of the logical and mathematical vocabulary. We see now that an appreciation of Duhemian holism makes that expedient superfluous.

Another notable consequence of Duhemian holism, noted by Dagfinn Føllesdal, is that science is drastically underdetermined by all possible observation. Or, bypassing the dubious notion of possible observation, we

can put the point thus: scientific systems of the world can differ profoundly and still be empirically equivalent, in the sense of implying all the same observation categoricals. This is clear from our freedom of choice in revoking one scientific hypothesis or another when an observation categorical goes wrong.

Can a good empiricist accept an invidious distinction between two empirically equivalent theories, declaring one of them true and the other false? Evidently he is bound to. Recalcitrant observations can be accommodated by unlike choices of what theoretical sentences to retract, and these unlike choices issue in incompatible theories, only one of which can be true.

There is a simple expedient, pointed out to me by Davidson, for resolving such contradictions between empirically equivalent theories. If a sentence is affirmed in the one theory and denied in the other, it will surely contain a term that is only loosely linked to observable criteria; for the two theories, after all, are empirically equivalent. Let us then exploit that slack by treating the term as two different terms, even changing its spelling in one of the theories. The sentence that was affirmed and denied gives way thus to two independent sentences. Pursuing this process, we can reconcile the two theories.

Can we accept the competing theories as simultaneously true, when thus reconciled? It would gratify our empiricist sensibilities to do so, and there was a period a few years ago when I thought we could. Why not just accept them jointly as a single tandem theory? I call this the *ecumenical* line.

Roger Gibson and Dagfinn Føllesdal favored my earlier *sectarian* line of reckoning one's own theory as true and the empirically equivalent alternatives as false. I think they are right, and I have reverted to my earlier sectarianism. I shall explain my reasoning, which I expect is much the same as theirs.

We are imagining a tandem theory consisting of two overlapping lobes. The one lobe is our original theory and the other is the rival theory, empirically equivalent to the first and rendered logically compatible by Davidson's device of construing some crucial theoretical terms as novel homonyms. The sentences that lack these novel terms present no problem; they either are already affirmed in the original lobe or can be happily added, for they treat of the same matters without contradiction. But the ones with the irreducible new terms constitute a gratuitous annex to the original theory. They add no new coverage of observation categoricals,

since the two theories in their entirety were empirically equivalent; yet they treat, ostensibly, of irreducibly new matters. It is as if some scientifically undigested terms of metaphysics or religion, say “essence” or “grace” or “Nirvana”, were admitted into science along with all their pertinent doctrine and tolerated on the ground merely that they contravened no observations. It would be an acquiescence in what Herbert Feigl called *danglers*. It would be an abandonment of the scientists’ quest for economy and of the empiricists’ standard of meaningfulness.

Let us come to terms, then, with the sectarian position. We know from the Duhemian consideration that there are bound to be alternative systems of the world, empirically equivalent to ours but logically incompatible with it. We know of some examples that present no problem, for we can convert them into our own by a systematic redefinition of terms. I think of Poincaré’s familiar example of a finite space in which the bodies that we call rigid shrink uniformly on moving away from the center; we know how to translate all this into our usual theory of our infinite space and properly rigid bodies.

But imagine rather a theory that we cannot reinterpret into our own but which we still have reason to believe is empirically equivalent to ours. Still, inconsistent as it is with our own, we must declare it false. Or, if we reconcile it by Davidson’s method, we must view it as inscrutable, for it is no longer in our language and we have no acceptable translation. But we must still recognize it as well *warranted*, if we are aware of its empirical equivalence to ours. We may even oscillate between it and ours, for the sake of an enriched perspective on nature. But whichever theory we are working in is the one for us to count as true, there being no wider frame of reference.

Panel on Reference

1986

Introduction

Suppose a strange bright form appears, sharply outlined in the dusk. A man points at it and says 'Look at that.' The apparition shimmers briefly and vanishes. The man was using his demonstrative pronoun to refer to the apparition. Here we have reference at its simplest and most conclusive. In other cases, lacking the right contrast and sharp outline, there can be doubt as to how much of the scene the pointing was meant to include. In still other cases, not so sudden and evanescent, there can be further doubt as to whether a mere momentary phase or rather some enduring object was being referred to, and, if the latter, there can be some doubt as to how the object should be extrapolated into the past and future. Faced with indeterminacies of these kinds, we have to resort to general terms to fix the intended references, even though the object be right there, pointed out, and referred to by means of a demonstrative. When I point and say 'This is Kirmani,' it is not clear whether I am referring to the man and giving his name or referring to his garment and identifying the fabric. I resolve the ambiguity by inserting a general term, saying 'This man is Kirmani,' or 'This caftan is Kirmani,' as the case may be.

However, when we resort thus to general terms a new kind of problem arises, and it again is one that can arise still in the presence of the demonstrative pronoun and the pointing finger. It is the problem of false description, the misapplication of the general term. I say, 'This caftan is Kirmani,' but I am wrong about its being a caftan; it is a burnoose.

When we move from demonstratives and ostension into more shadowy realms, these casualties of reference mount apace. In some cases, unlike my over-simple one of the Kirmani caftan, there may really be an object that

would fit the speaker's description, but it is not the object he meant. Shall he be said to have really referred to it without meaning to, or shall he be said to have referred to the one he meant to refer to? Or, finally, shall all reference by misdescription be declared void, and the containing sentence be declared void of truth value? Sir Peter Strawson will talk about that.

There is also an opposite problem, which goes back to Frege's example of the Morning Star and the Evening Star and probably on back to the Middle Ages. The Morning Star is the planet Venus, and so is the Evening Star; this we know. There is no question now of false description, no question of misapplication of terms. In our use of these singular terms we are referring to one and the same planet; all is evidently in order. If we put the one term for the other in a true sentence, the result should still be true; it says the same thing about the same thing. Yet, notoriously, this doesn't work. The sentence 'The Sumerians failed to identify the Morning Star with the Evening Star' is presumably true, but the substitution gives 'The Sumerians failed to identify the Evening Star with the Evening Star,' which would seem to be false.

Circumstances of this kind have led Professors Føllesdal and Kripke to postulate a special kind of singular term called a *genuine* singular term or *rigid designator*, which is immune to this anomaly. But just how are we to recognize such a term? Intuitively the idea would seem to be that a rigid designator picks out its object by its *essential* traits, but essence is hardly a notion to rest with.

Moreover, the problem of interchangeability that was illustrated by the Morning Star and Evening Star is a problem that arises in contexts of significantly unlike sorts. It can arise in epistemic contexts, as in my case of the Sumerians, and it can arise in connection with the logical modalities: necessity, possibility. It appears that a demarcation of the category of rigid designators that is narrow enough for modal contexts will still be too broad for epistemic ones.

Professor Føllesdal tells me that he will talk about rigid designators and that he will indicate a way of looking at them that is different from Kripke's.

There are some unsettling further thoughts about reference, hinging on what I have called proxy functions. Thus think of a scientific theory, purporting to treat of some domain of objects. Now choose a one-to-one mapping, any one you like, that carries each of these objects into some other. I call the mapping a proxy function. Next let us reconstrue each term

of the theory, coordinating it with the proxy function. That is, where f is the proxy function, let us reinterpret 'dog' to mean ' f of a dog'; and correspondingly for every term. If along with transforming the ontology we reinterpret each term in this way, all statements will keep their old truth values. Moreover, all empirical support for the theory will remain undisturbed. For, consider the nature of empirical support. It resides ultimately in sensory stimulation that has been associated with appropriate sentences, appropriate strings of sounds; and those strings of sounds remain unchanged, down to the last phoneme, under our shift of ontology and our reinterpretation of terms.

It thus emerges that what matters for science is not ontology, but structure, together with sensory stimulation. The purported objects serve merely as neutral nodes in the structure. There is no real difference between holding one ontology and holding another on which it can be mapped. When we say of a Francophone that he uses the word 'pierre' to refer to stones, we are only saying how we are pairing his nodes with ours: we are translating his word 'pierre' by our word 'stone.' When we say that *we* use *our* word 'stone' to refer to stones, we are sounding an empty tautology.

Indeterminacy of Translation Again

1987

Over twenty-five years my *Word and Object* has sparked frequent criticism of my thesis of indeterminacy of translation and occasional evidence of misunderstanding, to which I have responded only sporadically and in scattered places. Burton Dreben has now drawn me back into fruitful discussions of the subject and urged me to undertake a succinct over-all clarification.

Critics have said that the thesis is a consequence of my behaviorism. Some have said that it is a *reductio ad absurdum* of my behaviorism. I disagree with this second point, but I agree with the first. I hold further that the behaviorist approach is mandatory. In psychology one may or may not be a behaviorist, but in linguistics one has no choice. Each of us learns his language by observing other people's verbal behavior and having his own faltering verbal behavior observed and reinforced or corrected by others. We depend strictly on overt behavior in observable situations. As long as our command of our language fits all external checkpoints, where our utterance or our reaction to someone's utterance can be appraised in the light of some shared situation, so long all is well. Our mental life between checkpoints is indifferent to our rating as a master of the language.

There is nothing in linguistic meaning, then, beyond what is to be gleaned from overt behavior in observable circumstances. In order to exhibit these limitations, I propounded the thought experiment of radical translation. The "source language", as the jargon has it, is Jungle; the "target language" is English. Jungle is inaccessible through any known languages as way stations, so our only data are native utterances and their concurrent observable circumstances. It is a meager basis, but the native speaker himself has had no other.

Our linguist would construct his manual of translation by conjectural extrapolation of such data, but the confirmations would be sparse. Usually the concurrent publicly observable situation does not enable us to predict what a speaker even of our own language will say, for utterances commonly bear little relevance to the circumstances outwardly observable at the time; there are ongoing projects and unshared past experiences. It is only thus, indeed, that language serves any useful communicative purpose; predicted utterances convey no news.

There are sentences, however, that do hinge pretty strictly on the concurrent publicly observable situation: sentences like 'It's raining' or 'That's a rabbit', which I call *observation sentences*. Jungle sentences of this sort are our linguist's entering wedge. He tentatively associates a native's utterance with the observed concurrent situation, hoping that it might be simply an observation sentence linked to that situation. To check this he takes the initiative, when the situation recurs, and volunteers the sentence himself for the native's assent or dissent.

This expedient of query and assent or dissent embodies, in miniature, the advantage of an experimental science such as physics over a purely observational science such as astronomy. To apply it the linguist must be able to recognize, if only conjecturally, the signs of assent and dissent in Jungle society. If he is wrong in guessing those signs his further research will languish and he will try again. But there is a good deal to go on in identifying those signs. For one thing, a speaker will assent to an utterance in any circumstance in which he would volunteer it.

Our linguist then goes on tentatively identifying and translating observation sentences. Some of them are perhaps compounded of others of them, in ways hinting of our logical particles 'and', 'or', 'but', 'not'. By collating the situations that command the native's assent to the compounds with the situations that command assent to the components, and similarly for dissent, the linguist may get a plausible line on such connectives.

Unlike observation sentences, most utterances resist correlation with concurrent stimulations that the linguist can share. Taking the initiative he may volunteer and query such a sentence for assent or dissent in various situations, but no correlation with concurrent stimulation is forthcoming. What next?

He can keep a record of these unconstrued sentences and dissect them. Some of the segments will have occurred also in the already construed observation sentences. He will treat them as words, and try pairing them off

with English expressions in ways suggested by those observation sentences. Such are what I have called analytical hypotheses. There is guesswork here, and more extravagant guesswork to follow. The linguist will turn to the unconstrued, nonobservational sentences in which these same words occurred and he will project conjectural interpretations of some of those sentences on the strength of these sporadic fragments. He will accumulate a tentative Jungle vocabulary, with English translations, and a tentative apparatus of grammatical constructions. Recursion then sets in, determining tentative translations of a potential infinity of sentence. Our linguist keeps testing his system for its efficacy in dealing with natives, and he goes on tinkering with it and guessing again. The routine of query and assent that had been his standby in construing observation sentences continues to be invaluable at these higher and more conjectural levels.

Clearly the task is formidable and the freedom for conjecture is enormous. Radical translation is avoided in practice by finding someone who can interpret the language, however haltingly, into a somewhat familiar one. But it is only radical translation that exposes the poverty of ultimate data for the identification of meanings.

Let us consider, then, what constraints our radical translator can bring to bear to help guide his conjectures. Continuity is helpful: successive utterances may be expected to have some bearing on one another. When several such have been tentatively interpreted, moreover, their interrelation itself may suggest the translation of a linking word that will be helpful in spotting similar connections elsewhere.

The translator will depend early and late on psychological conjectures as to what the native is likely to believe. This policy already governed his translations of observation sentences. It will continue to operate beyond the observational level, deterring him from translating a native assertion into too glaring a falsehood. He will favor translations that ascribe beliefs to the native that stand to reason or are consonant with the native's observed way of life. But he will not cultivate these values at the cost of unduly complicating the structure to be imparted to the native's grammar and semantics, for this again would be bad psychology; the language must have been simple enough for acquisition by the natives, whose minds, failing evidence to the contrary, are presumed to be pretty much like our own. Practical psychology is what sustains our radical translator all along the way, and the method of his psychology is empathy; he imagines himself in the native's situation as best he can.

Our radical translator would put his developing manual of translation continually to use, and go on revising it in the light of his successes and failures of communication. And wherein do these successes and failures consist, or how are they to be recognized? Successful negotiation with natives is taken as evidence that the manual is progressing well. Smooth conversation is further favorable evidence. Reactions of astonishment or bewilderment on a native's part, or seemingly irrelevant responses, tend to suggest that the manual has gone wrong.

We readily imagine the translator's ups and downs. Perhaps he has tentatively translated two native sentences into English ones that are akin to each other in some semantic way, and he finds this same kinship reflected in a native's use of the two native sentences. This encourages him in his pair of tentative translations. So he goes on blithely supposing that he is communicating, only to be caught up short. This may persuade him that his pair of translations was wrong after all. He wonders how far back, in the smooth-flowing antecedent conversation, he got off the beam.

Considerations of the sort we have been surveying are all that the radical translator has to go on. This is not because the meanings of sentences are elusive or inscrutable; it is because there is nothing to them, beyond what these fumbling procedures can come up with. Nor is there hope even of codifying these procedures and then *defining* what counts as translation by citing the procedures; for the procedures involve weighing incommensurable values. How much grotesqueness may we allow to the native's beliefs, for instance, in order to avoid how much grotesqueness in his grammar or semantics?

These reflections leave us little reason to expect that two radical translators, working independently on Jungle, would come out with manuals acceptable to both. Their manuals might be indistinguishable in terms of any native behavior that they gave reason to expect, and yet each manual might prescribe some translations that the other translator would reject. Such is the thesis of indeterminacy of translation.

My much-discussed example *gavagai*¹ illustrates the indeterminacy of translation only of terms, not of sentences. When seen rather as an observation sentence, *Gavagai* is keyed directly to determinate stimulus situations that are open to empirical investigation and afford the firmest of

1. *Word and Object* (Cambridge, MA: MIT Press, 1960), 29–45.

checkpoints. Likewise my example of Japanese classifiers² is a matter only of terms and not of sentences. More extravagant examples, having to do still with terms, are afforded by proxy functions.³ But my thesis of indeterminacy of translation applies first and foremost to sentences, holophrastically conceived; and this I despair of documenting. Radical translation is a near miracle, and it is not going to be done twice to the same language. But surely, when we reflect on the limits of possible data for radical translation, the indeterminacy is not to be doubted.

The point of my thought experiment in radical translations was philosophical: a critique of the uncritical notion of meanings and, therewith, of introspective semantics. I was concerned to expose its empirical limits. A sentence has a meaning, people thought, and another sentence is its translation if it has the same meaning. This, we see, will not do.

The critique of meaning leveled by my thesis of indeterminacy of translation is meant to clear away misconceptions, but the result is not nihilism. Translation remains, and is indispensable. Indeterminacy means not that there is no acceptable translation, but that there are many. A good manual of translation fits all checkpoints of verbal behavior, and what does not surface at any checkpoint can do no harm.

Nor, in scouting the old notion of meanings of words and sentences, do I repudiate semantics. There is much useful work, done and to be done, regarding the manner and circumstances of the use of words. Lexicography is its conspicuous manifestation, and there is much scope also for refinements of semantic theory. But I would not seek a scientific rehabilitation of something like the old notion of separate and distinct meanings; that notion is better seen as a stumbling block cleared away. In later years indeed it has been more of a stumbling block for philosophers than for scientific linguists, who, understandably, have simply found it not technically useful.

Some of my readers have had difficulty in seeing how the thesis of indeterminacy of translation is anything but a special case of the thesis that natural science is underdetermined by all possible observation. This, in turn, suggested⁴ Pierre Duhem's perception that when we revise a theory

2. *Ontological Relativity and Other Essays* (New York: Columbia, 1969), 35–38.

3. *Theories and Things* (Cambridge, MA: Harvard, 1981), 19–22.

4. Quine changed the published “is a corollary of” to “suggested” on his copy of the printed paper.

in the light of a recalcitrant observation we are free to choose which component sentences of the theory to revoke.

The indeterminacy of translation differs from the underdetermination of science in that there is only the natives' verbal behavior for the manuals of translation to be right or wrong about; no claims are laid regarding hidden neural mechanisms. If translators disagree on the translation of a Jungle sentence but no behavior on the part of the Jungle people could bear on the disagreement, then there is simply no fact of the matter. In the case of natural science, on the other hand, there is a fact of the matter, even if all possible observations are insufficient to reveal it uniquely. The facts of nature outrun our theories as well as all possible observations, whereas the traditional semantics outruns the facts of language.

In thus contrasting the underdetermination of natural science with the indeterminacy of translation I have taken a realistic view of nature, which indeed I hold. But I have elsewhere drawn the contrast without the realism, in the following way. Natural science, we again assume, is underdetermined by all possible observation. However suppose that we have settled for one of the many over-all theories of nature that fit all possible observation. Translation remains indeterminate, even relative to the chosen theory of nature. Thus the indeterminacy of translation is an indeterminacy additional to the underdetermination of nature.

In closing I want to add an unrelated remark on indeterminacy, in the light of recurring evidence of misinterpretation. My indeterminacy applies only to translation, and in no way to grammaticality. I have represented grammaticality as subject to rounding out and rounding off at its limits, but apart from that I see grammaticality, unlike translation, as adequately determined by behavioral dispositions. Structurally unlike codifications of grammaticality can indeed be equivalent in output, but that goes without saying. Incidentally, semantics likewise continues to be, I repeat,⁵ a vital domain of inquiry. What I have challenged is just an ill-conceived notion within traditional semantics, namely, sameness of meaning.

5. Quine changed the published "in my view" to "I repeat" on his copy of the printed paper.

Mind, Brain, and Behavior

1989

There is one time-honored philosophical issue that seems to me to have been put to rest; namely, the mind-body problem. For Descartes and others, mind and body were two substances. Their interaction posed a problem, in view of the physical law of conservation of energy—a problem also of the mechanism of interaction. Nowadays people of a scientific temper seem generally agreed on a monistic, materialistic stance: thinking, wishing, dreading, suffering, feeling are states or activities of the body. More specifically they are neural activities, or states of nerves. Patterns of behavior they are not, but physical they are. A mental state or event is one kind of bodily state or event.

Where behavior fits into the picture is as our evidence or outward sign of the mental state or event, which itself is neural. We need the behavioral description in order even to specify which neural state or event we are talking about; for we do not know enough about the neural mechanisms to be able to specify the state or event in neurological terms. The situation is like that of a disease whose microbe or mechanism has not yet been determined; we specify the disease by its outward manifestation, but the disease itself is not the manifestation.

Another and more time-honored way of specifying a mental state or event is by the traditional mentalistic terminology, which is keyed to introspection. Behaviorism shuns this alternative simply because it is undisciplined, lacking intersubjective controls. Ironically, though, the mentalistic idiom is itself behavioristic as we first acquire it. We learn our first mentalistic terms as behavioral terms—there is no other way. Thus consider the simplest case, such a word as “pain.” Pain is a private thing, certainly; I know, better than any one else, whether I am in pain. But how did I learn

what to call it? I learned the word from someone else, on an occasion when she could see by outward signs or circumstances that I was in pain. I associated the word with my pain, and continued to apply it to recurrences of that inner state, regardless of whether there were again any outward signs for other persons to see, but my acquisition of the word was dependent on outward signs. The use of the word is acquired in the light of behavior and then extrapolated by introspection.

So it is with our whole mentalistic lexicon—our words for pleasure, joy, grief, dread, and the rest. As a first move, someone must supply me with the word on the strength of my behavioral manifestations of my state of mind. Or in some cases my mentor may volunteer the word to describe her own state, and I may learn the word by noting the behavioral signs of her state. Either way, the word is learned through its correlation with the behavioral evidence for the mental state.

Given then its behavioristic underpinnings, what is wrong with the old mentalistic jargon? Simply its runaway introspective extrapolation. It gets out of hand, beyond intersubjective access.

Mental states and events are neural states and events, I have been saying, and behavioral description is just a means of specifying and spotting them [—an improvement, for this purpose, on the traditional mentalistic and introspective way of specifying and spotting them].¹ Is this all we can say for behaviorism, and must all theory be left to the neurologist?

Certainly not, as Skinner has long stressed. There is strong theoretical reason for shortcutting the neural wellsprings and seeking laws or uniformities within the behavioral level. The reason is that the uniformities and significant correlations in psychology have to do with the function or outcome of an animal's movements, and are insensitive to differences in the anatomical details of the movements as long as they serve the same end. Running the maze is the thing, and no matter which muscles are flexed in what order, or which neurons are fired to flex them. Attention to the anatomical detail of implementation would be counterproductive, obstructing the explanatory generalities and obscuring the woods with trees.

Another and riskier way of putting the matter is that psychological laws or generalities hinge on grouping the animal's movements according to goal or purpose rather than neuron and muscle fiber. Talk of purpose is a

1. This portion of the sentence appeared in the original published version but is marked for deletion in Quine's personal copy of the printed paper.

risky reversion to unregenerate mentalism, but Darwin has taught us how to meet that risk. At the genetic level—which is not my present concern—Darwin accounted for seemingly goal-directed anatomy and goal-directed instinct by straightforward causality in natural selection and the survival of the fittest. At the level of learning or habit formation, which is my present concern, there is an accommodation of goal or purpose quite parallel to Darwin's. It is the familiar process of reinforcement and feedback, as in military fire control; what Wiener called cybernetics.

A few minutes ago I was stressing that mental states and events are events and states of nerves. Now I am supporting Skinner's strategy of shunning the nerve net and resting with self-contained behavioral correlations in a cybernetic vein. Is there no reconciliation? Ultimately and in principle there surely is. What is to be shunned in the nerve nets is the innervation of muscle fibers. There are nerve centers in the cerebral cortex, however, that distribute the flow of neural impulses along functional or goal-directed lines such as were conditioned by past reinforcement and feedback. These are the mysterious centers that neurobiologists such as Hubel, Wiesel, the Churchlands, and Crick have only in recent years begun to explore. An eventual understanding of them will bring behavioral psychology and neurology effectively together. It is plausible that the autonomous findings of behavioral psychology might meanwhile afford some hints to the neurologists in their framing of hypotheses regarding the cortical centers. And the behaviorists might reasonably continue on their merrily autonomous way even after the cortical mechanisms are better understood—though welcoming any light that might reach them from above.

Before closing I want to touch again on unregenerate mentalistic talk, for there is more to say of it. I pointed out that it is acquired on a behavioristic basis. We noted further that even the seemingly irresponsible references to final cause or purpose are in large part legitimized along the quasi-Darwinian lines of cybernetics. But we noted also that mentalistic talk promptly escapes the discipline of its behavioristic beginnings, by uncontrolled extrapolation along introspective lines. Much the same is true, notoriously, of teleological talk of purpose: It far outruns what can be rationalized along cybernetic lines.

The five mentalistic words that I enumerated earlier—"pain," "pleasure," "joy," "grief," "dread"—are atypically well behaved. Despite their unchecked extrapolation along introspective lines, I expect that the neurology of the five states thus designated could be reasonably approximated with-

out having to master the subtleties of the most complex cortical centers. But the neurology of purposive action does involve those obscure centers, I suggested, and is a perhaps remote agendum for the future.

As for mentalistic *talk* of purpose, much of it exceeds the scope of all possible physicalistic discourse, either behavioral or neurological. Such is the way in general of mentalistic idioms of the so-called *intensional* kind. The ascriptions of purpose are examples, and further examples are “thinking of Vienna,” “waiting for Lefty,” “hoping to see a lion,” “wondering who wrote *Waverley*.” Commonly an intensional sentence will contain another sentence as a part, somewhat in the manner of indirect discourse; thus “He hopes that he will see a lion,” “He wonders whether Scott wrote *Waverley*.”

Philosophers from Franz Brentano to Roderick Chisholm and Donald Davidson have argued plausibly that the intensional idiom is mostly not translatable into the physicalistic. If the old mind-body problem has dissolved as a problem of the relation between two substances, it has left us with a problem of the relation between two idioms, two types of language: the mentalistic, particularly at its intensional extreme, and the physicalistic. We can agree that both idioms treat of people as physical objects, but the idioms are related to each other as oil and water.

As Davidson has argued,² the irreducibility of the intensional idiom hinges on the distinction between particular and general. If at some particular moment some particular man thinks about Vienna, that event is a particular event in his body, presumably in his brain cells, and nothing more. It could be described exhaustively in neurological terms, if all were known. But there is no *general* neurological description, known or unknown, nor any behavioral one, that would cover *all* and only the cases of thinking about Vienna. The phrase “thinking about Vienna” is a general term, grouping together under it a wide range of mental events; each of these mental events is a separate physical event, but no neurological or behavioral description, however general, covers exactly that same *range* of events.

We saw that mentalistic language is acquired behavioristically. This applies to the intensional idioms along with the rest. We saw also, more in detail, how some of the mentalistic terms would have been acquired; “pain” was the example. How might it have been with the intensional idioms?

Let us think about the cases that embed a sentence in a sentence, after

2. Donald Davidson, *Action and Events* (Oxford: Clarendon, 1980), 214–225.

the pattern of indirect discourse; for the other intensional idioms can in large part be paraphrased into such a form, as we saw in the case of "hoping to see a lion." Now there is a crucial step in the handing down of language that depends critically on a mastery of certain cases of this sentence-embedding form. Thus consider how the child might learn the sentence "The curtain is blowing." Having picked it up on one occasion, by mimicry perhaps, he tries applying it on a later occasion. His mother approves and reinforces the performance only if she sees that he is facing in the right direction and can observe that the curtain is blowing. She must be in a position to say, in effect, to herself anyway, the second-order sentence "He sees that the curtain is blowing," if she is to teach him the first-order sentence. Essentially this same situation constantly recurs, clearly, in handing language down the generations. An aptitude for intensional idioms of the sentence-embedding type must therefore be very nearly as old as language itself. These idioms would not be learned as early in life as some of the first-order sentences, but adults would have had to become familiar with some of them, or with the circumstances appropriate to them, by the time they were ready to foster language in their offspring.

Some of the intensional idioms under consideration are thus of a piece with the mechanism by which language itself survives. Further locutions along the same lines would then drift in easily by analogy. Quite apart from the handing down of language, moreover, mentalistic language in the intensional mold is clearly indispensable to daily intercourse. But it remains utterly unscientific for all that, and has no place in the literal and austere formulation of one's theory of the world. For this reason the distinctive systems of formal logic that have been devised to accommodate intensional idioms have had little appeal for logicians who, like me, are interested in logic only as an integral part of our theory of the world and not as part of a study of language. It is in banning the irreducibly intensional idioms from the apparatus of rigorous scientific description and explanation that the behaviorist discipline makes its most drastic inroads and renders perhaps its greatest service.

The Elusiveness of Reference

1990

People are said to refer to things. Alternatively, words and phrases are said to refer to things. These words or phrases may be proper names, singular descriptions, general terms, variables, or mere demonstrative pronouns. Reference is the topic of quite a variety of questions and puzzles, and I propose to survey them. Their variety corresponds in part to the variety of avenues of reference that I have just enumerated.

The speaker is the agent of reference, and the spoken word is his instrument. Either may be said to do the referring; there is no harmful ambiguity here, as long as we remember that the spoken word that counts is the token or event of utterance, and not the type. The same linguistic form can vary in reference from speaker to speaker and from occasion to occasion.

Perhaps the most primitive and least problematic case of reference is the case where the speaker points at a conspicuous, unfamiliar, and sharply outlined apparition of short duration and says: 'Look at that'. Each of these four traits is pertinent: conspicuous, unfamiliar, sharply outlined, fleeting. Failure of any of the four can raise one or another question. If the apparition is not conspicuous and sharply outlined, the hearer may wonder just how broad a portion of the scene the speaker meant to indicate. If it is not unfamiliar and of short duration, the hearer may wonder whether a mere momentary phase was intended or rather some continuing process or enduring object and, in that case, he may also wonder just how to extrapolate the object into the past or future: did the speaker mean just the tadpole or the eventual frog as well?

Faced with indeterminacies of these kinds, we have to resort to general terms to fix the intended reference, even though the object be right there, pointed out, and designated by a demonstrative. To resolve the ambiguity

regarding the tadpole I may say: 'This animal' or 'This tadpole' according as I mean to include the eventual frog in my reference or not.

Pointing without resort to general terms can lead to a new kind of problem even in the presence still of the demonstrative pronoun and the pointing finger. I point at a chapel and say 'This baptistery is Romanesque', and it is indeed Romanesque, but it turns out not to be a baptistery. Has the intended reference really taken place at all? Here perhaps we can draw a useful distinction after all between reference by speaker and reference by spoken word. We can say that I did indeed refer to the chapel in question, but that my words 'this baptistery' do not. Or we might prefer still to say that my utterance, 'this baptistery' as token, did refer still to the chapel in question, though the general term 'baptistery' does not refer to it; does not denote it. I have no preference to one rather than another here, and no clear preference as to whether my utterance 'This baptistery is Romanesque' should be rated as true, false, or neither. Rules to the purpose can be adopted by convention when they matter, if ever.

Thus far I have been considering cases where the intended object is present and pointed at and a demonstrative pronoun is used. When we move away from demonstratives and ostension into more shadowy realms, the casualties of reference mount apace. Where ostension is not feasible, the speaker may resort to a proper name or singular description; and in so doing he may go wrong in either of two ways. It may happen, as in the case of the name 'Pegasus' or the description 'the king of France', that there is no such thing, or it may happen, as in Keith Donnellan's example of the man with the martini, that the name or description designates something other than what the speaker intended. When there is no such object, shall we rule that the speaker's sentence is neither true nor false? And where there is an object, but an unintended one, shall we rule that it was really referred to, or that the intended one was referred to? Again, rules to the purpose can be adopted by convention when they matter, if ever.

There is also a very different but equally familiar problem, which goes back to Frege's example of the Morning Star and the Evening Star and probably on back to the Middle Ages. The Morning Star is the planet Venus, and so is the Evening Star. There is no question here of false description, no question of misapplication of terms. In our use of these singular terms we are referring to one and the same planet. If we put the one term for the other in a sentence, the result should still be true; it says the same thing about the same thing. Yet, notoriously, this doesn't work. The sen-

tence 'The Babylonians discovered that the Morning Star is the Evening Star' is presumably true, but the sentence 'The Babylonians discovered that the Morning Star is the Morning Star' is false. Evidently the terms 'Morning Star' and 'Evening Star' have not been used in these examples merely to refer to the planet.

A distinction thus emerges between occurrences of a term that are purely referential and occurrences that are not. The law of substitutivity of identity, or of putting equals for equals, holds in general only for purely referential occurrences. Similarly for the law of existential generalization, ' $Fa \rightarrow \exists xFx$ '. If ' Fa ' stands for a true sentence in which the term ' a ' serves merely to refer to an object ' a ' then surely ' $\exists xFx$ '. But if ' Fa ' stands for the true sentence 'There is no such thing as Pegasus', we cannot conclude that ' $\exists x$ '(there is no such thing as ' x ').

Or take Robert Sleight's example of the shortest spy. I believe in spies, and I believe there is a shortest; so I believe that the shortest spy is a spy. Yet it would seem wrong to infer that ' $\exists x$ '(I believe that ' x ' is a spy), for this suggests that I have the goods on someone. Evidently then 'the shortest spy' does not occur purely referentially in the context 'I believe that the shortest spy is a spy'.

I have been classifying occurrences of terms as purely referential and otherwise. Føllesdal phrases the matter rather as a classification of the terms themselves. A genuine singular term, as he calls it, serves simply to designate its object, and is unswerved by context, unswerved by changes in the state of the object or in our belief regarding it. Substitutivity of identity and existential generalization hold for such terms without restriction. Kripke speaks of such terms more suggestively as rigid designators.

How do we spot such a term? A designator is rigid, Kripke tells us, if it names the same object in all possible worlds in which the object exists. This definition estranges those of us who are put off by the mystique of possible worlds.

There is trouble also apart from that; namely, different sorts of examples call for different brands of possibility. Sleight's example hinged on failure of existential generalization of an epistemic context 'I believe so and so'. To qualify for existential generalization in such a context, the term 'the shortest spy' would have had to be *epistemically* rigid: it would have had to name the same person in all epistemically possible worlds. Similarly the example of the Morning Star and the Evening Star hinged on failure of interchangeability in an epistemic context: 'The Babylonians discovered so-and-so'. To

qualify for interchange in such contexts, two designators would need to be epistemically rigid: they should name the same object in all epistemically possible worlds. Another familiar example, the one about the number 9 and the number of the planets, hinges rather on a modal context, necessity: the true sentence '9 is necessarily square' is turned false by substituting 'the number of the planets'. Designations qualifying for interchange in such contexts should be logically or metaphysically rigid, naming the same object in all logically possible worlds. Other examples may call for still further brands of possibility.

The appeal to possible worlds can be avoided by appealing rather to essence. A designator is rigid if it connotes only essential properties of its object. However, the problem of multiple brands of possibility then remains with us still, in another guise: multiple brands of essence. Designators interchangeable in epistemic contexts should connote only epistemically essential properties of their shared object; correspondingly for modal contexts; and correspondingly for any other sort. I hardly need add that I am not happy anyway with the idea of essential property, or the idea of connoting.

Føllesdal is right, I think, in seeing no need of either sort of apparatus for classifying terms as rigid or otherwise. It is merely a matter, I presume, of how we choose to use a term: whether to use it rigidly or not: whether to use it *de re* or *de dicto*. The sentence 'The Babylonians discovered that the Morning Star is the Evening Star' is true if we take the terms unrigidly, or *de dicto*, but false if we take them rigidly, or *de re*; taking them rigidly amounts to saying that the Babylonians discovered that Venus is Venus.

I have long argued that for some purposes, both philosophical and logical, it is well to treat all proper names as singular descriptions. What had been an irreducibly proper name, say 'Socrates', gives way to an irreducibly proper predicate, 'is-Socrates', which is true of only the one individual; and the name is then derived from it as a description, ' $\lambda x(x \text{ is-Socrates})$ '. One may wonder, now, whether the idea of rigid designator conflicts with this plan, especially when one thinks in Føllesdal's terminology; for Føllesdal speaks of rigid designators simply as genuine singular terms, as opposed to descriptions.

But in fact there is surely no conflict here. We can preserve rigidity of the designator 'Socrates' even when reparsing it as a description, ' $\lambda x(x \text{ is-Socrates})$ '; we have merely to reckon 'is-Socrates' as rigid in turn, a rigid denotator. Rigidity makes just as much sense for predicates or general

terms as for singular terms, and regardless even of whether the general term is true of just one thing or many. Recurring to Kripke's possible worlds for the sake of argument, we could say that a general term is a rigid denotator if it is true of just the same objects in all possible worlds in which they exist.

Though rigidity makes equally good sense for general and singular terms, it is no accident that its usual application is to singular terms. The point of the notion has been to implement the substitutivity of identity and existential generalization, both of which have to do strictly with singular terms.

Turning my back on possible worlds and essences a few minutes ago, I said that I prefer Føllesdal's casual approach to the classification of terms as rigid and otherwise: it is just a matter of choosing whether to use a term rigidly or not, whether *de re* or *de dicto*. We can change our decision from occasion to occasion, using the term now *de re* and now *de dicto* in one and the same context. I illustrated this contrast a few minutes ago in connection with the Evening Star and the Morning Star, and it can be seen again in terms of the number of the planets. *De re*, the number of the planets is indeed necessarily square;¹ *de dicto* it is not.

Thus a term may have both rigid or *de re* use and *de dicto* use, and it may also have, in my old terminology, purely referential and not purely referential occurrences. How are these two dualities related? It comes to this: a not purely referential position is a position where rigidity could matter. It is in those positions, and only those, that there is a material difference between taking a term *de dicto* and taking it *de re*. In purely referential position the distinction between taking a term *de re* and taking it *de dicto* loses its point.

My first topic this evening was the miscarriage of reference. My second topic, which I have now nearly concluded, might be called impurity of reference: the breakdown of substitutivity and existential generalization in not purely referential positions. A term picks out its object by means of descriptive content or other connotations, even when the term does not have the form of a singular description, and this descriptive or connotational matter is what can block substitutivity and instantiation when the term is not in purely referential position. To treat a term *de re*, or as rigid, is to abstract from these descriptive or connotational trappings and thus insure

1. The number of planets was a square at the time this paper was written (although not at the time of this book's publication).

substitutivity and existential generalization even when the position is not purely referential.

This move, rigidification by abstraction from trappings, seems to me to be rather neatly realized by a French usage: and there is a similar situation in Japanese. My French example, which I have used elsewhere, is: *Ma pauvre tante, elle ne sort jamais*. 'My poor aunt, she never goes out'. The construction is inelegant in English, but seems to be preferred in French. When I say 'My poor aunt never goes out' I describe my aunt in three ways: she is pitiable, she is my aunt, and she never goes out. Two of these three characterizations, however, are posed in the subject term, and must be set aside if we want rigidity. In the French construction they are visibly set aside; the referential function of the subject term is isolated in the bare and neutral *elle*. The bare pronoun is the prototype of rigid reference, being wholly without descriptive trappings; all descriptive matter is relegated to the context. In logic the rigid vehicle corresponding to the pronoun is the bound variable.

I cannot call the variable a rigid designator, for it does not designate: it takes things as values. But rigid it is; it takes its values strictly *de re*, for it has no *dictum* to draw on *de dicto*. This is why the laws of substitutivity of identity and existential generalization stand forth as logical laws without restriction when expressed with variables:

$$\forall x \forall y (x = y \rightarrow Fx \leftrightarrow Fy),$$

$$\forall y (Fy \rightarrow \exists x Fx)$$

But we must beware of substituting singular terms for the universally quantified variables here, unless the position in '*Fx*' is purely referential or the substituted singular terms are rigid designators. This restriction on substitution reflects the fact that universal instantiation is dual to existential generalization, and subject to the same constraints: the position must be purely referential or the instancial term must be rigid.

I have talked of miscarriage of reference and of impurity of reference. Questions arise also about the channels of reference; for surface grammar is not of itself sufficient to determine whether an expression refers to anything, or even whether it purports to refer to anything. Ideally a singular term refers to an object, by designating it. Sometimes, as in the case of 'Pegasus', it purports to refer but fails, for lack of an object. And sometimes we hesitate to say of a singular term that it even purports to refer. We bandy such singular terms as 'the burden of proof' or 'the crux of the

problem' with no thought of an object of reference, not even a make-believe object like Pegasus.

One proposal might be that we recognize reference only to concrete things and dismiss any thought of reference to abstract objects. But this is too cheap and empty a victory for nominalism. Surely we do make some substantial distinction between discourse that scrupulously and effortfully adheres to a concrete ontology and discourse that unwittingly draws on a prodigal universe of sets or properties or other abstract objects. Where does the difference lie?

The answer lies, again, in the bound variable, the pure and unencumbered medium of rigid reference. The objects that we recognize as existing are all and only the objects, concrete or abstract, that we recognize as values of our variables. This is evident, indeed, from the verbal reading of our quantifiers: 'everything x is such that', 'something x is such that'.

A singular term, then, is reckoned as designating or referring to an object just in case it is substitutable for the variable for purposes of universal instantiation or existential generalization.

The objects assumed by a theory, according to this standard, are not ordinarily just the objects that are referred to by singular terms of the theory. The values of the variables may be expected to outrun the designation of the singular terms, and this is as it should be.

This standard of ontological commitment is valid only for a theory that is couched in the classical quantificational form of first-order predicate logic, without supplementary syntactical structure. For the structures could complicate the ontological accounting by raising problems of foreign exchange. Putnam has pointed out, for instance, that some of the power gained by adding higher-order sets to an ontology can be gained alternatively by admitting a modal operator.

Still the proposed standard carries over to theories of other logical form, so long as we are agreed on how to translate them into the form of first-order predicate logic. In this way it carries over to theories couched in what I call predicate-functor logic, which is devoid of variables. For theories in that form the result is simply that the ontology comprises all objects that fulfill predicates of the theory; in other words, all denotata of general terms.

The standard could be phrased in that way even for theories in classical quantificational form; it comes to the same thing as appealing to the values

of the variables. The version in terms of variables strikes me as more basic, however. The distinctive feature of a predicate is that it attaches to a variable to form an open sentence; and the variable is, we saw, the pure and unencumbered medium of rigid reference.

The pure and unencumbered use of the bound variable, in turn, is not in quantification, but in the 'such that' construction, to form a complex general term: 'x such that ... x ...'. It is here that the variable plays its purely combinatory role of permuting or linking various referential positions. The quantificational ideas of 'all' and 'some' are extraneous to the variable; they and other uses of bound variables can conveniently be rendered as parasitical on their basic use in 'such that' clauses.

The 'such that' clause in turn is just a stream-lined version of the relative clause, and the bound variable in the 'such that' clause answers to the relative pronoun and the pronouns of cross-reference to the relative pronoun. Accordingly I see the relative clause as the cradle of reference. Conception, no; birth, no; but cradle. Conception and birth are gradual and indecisive in the case of reference, but reference is in full swing with the emergence of the relative clause.

Again our criterion is parochial, as was the bound variable. Both criteria are extensible to other languages just insofar as we settle on translations of those languages into our own. I find this acceptable. If we cannot see how to translate our crucial idioms of reference into the foreign language, of course we cannot see what to count as reference in that language. The very notions of reference and ontology may be inappropriate to that language.

Some of my readers have found this disturbing. They have protested that the wildest of aliens, and even dumb animals, reify things. Animals react distinctively to particular things and kinds of things, thus recognizing their existence and incorporating them into their ontology. In a recent paper in the *Journal of Philosophy*² I recognized this want and proposed to accommodate it under the head not of ontology but of ideology, in a somewhat special and somewhat extended sense: *perceptual ideology*.

["If the subject is disposed to react to the presence of any and every cat in some manner in which he is not disposed to react to anything but cats, then I shall reckon the term 'cat' to his ideology, though it is strictly my

2. "Ontology and Ideology Revisited," *Journal of Philosophy* 80, no. 9 (September 1983): 499–502; reprinted in this volume.

term, with which he may or may not be conversant (. . .) Again if he is disposed to react to Geach's Tabitha as to no other cat or noncat, I shall reckon 'Tabitha' to his ideology, though he be ignorant of the name."³

We could even derive a notion of perceptual ontology independent of questions of values of bound variables. Along with reckoning the term 'cat' or 'Tabitha' to the creature's ideology, we could reckon all cats, or specifically Tabitha, to his perceptual ontology. But this would be idle; perceptual ideology already tells the story. It could be more usefully extended in another direction: to comprise not mere terms, such as 'cat' and 'Tabitha', but whole occasion sentences, such as 'It's raining', 'The sun is setting', 'The chill factor is below zero Fahrenheit'. Each such sentence belongs to the creature's perceptual ideology if he reacts distinctively in the described situation.

[I quote again from "Ontology and Ideology Revisited":

"For behavioral psychology and ethology we should look merely to perceptual ideology as the key idea, and for a properly philosophical theory of reference and ontology we should continue to look to the values of the variables. It is here that we can take account of the reification not only of perceptible things and situations but also of microphysical and abstract objects. It is here that we can make sense of ventures in ontological economy, and of the place of ontology in Tarski's recursive definition of satisfaction and truth." <See footnote 3.>]

I have argued that reference is at hand, full blown and unmistakable, only with mastery of the relative clause. I argued the point more fully in *The Roots of Reference*. The grammar of the relative clause is a sophisticated development, standing in no direct correlation with the sensory stimulation that is the ultimate source of evidence for scientific theories. Various of our sentences are conditioned to sensory stimulation in varying degrees, and it is through them that scientific discourse imbibes its sensory evidence. The sentences that are thus conditioned to stimulation are what I call observation sentences. They are conditioned as wholes. The referential status of their component words is indifferent to that connection. It hinges rather on the articulated structure of the eventual scientific theory. If we are to understand scientific evidence, as well as language learning, it is imperative that we see the interface between language and stimulation as con-

3. These paragraphs appeared in the original published version but are marked for deletion in Quine's personal copy of the printed paper.

sisting not of observation terms but of observation sentences, one-word sentences though many of them be. Terms and their objects are adjuncts of the elaborated superstructure.

Once we appreciate this, it becomes evident that all references can be revised and reshuffled at will without rewriting any of our science, falsifying any of its sentences, or disturbing any of its observational evidence. We have merely to reinterpret all terms, singular and general, by what I call a *proxy function*, namely, an arbitrary one-to-one transformation that carries each of the old objects into a new one.

My first topic was miscarriage of reference. My second was impurity of reference. My third was the channel of reference, which I identified with the bound variable, or pronoun. Now we are on the fourth: the inscrutability of reference. For science, for all knowledge and all evidence, any objects will do as well as any others. All we need are neutral values of variables, neutral nodes in a network of sentences that connects sensory stimulations with other, future sensory stimulations.

It has been a long-standing platitude that intension determines extension; that meaning determines reference. Terms with the same reference may have different meanings, as witness '9' and 'the number of the planets' <see footnote 1> or 'Scott' and 'the author of *Waverley*', but obviously, it was said, terms with the same meaning must have the same reference. How does that stand now? I have no definition of meaning, but whatever goes into meaning must be traceable ultimately to the associations of our linguistic forms with sensory stimulation and with one another. Proxy functions leave these connections unchanged while revising reference across the board. Yet everyone said that meaning determined reference. How can everyone have been so wrong? I think the solution is evident: meaning determines reference within each fixed ontology.

A more serious question is raised by my characterization of objects, two minutes ago, as neutral nodes in a network that connects stimulations with stimulations. Am I exempting the stimulations from this neutrality? Are they my real ontology?

The answer turns on a vital distinction between ontology and the epistemology of ontology. My position is naturalism. What there is, is a question of our overall scientific system of the world. The stimulations are indeed real denizens of my ontology, but so are sticks, stones, electrons, numbers, functions, and all the other values of the variables in a true account of how things are. Proxy functions figure only in another context, where we are

not saying what there is, but talking about our evidence for what there is. Proxy functions show that the ontology is underdetermined by evidence, but not that it is false. Ontological questions are to be answered within our evolving scientific system, and what proxy functions show is just that the answer must be in large part arbitrary.

Suppose then that someone propounds a theory of the world that is just like ours in structure and evidence, but disagrees in ontology, the two being reconcilable by applying a proxy function. Knowing all this, and holding our theory to be true, must we say that his is false? As empiricists we hate to, knowing that his evidence is the same as ours. There is an easy escape: we can declare his meanings to be different from ours, and restore agreement by translating his terms through the conciliatory proxy function.

Proxy functions reveal one way in which theories can conflict with each other and still be empirically equivalent, fitting all the same observations. It is a particularly benign case, we see, readily reconciled. But what about other cases? Suppose two empirically equivalent theories which cannot, so far as we can see, be rendered identical by reinterpretation of words. We can still accept both as true, if there is no logical contradiction between them. We can recognize them as two faithful descriptions, along very different conceptual lines, of one and the same world.

But what if there is a contradiction between them? This proves to be a red herring. There must be a sentence that is affirmed in one theory and denied in the other, but it must be a sentence whose truth value is undetermined by all possible observations, for the two theories are supposed to be empirically equivalent. At least one word in that sentence, then, must be a theoretical term whose conditions of application are not wholly fixed by empirical criteria. We can exploit its slack by treating it rather as two homonyms, one in the one theory and one in the other, and changing its spelling in one of the theories. The sentence that was affirmed in the one theory is no longer the one that was denied in the other; the contradiction is resolved. Similar adjustments elsewhere make the two theories logically compatible, and we can accept both theories as true. I owe this to Donald Davidson.

These last reflections have had to do with the general case of empirically equivalent but dissimilar theories; they have not borne specifically on ontology or reference, but they lead to a final sobering thought on that topic. Imagine again two empirically equivalent and logically compatible but dis-

similar systems of the world, and no visible way of bringing them into coincidence by reinterpretation of terms. Both are couched in the structure of classical quantification logic. As explained, we can accept both theories as true descriptions, along unlike conceptual lines, of the same world. But suppose the ontologies are unlike, and not identifiable through reinterpretation by any proxy function. Evidently then we shall have to regard the bound variables and quantifiers of the one theory as irreducibly different in meaning from those of the other, and thus acquiesce in a tolerantly pluralistic attitude toward ontology itself. This reflection debases ontology as the mere doctrine of proxy functions did not. That doctrine had to do only with the theory of evidence, and did not interfere with the ontological doctrine within our system of the world. This last reflection does; it bespeaks an ontological relativism even at the level where we are presuming to say what is true about the world.

The Phoneme's Long Shadow

1990

Reading of the *emic* and the *etic* recalls crotchety Joshua Whatmough. Along with being an authority on Gaulish, he was a pedantic classicist. As such he refused to say *phonemic*. The adjective was *phonematic*. I expect we all share his revulsion to some degree, but there is nothing for it. *Phonemic* is here to stay.

A more glaring specimen has reared its ugly head in the present volume—just barely, but I do mean “ugly head”—and that is (ugh!) *negentropy*, a fallout from physics. A physicist in need of a new word should be required to consult a literate colleague. *Entropy*, *ectropy*: What could be more pat? Pike did take refuge in *antientropy*, which affords some relief.

But so much for morphological aesthetics. I must get down to substance, the phoneme as such.

Besides providing the very backbone of phonetic theory, the phoneme idea is gratifying in its clarity. This is not evident at first, for the idea is usually defined with the help of the nebulous idea of meaning. Sounds belong to the same phoneme, for a language, if they can be interchanged without changing any meaning. However, there is happily a direct criterion in *assent*. Switching allophones in a sentence must not alter an informant's disposition to assent to it.

The phoneme inspired Pike, we see, to analogies that prove central to cultural anthropology. Let us consider what traits of the phoneme invite analogy. The phoneme reflects a neat blending of nature and nurture; for each of its allophones is specifiable in purely acoustical or physiological terms, but they are grouped as a phoneme only by force of culture. The phoneme is *culture-bound* in a remarkably apt sense of the phrase: Culture is what binds its allophones together as a phoneme.

The binding relation is the relation between fellow allophones: the relation, we may say, of homeophony. The relation holds relative to one culture and not another. Unaspirated /p/ and aspirated /p^h/ are homeophones for English and not for Hindi.

The generalized concept of *emic* that this suggests is relativity to a culture. Homeophony (unlike homophony) is emic, and consequently phoneme is emic. On the other hand, each allophone is etic: specifiable absolutely rather than just relative to one or another culture.

Let us look now for exact analogues. How about *nation* for phoneme, *citizen* for allophone, *compatriot* for homeophone, and *person* for sound? Not good. The compatriot relation is not emic—not in the manner of homeophony. Two citizens of a nation count as compatriots out and out, without our stopping to say what nation; but two allophones of a phoneme can be called homeophones only subject to specification of the language. The relation of compatriots is flatly dyadic, relating people in pairs. It is etic, it would seem, despite its ulterior dependence on nations. On the other hand, homeophony is covertly triadic, in its emicity, relating two sounds and a language. The language, the culture, persists here as an indispensable parameter, however tacit, unlike nation in the compatriot case.

A feeble but faithful analogy of homeophony is equality of economic value, as applied indiscriminately to commodities and moneys. If we take a momentary state of a market as our analogue of a language or society, then each economic value corresponds to a phoneme, each assortment of commodities and/or currency corresponds to a sound, and equality in value corresponds to homeophony. Two assortments are of equal value relative to one or another state of a market, just as two sounds are homeophonic relative to one or another language.

If one insists thus on strict analogies, one seems to see as keynote of the emic a tacit dependence on an indispensable cultural parameter: language in the paradigm case, market in the analogue. Complete the defectively emic dyadic relation by supplying the parameter and you have an etic triadic relation. The dyadic relation of homeophony between /p/ and /p^h/, which is the very paradigm of the emic, has the triadic “/p/ and /p^h/ are homeophones in English” as its etic analysis.

The painfulness of my one faithful analogue of the phonemic-phonetic contrast suggests that the utility of an emic-etic contrast for anthropological studies must have depended on looser analogies. Indeed it has, and this circumstance is reflected in the uncertainties and divergences in the use of

that terminology by Harris and Pike. But there is no questioning the importance of the distinctions that have figured in their use of it or in their quandaries over it.

One of those distinctions is the contrast between the mental and the material. It has dominated metaphysics, and it dominates one's spontaneous outlook as well. Because of the importance of translation in understanding an alien culture, the dialectic of the mental and the material has a special bearing on social anthropology; for the mental and the material meet head on in translation.

We must think here of what I call *radical* translation. The translator goes into the bush with no interpreter and no line on a related language. He is in the position of Pike when he does his public monolingual demonstrations. The data are behavioral, material. Yet translation purports to be a matching of the meanings of sentences, and meanings purport to be mental.

The translator empathizes. He projects himself into the Kalaba's sandals, as I put it in an allusion to Pike 36 years ago (Quine 1953, 60–63). But all his evidence for his empathy is still on the behavioral side, still material.

Nor is this a predicament peculiar to the radical translator. The infant learning the language, on its own or under parental guidance, has nothing better to go on. We must recognize that our notion of the mental life of other individuals, local or exotic, is only a projection from behavioral data by our own subjective lights. Language being social, nothing can count as linguistic meaning beyond shared dispositions to observable behavior in observable circumstances. We may go on blithely assuming a predominant matching of thoughts, but only for want of evidence to the contrary.

It is all too easy to finesse behavioral criteria by unwarranted imposition of one's own categories. There is a case in point, I suspect, in linguistic universals. A struggling radical translator will naturally and rightly impose familiar grammatical concepts on the exotic language where practicable, and will even warp them a bit and extend the familiar terminology as required. Similar tendencies have facilitated translation down through the ages and are probably responsible for much of the apparent similarity of structure from language to language. If on a global outlook some construction strikes a linguist as universal, then, let him look to his criteria. Only behavioral criteria, applicable prior to any translation manual, should carry conviction of universality unimposed.

Another important distinction that has been associated with the emic

and etic is the distinction between the insiders' conceptual apparatus and the conceptual apparatus by which we study it. It is of course a crucial one for anthropology, whether we stretch emic-etic to cover it or find other handy terms. Its importance was brought home to me in a book about belief by a respected anthropologist. His account was vitiated by a recurrent confusion, it seemed to me, between what the insiders believed and what they meant by the word that the author translated as "believe."

I have sensed at several points in this volume a misdirected modesty, a nonpartisan readiness to look upon one's own apparatus of conceptual analysis as just another cultural manifestation on a par, *sub specie aeternitatis*, with the insiders' mumbo jumbo. In a way this is true, but with a palpable difference. Nagged by the difference, one and another of my fellow contributors allude uncomfortably to one or another reassuring canon of scientific method, or, in Lett's case (Chapter 9), to "the community of scientific observers." Lett presses the defense with laudable excoriations of pseudoscience. But it is in the initial evenhandedness that I find misplaced objectivity.

In treating of the relation between our outside apparatus of culture analysis and the inside culture to which we apply it, I would speak flatly of fact or reality on the one hand, culture on the other. All science is fallible, as scientists are aware, but nothing is gained by affixing "probably" or "from our point of view" to its progressive deliverances. The physicist tells us how it is, subject to possible future correction; and this latter caveat, being invariable, is rightly left tacit. The scientist's fingers are permanently crossed, however slightly, be he physicist or anthropologist. And yet there remains, between outside and inside, a vital asymmetry. Our provisional but responsible commitment to our science extends to what we say about the exotic culture, and does not extend to what the insiders say within it.

Reference

- Quine, W. V. 1953. *From a Logical Point of View*. Cambridge, MA: Harvard University Press.

Three Indeterminacies

1990

Three theses of indeterminacy have figured conspicuously in my writings: indeterminacy of translation, inscrutability of reference, and underdetermination of scientific theory. Each of the three presupposes a distinctive further topic with concerns of its own. The topic presupposed by translation is stimulation. The topic presupposed by the inscrutability of reference is reification. The topic presupposed by underdetermination of science is empirical content. A further topic still, pursuant on the underdetermination of science, is truth of empirically equivalent theories. So I have seven interlocking topics. Two of them, stimulation and empirical content, are inseparable from an eighth: observation. In my fancy ball game I propose to touch all eight bases, beginning with the eighth.

1. Observation Sentences

When epistemology rounded the linguistic turn, talk of observable objects gave way to talk of observation terms. It was a good move, but not good enough. The terms ‘Stick’, ‘Stone’, ‘Milk’, ‘Mama’ are all very well, but many utterances that register observations are not terms, not announcements of objects; witness ‘It’s cold’, ‘It’s raining’. Another drawback was that in taking terms as a starting point we finesse reification; we concede objective reference out of hand, without considering what it is for or what goes into it.

It was a further improvement, therefore, to switch from observation terms to observation sentences. The old examples still carried over, since ‘Stick’, ‘Stone’, ‘Milk’, and ‘Mama’ qualify equally as one-word sentences, on a par with ‘It’s raining’ and ‘It’s cold’. They all qualify as occasion sentences,

true on some occasions, false on others. Selective assent to them can be conditioned to appropriate sensory stimulation.

When the child begins to speak of his surroundings, observation sentences will be his medium; for they can be learned without help of earlier language. Some of them—'Stick', 'Stone', etc.—do take on the status of terms when reification has set in. I have argued that the clearest manifestation of that stage is the bound variable, or its prototype the pronoun in relative classes. See §§4 and 6.

But let us linger over the observation sentences. It has become fashionable to stress the relativity of observation. When scientists marshal and check their own data or one another's, certainly they press no farther than is needed to assure agreement among witnesses conversant with the subject; for they are reasonable men. 'The mixture is at 180°C' and 'Hydrogen sulfide is escaping' are observational enough for any of them, and more recondite reports are observational enough for some. The practical notion of observation is thus relative to one or another limited community. An observation sentence is an occasion sentence that the members of the community can settle by direct observation to their joint satisfaction.

For philosophical purposes we can probe deeper, however, and reach a single standard for the whole speech community. Observable in this sense is whatever would be attested to on the spot by any witness in command of the language and his five senses. If scientists were perversely to persist in demanding further evidence beyond what sufficed for agreement, their observables would reduce for the most part to those of the whole speech community. Just a few, such as the indescribable smell of some uncommon gas, would resist reduction.

Observation sentences even in this ultimate sense are reports not of sense data still, but of ordinary external circumstances, as my first six examples show. Many are nevertheless learned by direct conditioning to sensory stimulation, and all of them could be. Hence their epistemological significance as a link between our sensory stimulation and our theories about the world.

And hence also their semantic significance for radical translation. They are the entering wedge into cognitive language for the translator, as well as for the child on his native heath. Other utterances—greetings, commands, questions—will figure among the early acquisitions too, but the first declarative sentences to be mastered are bound to be observation sentences of the jungle language with observation sentences of his own that have the

same *stimulus meaning*. That is to say, assent to the two sentences should be prompted by the same stimulations; likewise dissent.

2. Stimulation

It would seem then that this matching of observation sentences hinges on sameness of stimulation of both parties, the linguist and the informant. But an event of stimulation, as I use the term, is the activation of some subset of the subject's sensory receptors. Since the linguist and his informant share no receptors, how can they be said to share a stimulation? We might say rather that they undergo *similar* stimulation, but that would assume still an approximate homology of nerve endings from one individual to another. Surely such anatomical minutiae ought not to matter here.

I was expressing this discomfort as early as 1965.¹ By 1981 it prompted me to readjust my definition of observation sentence. In my original definition I had appealed to sameness of stimulus meaning between speakers,² but in 1981 I defined it rather for the single speaker, by the following condition:

If querying the sentence elicits assent from the given speaker on one occasion, it will elicit assent likewise on any other occasion when the same total set of receptors is triggered; and similarly for dissent.³

Then I accounted a sentence observational for a whole community when it was observational for each member. In this way the question of intersubjective sameness of stimulation could be bypassed in studies of scientific method, I felt, and deferred to studies of translation. There it continued to rankle.

The question was much discussed in the course of a closed conference with Davidson, Dreben, and Føllesdal at Stanford in 1986.⁴ Two years later, at the St. Louis conference on my philosophy,⁵ Lars Bergström observed that even my bypassing of the question within studies of scientific method was unsuccessful, since a sentence could be observational for each of vari-

1. E.g., in a lecture "Propositional Objects," published in *Ontological Relativity and Other Essays*.

2. Thus *Word and Object*, 43.

3. *Theories and Things*, 25.

4. July 14–17, supported by Stanford's Center for the Study of Language and Information.

5. "Perspectives on Quine," Washington University, April 9–13, 1988.

ous speakers without their being disposed to assent to it in the same situations. It is odd that I overlooked this, for already in 1974 I had remarked in effect that the fisherman's sentence 'I just felt a nibble' qualifies as observational only for the individual and not for the group.⁶

At the Stanford conference, Davidson proposed providing for intersubjective likeness of situation by locating the stimulus not at the bodily surface but farther out, in the nearest shared cause of the pertinent behavior of the two subjects. Failing a rabbit or other body to the purpose, perhaps the stimulus would be a shared situation, if ontological sense can be made of situations. But I remain unswerved in locating stimulations at the nearest point, for my interest is epistemological, however naturalized. I am interested in the flow of evidence from the triggering of the senses to the pronouncements of science; also in the rationale of reification, and in the credentials, if any, of the notion of cognitive meaning. It is these epistemological concerns, and not my incidental interest in linguistics, that motivate my speculations on radical translation. Hence, indeed, my disregard of literary or poetic aspects of translation. And thus it is that the subject's reification of rabbits and the like is for me decidedly a part of the plot, not to be passed over as part of the setting. My own acceptance of nerve endings and indeed of rabbits, on the other hand, is part of the naturalistic setting of my inquiry.

The view that I have come to, regarding intersubjective likeness of stimulation, is rather that we can simply do without it. The observation sentence 'Rabbit' has its stimulus meaning for the linguist, and the observation sentence 'Gavagai' has its stimulus meaning for the informant. The linguist observes natives assenting to 'Gavagai' when he, in their position, would have assented to 'Rabbit'. So he tries assigning *his* stimulus meaning of 'Rabbit' to 'Gavagai' and bandying 'Gavagai' on subsequent occasions for his informant's approval. Encouraged, he tentatively adopts 'Rabbit' as translation.

Empathy dominates the learning of language, both by child and by field linguist. In the child's case it is the parent's empathy. The parent assesses the appropriateness of the child's observation sentence by noting the child's orientation and how the scene would look from there. In the field linguist's case it is empathy on his own part when he makes his first conjecture about 'Gavagai' from the native's utterance and orientation, and again

6. "The Nature of Natural Knowledge," 72.

when he queries 'Gavagai' for the native's assent in a promising subsequent situation. We all have uncanny knack for empathizing another's perceptual situation. The knack is comparable, almost, to our ability to recognize while unable to sketch or describe them.

Empathy guides the linguist still as he rises above observation sentences through his analytical hypotheses, though here he is trying to project into the native's associations and grammatical trends rather than his perceptions. And much the same must be true of the growing child.

As for the lacuna that Bergström noted in my definition of observation sentence my new line would be to retain my 1981 definition of observation sentences for the single speaker, and then account a sentence observational for a group if it is observational for each member *and* if each would agree in assenting to it, or dissenting, on witnessing the pertinent occasion. What counts as witnessing the occasion will be judged, as in the translation case, by projecting oneself into the witness's position.

3. To Each His Own

A pioneer manual of translation has its utility as an aid to negotiation with the native community. Success in communication is judged by smoothness of conversation, by frequent predictability of verbal and non-verbal reactions, and by coherence and plausibility of native testimony. It is a matter of better and worse manuals rather than flatly right and wrong ones. Observation sentences continue to command the firmest agreement between rival manuals of translation; but their distinctive factuality is blurred now by the disavowal of shared stimulus meaning. What is utterly factual is just the fluency of conversation and the effectiveness of negotiation that one or another manual of translation serves to induce.

Discussion with Dreben helped to clarify these consequences of my new stance. In *Word and Object* I had already pointed out that communication presupposes no similarity in nerve nets; verbal behavior is inculcated on the strength only of surface stimulation. Such was my parable of the trimmed bushes (p. 8), alike in outward form but wildly unlike in their inward twigs and branches. Save the surface, in the painter's words, and you save all. But now, leaving the surface itself to Sherwin-Williams' tender mercies, I give the individual yet wider berth. His privacy widens apace.

Unlike Davidson, I leave the stimulations at the subject's surface, and private stimulus meaning with them. But they may be as idiosyncratic, as

all I care, as the subject's internal wiring itself. What floats in the open air is our common language, which each of us is free to internalize in his peculiar neural way. Language is where intersubjectivity sets in. Communication is well named.

A manual of Jungle-to-English translation constitutes an inductive definition of a translation together with a claim that it correlates sentences compatibly with the behavior of all concerned. The thesis of indeterminacy of translation is that these claims on the part of two manuals might both be true and yet the two translation relations might not be usable in alternation, from sentence to sentence, without issuing in incoherent sequences. Or, to put it another way, the English sentences prescribed as translations of a given Jungle sentence by the two rival manuals might not be interchangeable in English contexts.

The use of one or the other manual might indeed cause differences in speech afterward, as remarked by Robert Kirk in connection with the idioms of propositional attitude; but the two would do equal justice to the status quo.

The indeterminacy of translation is unlikely to obtrude in practice, even in radical translation.⁷ There is good reason why it should not. The linguist assumes that the native's attitudes and ways of thinking are like his own, up to the point where there is contrary evidence. He accordingly imposes his own ontology and linguistic patterns on the native wherever compatible with the native's speech and other behavior, unless a contrary course offers striking simplifications. We could not wish otherwise. What the indeterminacy thesis is meant to bring out is that the radical translator is bound to impose fully as much as he discovers.

4. Syntax

Readers have supposed that I extended my indeterminacy thesis to syntax. This puzzled me until I became aware, recently, of a subtle cause of the misconception.

In *Word and Object* (pp. 53, 68–72) I claimed that our distinctive apparatus of reification and reference is subject to indeterminacy of translation. The apparatus includes pronouns, '=', plural endings, indeed whatever

7. See Levy, 594–600, for an artificial example of indeterminacy of translation based on geometry and physics.

serves the logical purposes of quantifiers and variables. But these devices, some of my readers have reasoned, are part of what syntax is about. So indeterminacy, they supposed, extends to syntax.

The business of syntax is the demarcation of strings of phonemes proper to the language. More than one battery of grammatical constructions and vocabulary will probably be capable of generating the same total output of strings, but in this freedom there is no indeterminacy analogous to that of translation. Indeterminacy of translation consists rather in conflict in the outputs themselves.⁸

What misled those readers was the indeterminacy of translation of pronouns and other referential devices. But that indeterminacy was only over whether to equate certain Jungle locutions to these devices or to something else. The translator will accommodate them anyway, whatever his translations. He may or may not call them pronouns, plurals, quantifiers, etc., according as he thinks in terms of one or another manual of translation. The difference will be only verbal or, at most, a choice of one syntactic structure rather than another for generating one and the same total output of Jungle strings.

A factor that may have contributed to the misunderstanding is variation in the use of the word 'grammar', which is sometimes applied quite broadly. With this in mind I have now used the word 'syntax'.

5. Inscrutability of Reference

A thick and imposing periodical on the philosophy of language is published in Spanish twice a year in the Canary Islands under the title *Gavagai*. A book by David Premack, on his language experiments with chimpanzees, came out lately under the title *Gavagai*. Hubert Dreyfus has California vanity plates on his Volkswagen Rabbit that spell 'Gavagai'. The word has become the logo of my thesis of indeterminacy of translation, and now it is making its way in a wider world.

Ironically, indeterminacy of translation in a strong sense was not what I coined the word 'Gavagai' to illustrate. Seen as a term, the word illustrated inscrutability of reference. Seen as a sentence, the word did not illustrate indeterminacy of holophrastic translation of sentences, which is the more

8. The syntactician may indeed exercise some freedom in setting the limits of the language, but only marginally. See *From a Logical Point of View*, 53–55.

sweeping thesis; for 'Gavagai' is an observation sentence, translatable as '(Lo, a) rabbit' firmly enough.

Translation of 'Gavagai' as '(Lo, a) rabbit' is insufficient to fix the reference of 'gavagai' as a term; that was the point of the example. The inscrutability of reference, in that sense, was scarcely debatable. It was factually illustrated by the Japanese and Chinese classifiers (*Ontological Relativity*, pp. 35–8), and there was a more abstract argument for it in proxy functions (e.g., *Theories and Things*, pp. 19–22).

The phrases 'inscrutability of reference' and 'ontological relativity' dominated my account of these matters, and kindly readers have sought a technical distinction between them that was never clear in my own mind. But I can now say what ontological relativity is relative to, more succinctly than I did in the lectures, paper, and book of that title. It is relative to a manual of translation. To say that 'gavagai' denotes rabbits is to opt for a manual of translation in which 'gavagai' is translated as 'rabbit', instead of opting for any of the alternative manuals.

And does the inscrutability or relativity extend also somehow to the home language? In *Ontological Relativity* I said it did, for the home language can be translated into itself by permutations that depart materially from the mere identity transformation, as proxy functions bear out. But if we choose as our manual of translation the identity transformation, thus taking the home language at face value, the relativity is resolved. Reference is then explicated in paradigms analogous to Tarski's truth paradigm; thus 'rabbit' denotes rabbits, whatever *they* are, and 'Boston' designates Boston.

6. Reification

We and our fellow mammals have a robust sense of the reality of gross bodies around us. It hinges on salient, integrated patches and sharp edges in the visual field, reinforced by correlated tactual or olfactory stimulations, and subject mostly to only gradual distortion over time. Here is the root, surely, of reification. To begin with, however, the recurrent presentation of a body is much on a par with other similarities of stimulation that do not prompt reification. Recurrent confrontation of a ball is on a par at first with mere recurrent exposure to sunshine or cool air: the question whether it is the same old ball or one like it makes no more sense than whether it is the same old sunbeam, or the same old breeze.

True, an infant is observed to expect a steadily moving object to reap-

pear after it pauses behind a screen; but this all happens within the specious present, and is evidently a matter rather of simple extrapolation than of anything worthy of the name of reification. Again a dog's recognition of a recurrent individual is beside the point; the dog is responding to a distinctive odor or other trait, unavailable in the case of qualitatively indistinguishable balls.

To us the question whether we are seeing the same old ball or penny or just a similar one is meaningful even in cases where it remains unanswered. It is here that the reification of bodies is full blown. This venerable theory of the persistence and recurrence of bodies is characteristic of the use of reification in integrating our system of the world. If I were to try to decide whether the penny now in my pocket is the one that was there last week, or just another like it, I would try to reconstruct the simplest, most plausible account of my interim movements, costumes, and expenditures.

The things reified serve as nodes or reference points in the structure that we have progressively elaborated as our theory of nature. Bodies were the main thing, but subsequently we gained further links for our growing theoretical structure by reifying numbers and other abstract objects, and reifying also imperceptible concrete ones. The technical contribution of all such reifications seems to consist in a tightening of implications between such observations and high theory, by reinforcing loose truth-functional relations. See "Events and Reifications," pp. 169–71.

The lesson of proxy functions is that all the technical service rendered by our ontology could be rendered equally well by any alternative ontology provided only that we can express a one-to-one correspondence between the two. The gross bodies themselves, charter members of our ontology, could thus be superseded by proxies and not be missed. The primordial visual patches would still prevail in all their salience and integrity, and the same old observation sentences would be keyed to them as before. But the denotation of the terms would be changed, and the values of the variables.

7. Experiment

The indeterminacy of translation of sentences as wholes is thus one thing and the indeterminacy of translation of those very same expressions or others in the capacity of terms is quite another, namely, ontological relativity, or the inscrutability of reference. Now the underdetermination of scientific theory is yet a third thing. Roughly stated, it is that different theories of the world can be empirically equivalent.

Empirical equivalence has to do with the relation of theory to observation. That relation stands forth most explicitly in experiment, so let us look into that. The scientist has a backlog of accepted theory, as is considering a hypothesis for possible incorporation into it. The theory tells him that if the hypothesis under consideration is true, then, whenever a certain observable situation is set up, a certain effect should be observed. So he sets up the situation in question. If the predicted effect fails to appear, he abandons his hypothesis. If the effect does appear, his hypothesis may be true and so can be tentatively added to his backlog of theory.

Thus suppose a team of field mineralogists have turned up an unfamiliar crystalline mineral of a distinctively pinkish cast. They speak of it provisionally as *litholite*, for want of a better name. One of them conjectures its chemical composition. This is the hypothesis, of which I shall spare myself the details. From his backlog of chemical lore he reasons that if this chemical hypothesis is true, than any piece of litholite should emit hydrogen sulfide when heated above 180°C. These last provisions are the observables; for our mineralogist and his colleagues know litholite when they see it and hydrogen sulfide when they smell it, and they can read a thermometer.

The test of a hypothesis thus hinges on a logical relation of implication. On one side, the theoretical, we have the backlog of accepted theory plus the hypothesis. This combination does the implying. On the other side, the observational, we have an implied generality that can be directly tested, directly challenged—in this case by heating some of the pink stuff and sniffing.

8. Observational Categoricals

A generality that is compounded of observables in this way—‘Whenever this, that’—is what I call an *observational categorical*. It is compounded of observation sentences. The ‘Whenever’ is not intended to reify times and quantify over them. What is intended is an irreducible generality prior to any objective reference. It is a generality to the effect that the circumstances described in the one observation sentence are invariably accompanied by those described in the other.⁹

9. The observation categorical is not to be confused with the observation *conditional*, a less fruitful notion that I ventured in 1975. The observation conditional is formed of two standing sentences each of which has been built upon an observation sentence with the aid of theory. See *Theories and Things*, 26–27.

Though compounded of two occasion sentences, the observational categorical is itself a standing sentence, fair game for implication by scientific theory. It thus solves the problem of linking theory logically to observation, as well as epitomizing the experimental situation.

That situation is where a hypothesis is being tested by an experiment. An opposite situation is equally familiar: a chance observation may prompt us to conjecture a new observational categorical, and we may invent a theoretical hypothesis to explain it. For example we might notice willows leaning over a stream. This suggests the observational categorical: 'When a willow grows at the water's edge it leans over the water.' This suggests, in turn, a theoretical hypothesis: 'A willow's root nourishes mainly its own side of the tree.' Taken together with prior bits of theory, such that roots get more nourishment from wetter ground, and that nourishment promotes the growth of branches, the hypothesis is found to imply the observational categorical. Other observation categoricals will be implied too, and the continued testing of the hypothesis would proceed by testing various of them, along with further testing of the one that happened to suggest the hypothesis.

It should be noted, however, that the observational categorical of this particular example exceeds my definition in a subtle way: it is not compounded of two self-sufficient observation sentences. Its second component, 'it leans over the water', refers back to the willow of the first component. What figures here may be called an *objectual* observation categorical, as distinct from a *primitive* one. The primitive one had the form 'Whenever this, that'—an irreducible generality, with no thought of quantification or bound variables. On the other hand the objectual observation categorical does hinge on quantification, or the equivalent. It is a sophisticated construction, 'Whatever is thus and so is also thus and so', or indeed 'Every S is P': 'Every willow at water's edge is a leaner over water.' It presupposes reification of willows.

Primitive observation categoricals, unlike objectual ones, are ontologically innocent expressions of generalized expectation. I would expect some of them to enter the child's repertoire soon after his first observation sentences and well ahead of predication and objective reference, or reification. They would be his first standing sentences.

To a degree, however, they can already simulate predication and objective reference. Thus contrast 'Tabby is white' and 'Whenever Tabby, white'. The former requires that each presentation of Tabby occupy a white part of

the scene, whereas the latter requires merely that Tabby appear only when white is somewhere around. The former sentence is an objectual observation categorical, and the latter a primitive one. However, the child can acquire 'Tabby is white' in the primitive manner of 'Whenever Tabby, white' without noticing the discrepancy for a while, because the scenes first associated with the observation sentence 'Tabby' will show Tabby at the salient focus, and those first associated with 'white' will be white at the salient focus. The cat and the white will tend to *register*, as printers say. Subsuming the one set of scenes under the other will determine the cat as white, nearly enough, after all. Primitive observation categoricals may even thus start the child down the path to reification.

I think the initial utility of reification lies in just such stacking of qualities: in conveying that the cat and the white are the same part of the scene.

9. Test and Refutation

The stimulations under which a speaker is disposed to assent to an observation sentence constitute what I have called the *affirmative stimulus meaning* of the sentence for him. If the affirmative stimulus meaning of one observation sentence is included in that of another, the observation categorical that is compounded of the observation sentences in that order qualifies as *analytic*. An example, mentioned to me by Thomas Tymoczko, is 'Where there are three sticks, there are two sticks.' Another example is 'If it's a robin, it's a bird', or briefly 'Robins are birds'. Tymoczko's example is a primitive observation categorical, despite its arithmetical character, and the other is objectual.

An analytic judgement for Kant was one whose subject contained its predicate. Our present variety, likewise based on inclusion, is thus somewhat reminiscent of his. Analytic observation categoricals so defined are stimulus analytic in the sense of *Word and Object* (pp. 65–66), and analytic also for the most part, I suppose, in the sense of *Roots of Reference* (pp. 78–80).

It is the other sort of observation categoricals, the *synthetic*, that calls for testing. It is tested for truth or falsity by pairs of observations. It is not conclusively verified by observations that are conformable to it, but it is refuted by a pair of observations, one affirmative and one negative—thus observation of litholite at 180°C but absence of hydrogen sulfide, or observation of riverine willows leaning away from the water. The primitive ob-

servation categorical 'When the sun comes up the birds sing' is refuted by observing sunrise among silent birds.

The observational test of scientific hypotheses, in turn, and indeed of sentences generally, consists in testing synthetic observation categoricals that they imply. Here again, as in the case of the observation categorical itself, there is no conclusive verification, but only refutation. Refute an observation categorical, by an affirmative and a negative observation, and you have refuted whatever implied it.

But let us recall that the hypothesis regarding the chemical composition of litholite did not imply its observation categorical single-handed. It implied it with the help of a backlog of accepted scientific theory. In order to deduce an observation categorical from a given hypothesis, we may have to enlist the aid of other theoretical sentences and of many common-sense platitudes that go without saying, and perhaps the aid even of arithmetic and other parts of mathematics.

10. Holism

In that situation, the falsity of the observation categorical does not conclusively refute the hypothesis. What it refutes is the conjunction of sentences that was needed to imply the observation categorical. In order to retract that conjunction we do not have to retract the hypothesis in question; we could retract some other sentence of the conjunction, any one, instead. This is the important insight called *holism*. Pierre Duhem made much of it early in this century, but not too much.

The scientist thinks of his experiment as a test specifically of his new hypothesis, but only because this was the sentence he was wondering about and is prepared to reject. Moreover, there are also situations where he has no preconceived notion hypothesis; he just happens upon an anomalous phenomenon. It is a case of his happening upon a counter-instance of an observation categorical which, according to his current theory as a whole, ought to have been true. So he looks to his theory with a critical eye.

Over-logicizing, we may picture the accommodation of a failed observation categorical as follows. We have before us some set *S* of beliefs that was found jointly to imply the false categorical. Implication may be taken here simply as deducibility by the logic of truth functions, quantification, and identity. (We can always provide for more substantial consequences by incorporating appropriate premisses explicitly into *S*.) Now some one or

more of the sentences in *S* are going to have to be rescinded. We exempt some members of *S* from this threat on determining that the fateful implication still holds without their help. Any purely logical truth is thus exempted, since it adds nothing to what *S* would logically imply anyway; and sundry irrelevant sentences in *S* will be exempted as well. Of the remaining members of *S*, we rescind one that seems most suspect, or at least crucial to our overall theory. We heed a maxim of minimum mutilation. If the remaining members of *S* still conspire to imply the false categorical, we try rescinding another and restoring the first. If the false categorical is still implied, we try rescinding both. We continue thus until the implication is defused.

But this is only the beginning. We must also track down sets of sentences elsewhere, in our overall theory, that imply these newly rescinded beliefs; for those must be defused too. We continue thus until consistency seems to be restored. Such is the mutilation that the maxim of minimum mutilation is meant to minimize.

In particular the maxim constrains us, in our choice of what sentences of *S* to rescind, to safeguard any purely mathematical truth; for mathematics infiltrates all branches of our system of the world, and its disruption would reverberate intolerably. If asked why he spares mathematics, the scientist will perhaps say that its laws are necessarily true; but I think we have here an explanation, rather, of mathematical necessity itself. It resides in our unstated policy of shielding mathematics by exercising our freedom to reject other beliefs instead.

So the choice of which of the beliefs to reject is indifferent only so far as the failed observation categorical is concerned, and not on other counts. It is well, we saw, not to rock the boat more than need be. Simplicity of the resulting theory is another guiding consideration, however, and if the scientist sees his way to a big gain in simplicity he is even prepared to rock the boat very considerably for the sake of it. But the ultimate objective is to choose the revision as to maximize future success in prediction: future coverage of true observation categoricals. There is no recipe for this, but maximization of simplicity and minimization of mutilation are maxims by which science strives for vindication in future predictions.

It is difficult to see how anyone can question holism, in the sense now before us. Grünbaum has indeed argued against holism, but in a sense that I have never entertained. He construes holism as claiming that when a prediction fails, we can always save the threatened hypothesis by so revising

the backlog of accepted theory that it, plus the threatened hypothesis, will imply the *failure* of the prediction. I recognize no such presumption. Inactivating the false implication is all that is at stake. Explaining the unexpected counter-observation is quite another step of scientific progress, which may or may not be made in the fulness of time.

Holism in this moderate sense is an obvious but vital correction of the naive conception of scientific sentences as endowed each with its own separable empirical content. Content is shared, even by mathematics insofar as it gets applied.

A set or conjunction of sentences inclusive enough to imply synthetic observation categoricals without outside help will be called *testable*. Some unconjoined single sentences qualify as testable, notably the synthetic observation categoricals themselves. For the most part, however, a testable set or conjunction of sentences has to be pretty big, and such is the burden of holism. It is a question of critical semantic mass.

11. Empirical Content

Scientists are apt to feel that my account of test and refutation is still vastly over-simplified, despite its commendable holism. In practice a failed prediction is often excused, and a hypothesis saved, by finding or assuming some interference. Such an excuse, however, can be analyzed in my terms as a discovery of a logical error in the deduction of the observation categorical that was tested. Some tacit qualification of *ceteris paribus* needed to be sorted out. Some of its tacit clauses would have been needed among the experimental conditions—hence in the antecedent of the observation categorical—for a rigorous clinching of the implication.

The scientist does not tabulate in advance the whole fund of theoretical tenets and technical assumptions, much less the common-sense platitudes and mathematical laws, that are needed in addition to his currently targeted hypothesis in order to imply the observation categorical of his experiment. It would be a Herculean labor, not to say Augean, to sort out all the premisses and logical strands of implication that ultimately link the broad laws of physics with observation, if or insofar as linked they be. But my concern is with the central logical structure of empirical evidence. The point of my schematism of observation categoricals is to suggest how, in principle, empirical science might proceed if, heaven forbid, all were ex-

explicitly set down. In fused phrases of Kant and Russell, it is a question of how our knowledge of the external world is possible.

Our construction affords us a definition of the empirical content of a sentence or set of sentences. We can define it as the set of all implied synthetic observation categoricals, plus such further observation categoricals as may be *stimulus-synonymous* to any of these. Two observation categoricals are stimulus-synonymous if their component observation sentences have the same stimulus-meanings.

Recalling the strictures of §2 on stimulus meaning, we note that it is a question now of stimulus synonymy for *each* pertinent speaker; not between speakers.

The conception of terms as vehicles of meaning is now well behind us. At the outset we shifted the burden to sentences, and now we see that for the most part we must settle it rather upon sets or long conjunctions of sentences. A sentence or set of sentences is devoid of empirical content unless it is testable. These words echo an old verification theory of meaning, but they gain new force from our holistic standard of testability.

Do our broadest scientific laws escape evidence altogether? Yosida writes (pp. 207–8) that they ‘may become out of fashion, . . . they are never refuted by direct observation, they are the old soldiers who never die but only fade away.’ Very well: if in a final full accounting this is the state of affairs, then those laws must be accounted devoid of empirical content on any reckoning, mine or another.

12. Equivalent Theories

It appears that distinct global theories of the world, even logically incompatible ones, may be empirically equivalent—that is, alike in empirical content. Poincaré’s example comes to mind. The one theory assumes infinite space in which the familiar rigid bodies move about without change in size. The other theory assumes a finite spherical space in which those bodies shrink uniformly as they move away from center. The two are empirically equivalent.

One may ask what to count as distinct theories and what to count merely as different formulations of a single theory. If two empirically equivalent theory formulations are interconvertible sentence by sentence through reinterpretation of terms, surely we should regard them as formu-

lating a single theory. If we just interchange the words 'proton' and 'electron' in a physics treatise, we already have two formulations that are related in that trivial way. On the other hand Poincaré's two theories are not thus interconvertible, since one of them singles out a central point of space while the other provides for no such singularity.

Anyway I no longer consider the question of identity of theories worth pausing over. We can just speak of drastically unlike theory formulations, irreducible one to the other by any known system of sentence-to-sentence reinterpretation, and let the notion of theory go. I shall keep the word, but it may be construed as short for 'theory formulation'.

There is an air of paradox in the empirical equivalence of logically incompatible theories. But Poincaré's example is a case of it, and so also, for that matter, is the trivial example where 'proton' is switched with 'electron'. There is an almost equally trivial adjustment, however, pointed out to me by Davidson, that suffices to resolve any logical conflict of empirically equivalent theories. Given their empirical equivalence, any sentence implied by the one theory and denied by the other will have contained some theoretical term that is not firmly clinched by empirical criteria; so we can exploit the slack by respelling the term within one of the theories as a distinct and independent term, thus resolving the inconsistency.

13. Rival Theories

It remains, then, to decide what line to take regarding a theory that is logically compatible with our overall system of the world and presumed empirically equivalent to it, but is not reducible to it sentence by sentence through any known reinterpretation. Four alternatives have detached themselves in recent years.

One is the hard line that I took in *Theories and Things*, pp. 21–22. Here we reckon our own theory true and the other false, even while believing them empirically equivalent. The argument is that the truth predicate itself makes sense only within our theory, there being no higher tribunal. We remain free to oscillate between the two theories, for a richer perspective of the world; but the one to reckon true is the one we are in at the time.

Still it offends our empiricist sensibilities to distinguish thus invidiously between empirically equivalent theories, if they are equally elegant as well. I took a second and opposite line in the next essay eight pages later (but countermanded it in a subsequent printing), namely, acceptance of both

theories as true. The above difficulty about the truth predicate is then met by fusing the empirically equivalent theories as a single tandem theory. A central portion, comprising logic and much else, will have been common to the two theories; outlying wings will be redundant in the sense that either, but not both, could be dropped without loss of empirical content. Dagfinn Føllesdal and Roger Gibson expressed misgivings over this tandem solution. I still find it good if neither wing uses terms foreign to the rest of the system, for the merger has then merely enriched our stock of truths, as we hope, about the world. If, on the other hand, one wing uses additional and irreducible terms, engendered perhaps by Davidson's method of resolving incompatibility, I see good reason to agree with Føllesdal and Gibson. The subtheory involving the alien terms is comparable to an annex centered on some such scientifically indigestible term as 'entelechy' or 'grace' or 'nirvana', its only excuse being that its annexation is indifferent to the empirical content of the overall theory. It offends against the scientists' ideal of economy and the empiricists' standard of meaningfulness.

So my present position is that the second or tandem line is good if neither of the given empirically equivalent theories contains terms irreducible to the other, and bad otherwise. For the latter case, then, a third line suggests itself: account the sentences neither true nor false that contain the alien terms, since our truth predicate makes sense only relative to our own language and to translations into it.

A fourth alternative, which now appeals to me most, was urged by Davidson in the aforementioned Stanford meeting. Here the thought is that economy, which counted against the tandem idea, is imperative only as an ideal of theory construction and not of language. Meaningful application of the truth predicate, on the other hand, extends to the whole language and is not limited to any particular theory formulation. Empirically equivalent and logically compatible theories can be accepted as true descriptions of the world even if one of them uses terms irreducibly alien to the other. There is no call to fuse them into a single redundant theory. Our language can embrace the full vocabularies of both theories, and our truth predicate can then apply to each on its separate merits.

This line is in the spirit of scientific practice, which for the most part generates theories of limited scope and reckons each as true. A total coherent system of the world remains the scientific and philosophical ideal, but, again, multiple ways of achieving it can all be accounted true. Economy

still remains imperative in each, as also in each of the commonplace theories of more modest scope.

The theories will differ from one another in their ontologies —the ranges of their variables. These divergences can be accommodated within the inclusive language by assigning to each theory a predicate that purports to be true of all and only the objects of that theory. The quantifiers used in that theory are then to be construed as tacitly restricted by that predicate. Quantifications ' $\forall x(\dots)$ ' and ' $\exists x(\dots)$ ' are to be read as

$\forall x(Fx \rightarrow \dots)$ and ' $\exists x(Fx \rightarrow \dots)$ ', where ' F ' stands for that predicate.

I am indebted to Burton Dreben, as usual, for helpful suggestions elicited by earlier drafts.

References

- Grünbaum, Adolf. 1962. "The Falsifiability of Theories." *Synthese* 14:17–34.
- Kirk, Robert. 1969. "Quine's Indeterminacy Thesis." *Mind* 78:607–8.
- Levy, Edwin. 1971. "Competing Radical Translations." *Boston Studies in Philosophy of Science* 8:590–605.
- Poincaré, Henri. 1929. *Science and Hypothesis*, ch. 4. New York: Science Press.
- Premack, David. 1986. *Gavagai! or the Future History of the Animal Language Controversy*. Cambridge, MA: MIT Press.
- Quine, W. V. 1953. *From a Logical Point of View*. Cambridge, MA: Harvard University Press (2nd edition revised 1996).
- . 1960. *Word and Object*. Cambridge, MA: MIT Press.
- . 1969. *Ontological Relativity and Other Essays*. New York: Columbia.
- . 1973. *The Roots of Reference*. La Salle, IL: Open Court.
- . 1975. "The Nature of Natural Knowledge." S. Guttenplan ed., *Mind and Language*. Oxford: Clarendon, 67–82. [reprinted in this volume]
- . 1981. *Theories and Things*. Cambridge, MA: Harvard University Press.
- . 1985. "Events and Reification." E. LePore and B. McLaughlin ed., *Action and Events*. Oxford: Basil Blackwell, 162–71.
- Yosida, Natuhiko. 1984. "Scientific Laws and Tools for Taxonomy." *Annals of Japanese Association for Philosophy of Science* 6:207–18.

Preface to *The Logic of Sequences*

1990

Whitehead and Russell's *Principia Mathematica*, with its 1925 update of introduction and appendices, was for me still the last word in mathematical logic in 1931 when I undertook this dissertation. Europe had meanwhile progressed in these matters, but without impact on Harvard. It was given me to quaff at European wellsprings only the year after my dissertation had served its academic purpose.

I touted the dissertation as a generalization of *Principia*. Classes and dyadic relations had been developed separately in *Principia*, and relations beyond the dyadic had been left aside. I consolidated and generalized the treatment by taking relations in effect as classes of sequences, these being generated by iterating the ordered pair. It is a good plan, and usual nowadays; but in retrospect it is scarcely surprising. The dissertation had its strengths, but they lay elsewhere.

A minor advance over *Principia*, pedantic perhaps, was a sharper and more mechanical regimentation of the notation of formal proofs. A more substantial advance was economy of definition. I found that many notations introduced by definition in *Principia* were equivalent to notations that were already available in the system and were equally brief, or nearly so. Every definition calls for proving a sheaf of formulas relating the new notation to what has gone before; so the elimination of dispensable notations was a way of eliminating masses of formulas from *Principia*, as well as obviating needless memorization of symbols, without loss of rigor or perspicuity.

A more pertinent advance philosophically was strict observance of the distinction between sign and object: between use and mention of expressions. Frege had been scrupulous in the matter, as I later learned. But

Whitehead and Russell had lapsed. A result was vagueness and obscurity in their notions of proposition and propositional function and in their treatment of them, especially with the onset of the “ramified” theory of types. But for those confusions, the authors might have appreciated from the start that either their “axiom of reducibility” was invalid or else both it and the ramification of the theory of types were superfluous. I made the point in pp. 5–8 of the dissertation.

I wrote above that in the dissertation I took relations *in effect* as classes of sequences. Actually in the dissertation I limited the term ‘class’ rather to classes of singles, and then reckoned pairs, triples, etc., as members of relations as distinct from classes. For classes and relations together I retained the term ‘propositional function’ of *Principia*, but construed it extensionally.

Quantification emerges as in Peano: ‘ $\exists x Fx$ ’ is ‘ $\exists \{x: Fx\}$ ’, where $\exists$ is a function taking classes as arguments. Similarly universal quantification emerges as ‘ $\forall \{x: Fx\}$ ’. This is a line that I have found advantages in reviving at long last in the fourth edition of *Methods of Logic* (Harvard, 1982) but with ‘ $\{x: Fx\}$ ’ ontologically deactivated now as an innocent general term or predicate, a ‘such that’ clause.

In the dissertation I sought economy of primitive operations. I acknowledged four: predication (membership), catenation (pairing), an inelegant little monster called superplexion, and finally assertion, which should be disregarded. A genuine fourth, abstraction, I followed *Principia* in failing to recognize; I accounted it defined by the spurious $\circ 1.01$ (p. 20). On the other hand, I could and should have joyfully eliminated pairing, through Norbert Wiener’s clever definition. His paper had appeared in the *Proceedings of the Cambridge Philosophical Society* in 1914, but it was noticed neither in the updated *Principia* of that Cambridge nor in Wiener’s and my Cambridge, Mass.

Tarski and Gödel had meanwhile reduced the primitive operations to the economical three that remain standard today: a truth function, quantification, and membership. Their papers appeared in 1931, unknown to me, while I was working on my dissertation.

My professors esteemed the dissertation to the point of voting to subsidize its publication by Harvard University Press—a rare measure at Harvard. I reworked it for publication over the next two years, and it came out in 1934 radically transformed into *A System of Logistic*. Here assertion is dropped, abstraction is recognized as a primitive operation, superplexion

gives way to mere “congeneration,” amounting in effect to inclusion, and predication is eliminated. I eliminated it by identifying a proposition $\Phi x_1 x_2 \dots x_n$ with the sequence $\Phi x_1 x_2 \dots x_n >$ —an expedient that has resurfaced in current literature under the name of ‘proposition surrogate.’ In this way I made pairing serve the double purpose of pairing and predication, or membership. In Europe I had indeed learned meanwhile of Wiener’s elimination of ordered pairs, but this economy was now precluded by my alternative economy, the elimination of membership.

A System of Logistic evolved as a revision of my dissertation, and the subsidy for publishing the dissertation. What was afoot was a single publication, not two, so I was happy to retain in *A System of Logistic* any passages of the dissertation that fitted satisfactorily with the extensive revisions. What we find, however, are two decidedly distinct books. There is significant overlap in the underlying thought, but I have found barely two pages of shared text. Page ii and parts of pages iii–iv of the original Preface to the dissertation reappear in pages 2–3 and 9 of the Introduction to *A System of Logistic*. They subsist here by permission of Harvard University Press.

For a current discussion of the dissertation and a couple corrections see Burton Dreben, “Quine,” and my accompanying comment, in *Perspectives on Quine (Philosophers and Their Critics, 6)*, ed. Robert Barrett and Roger Gibson (Oxford: Blackwell, 1990).¹

1. The title and date of the forthcoming cited book changed slightly after this preface was written. They are updated here.

Two Dogmas in Retrospect¹

1991

In retrospecting “Two Dogmas” I find myself overshooting the mark by twenty years. I think back to college days, 61 years ago. I majored in mathematics and was doing my honors reading in mathematical logic, a subject that had not yet penetrated the Oberlin curriculum. My new love, in the platonic sense, was Whitehead and Russell’s *Principia Mathematica*.

I was taken with the clear, clean incisiveness of its formulas. But this was not true of its long introduction to volume 1, nor of some of the explanatory patches of prose that were interspersed through the three volumes. In those pages and passages the distinction between sign and object, or use and mention, was badly blurred. Partly in consequence, there was vague recourse to intensional properties, or ideas, under the disarmingly technical name of propositional functions. These ill-conceived mentalistic notions paraded as the philosophical foundation for the clean-cut classes, truth functions, and quantification that would have been a far better starting point in their own right.

The distrust of mentalistic semantics that found expression in “Two Dogmas” is thus detectable as far back as my senior year in college. Even earlier I had taken kindly to John B. Watson’s *Psychology from the Standpoint of a Behaviorist*, which Raymond Stetson had assigned to us in his psychology class. Nor do I recall that it shocked any preconceptions. It chimed in with my predilections.

In 1931, my second year out of college, I was writing my PhD thesis un-

1. This is a revised version of a paper that I presented at the University of Toronto in December 1990 in memory of my presentation of “Two Dogmas of Empiricism” there in December 1950.

der Whitehead's sponsorship at Harvard. It is my first and latest book: first because I wrote it before the others, and latest because it was published only a few months ago. Garland Publishing Company included it in their recently conceived series of twentieth-century Harvard PhD theses in philosophy. They did a surprisingly tidy job of my painfully wrought typescript, replete with logical symbols. I'm tickled. But this is not a commercial. The book is expensive, and I can offer better ones.

The relevance of my PhD thesis to our present topic is that in it I reworked the foundations of *Principia Mathematica* in strictly extensional terms, and propositional functions to the winds. The term 'propositional function' carried over, but to denote strictly classes, as I would now phrase it.

I was not abetted in my extensionalism by the Harvard professors of that time. Whitehead, C. I. Lewis, H. M. Sheffer, and E. V. Huntington all were soft on intensions and introspective meanings. But a postdoctoral fellowship the next year took me to a kindred spirit in Czechoslovakia: the great Carnap. He was just finishing his *Logische Syntax der Sprache*, and I read it and discussed it with him as it issued from his wife's typewriter. He was setting his face steadfastly against modal logic and mentalistic talk of meanings. His little logic text *Abriss der Logistik* had likewise been impeccably extensional, as was his impressive application of mathematical logic to epistemology in *Der logische Aufbau der Welt*.

True, in *Logische Syntax* we find him making capital of a purported distinction between analytic and synthetic truths. He didn't see this as a reinstatement of mentalistic meanings. He saw it as a matter of truth by linguistic convention. I expressed misgivings already in our discussions there in Prague, March 1933. Neil Tennant has uncovered, in the Carnap archives in Pittsburgh, this entry in Carnap's unpublished jottings of the time. I translate Carnap's somewhat telegraphic German:

Quine, 31.3.33

He says after some reading of my "Syntax" MS:

1. *Is there a difference in principle between logical axioms and empirical sentences?* He thinks not. Perhaps I seek a distinction just for its utility, but it seems he is right: gradual difference: they are the sentences we want to hold fast.

Evidently Carnap allayed my misgivings for a while. Three expository lectures on Carnap that I gave at Harvard in 1934 were abjectly sequacious.

But my misgivings surfaced again in 1935 when I wrote “Truth by Convention.” I quote the end of the first paragraph:

[D]evelopments of the past few decades have led to a widespread conviction that logic and mathematics are purely analytic or conventional. It is less the purpose of the present inquiry to question the validity of this contrast than to question its sense.

Carnap wrote to me in 1936 expressing ‘very much interest’ in “Truth by Convention,” and added that he was ‘very keen of discussing it when we meet in April.’ So I have no record of his reaction, not remembering the discussion. There must have been frequent discussions, for he was around Harvard all that summer.

I was diverging from Carnap because his aloofness from intensions and mentalism, which had so appealed to me, had proved to be insufficiently austere. Ironically, in those same years Carnap came to welcome intensions increasingly. The trend began when Tarski persuaded him that his “Thesis of Syntax” was untenable: the thesis that ‘philosophy is the syntax of the language of science.’ *Semantics* was wanted, not just syntax. Tarski was right, in his own austere sense of semantics, namely model theory and his theory of truth. But Carnap went further, even embroiling himself in modal logic. I learned of this in 1938 from Hempel, and wrote to Carnap in dismay. He duly pondered my sermon, so he wrote in reply, and he went on to defend himself as follows:

Although we do not ordinarily like to apply intensional languages, nevertheless I think we cannot help analyzing them. What would you think of an entomologist who refuses to investigate flees [sic] and lice because he dislikes them?

Well, the fleas and lice proved addictive. By 1946 he was championing modal logic.

Carnap was again at Harvard in the fall and winter term of 1940–41, along with Tarski and Russell. Glorious days. Tarski and I argued persistently with Carnap over his appeal to analyticity in the opening pages of his work in progress, *Introduction to Semantics*.

You don’t write when you can meet and talk. I find nothing on analyticity and meaning in my correspondence with Carnap until 1943, when I was in Washington as a naval officer. I wrote him a long letter about his *Introduction to Semantics*. One issue was analyticity and another was my cri-

terion of ontological commitment as applied to abstract objects. The two issues were linked, for Carnap viewed his appeal to abstract objects as empty convention, and their quasi-existence analytic.

I had not thought to look on my strictures over analyticity as the stuff of revolution. It was mere criticism, a negative point with no suggestion of a bright replacement. I had not felt moved to follow "Truth by Convention" with more of the same. But word got around. Nelson Goodman had joined in some of the discussions of 1940–41 with Carnap, Tarski, and me. In June and July of 1947 a triangular correspondence on the issue developed among Goodman, Morton White, and me. In 1950 I was invited by the program committee of the American Philosophical Association to present a paper on the subject at the December meeting here in Toronto. Hence "Two Dogmas." It is remarkable that my most contested and anthologized paper was an assignment. The response was quick and startling. The paper appeared in *Philosophical Review* a few weeks after the Toronto meeting, and four months later there were symposia on it in Boston and at Stanford.

Looking back on it, one thing I regret is my needlessly strong statement of holism.

The unit of empirical significance is the whole of science. . . . Any statement can be held true come what may, if we make drastic enough adjustments. . . . Conversely . . . no statement is immune to revision.

This is true enough in a legalistic sort of way, but it diverts attention from what is more to the point: the varying degrees of proximity to observation, the example of the brick houses in Elm Street. In later writings I have invoked not the whole of science but chunks of it, clusters of sentences just inclusive enough to have critical semantic mass. By this I mean a cluster sufficient to imply an observable effect of an observable experimental condition.

I formulate the matter nowadays in terms of what I call *observation categoricals*. An observation categorical is a generalization of the form 'Whenever this, that' where 'this' and 'that' are observation sentences. I'll say more about observation sentences later. Now a cluster of sentences has critical semantic mass if it implies an observation categorical; and the experimental check of the cluster of sentences consists in testing the implied categorical by arranging for fulfillment of its observable protasis and seeing whether the apodasis is realized.

This is meant as a schematic caricature of the experimental method. In practice many sentences of the cluster are tacit and some are probabilistic or dependent on unspecified things being equal. But I think it catches the essence of experimental testing, and therewith of empirical content.

I have appealed here to *implication*: the cluster of sentences implies the observation categorical. In so doing I give logic a special status: logical implication is the link between theory and experiment. However, we remain free here to adjust and to vary the limits of what to count as logic.

Thus consider the law of excluded middle, which I cited in “Two Dogmas” to illustrate my claim that ‘no statement is immune to revision.’ If the prospect of simplification and clarification at the level of quantum physics were to prove so overwhelming as to induce us to abrogate the law of excluded middle, we could still do so. It would just mean setting the limits of what to count as logic very narrowly and reckoning the law of excluded middle to the clusters of defeasible hypotheses that do the implying.

In a footnote to “Two Dogmas” I noted Duhem’s priority in stressing holism. As a matter of curiosity, however, I might mention that when I wrote and presented “Two Dogmas” here forty years ago, and published it in the *Philosophical Review*, I didn’t know about Duhem. Both Hempel and Philipp Frank subsequently brought Duhem to my attention, so I inserted the footnote when “Two Dogmas” was reprinted in *From a Logical Point of View*. Another insert was page 35 on postulates, which was prompted by a paper by Richard Martin in defense of Carnap on analyticity. But that was explained in a footnote.

I think Carnap’s tenacity to analyticity was due largely to his philosophy of mathematics. One problem for him was the lack of empirical content: how could an empiricist accept mathematics as meaningful? Another problem was the *necessity* of mathematical truth. Analyticity was his answer to both.

I answer both with my moderate holism. Take the first problem: lack of content. Insofar as mathematics gets applied in natural sciences, I see it as sharing empirical content. Sentences of pure arithmetic and differential calculus contribute indispensably to the critical semantic mass of various clusters of scientific hypotheses, and so partake of the empirical content imbibed from the implied observation categoricals.

As for inapplicable parts of mathematics, say higher set theory, I sympathize with the empiricist in questioning their meaningfulness. We do

keep their sentences as meaningful, but only because they are built of the same lexicon and grammatical constructions that are needed in applicable mathematics. It would be an intolerably pedantic *tour de force* to gerrymander our grammar in such a way as to account the inapplicable flights ungrammatical while preserving the applicable part.

This does put it to us to ponder truth and falsity for the inapplicable sentences. In doing so we can get some guidance from a maxim that is already serving the natural scientist, namely Occam's razor: where choice is otherwise undetermined, opt for economy. This attitude is in keeping with my inclination to minimize the cleavage between mathematics and natural science. That tendency is abetted already by my point about shared empirical content, and also by my questioning the analytic/synthetic distinction.

What then about the other problem, that of the necessity of mathematical truth? This again is nicely cleared up by moderate holism, without the help of analyticity. For let us recall that when a cluster of sentences with critical semantic mass is refuted by an experiment, the crisis can be resolved by revoking one *or* another sentence of the cluster. We hope to choose in such a way as to optimize future progress. If one of the sentences is purely mathematical, we will not choose to revoke it; such a move would reverberate excessively through the rest of science. We are restrained by a maxim of minimum mutilation. It is simply in this, I hold, that the necessity of mathematics lies: our determination to make revisions elsewhere instead. I make no deeper sense of necessity anywhere. Metaphysical necessity has no place in my naturalistic view of things, and analyticity hasn't much.

Analyticity undeniably has a place at a common-sense level, and this has made readers regard my reservations as unreasonable. My threadbare bachelor example is one of many undebatable cases. It is intelligible and often useful in discussion to point out that some disagreement is purely a matter of words rather than of fact. The point can commonly be sustained and acted upon by a paraphrase that circumvents a troublesome word. Often in talking with a foreigner we recognize some *impasse* as due to his having mislearned an English word rather than to his having a bizarre view of the subject matter. This is a bit of practical psychology at which we are all adept.

In *Roots of Reference* I proposed a rough theoretical definition of analyticity to fit these familiar sorts of cases. A sentence is analytic for a native speaker, I suggested, if he learned the truth of the sentence by learning the

use of one or more of its words. This obviously works for 'No bachelor is married' and the like, and it also works for the basic laws of logic. Anyone who goes counter to *modus ponens*, or who affirms a conjunction and denies one of its components, is simply flouting what he learned in learning to use 'if' and 'and.' (I limit this to native speakers, because a foreigner can have learned our words indirectly by translation.)

I also recommended improving this rough definition by providing for deductive closure, so that truths deducible from analytic ones by analytic steps would count as analytic in turn. All logical truths in my narrow sense—that is, the logic of truth functions, quantification, and identity—would then perhaps qualify as analytic, in view of Gödel's completeness proof.

If the logical truths are analytic—hence true by meanings of words—then what are we to say of revisions, such as imagined in the case of the law of excluded middle? Do we thereby change our theory or just change the subject, change the meaning of our words? This has been a recurrent challenge, and my answer is that in elementary logic a change of theory *is* a change of meaning. Repudiation of the law of excluded middle would be a change of meaning, and no less a change of theory for that.

You could also have a change of meaning in elementary logic without change of theory. That would be if the word 'and' were merely put to use in place of 'or' and vice versa: no change of theory. But in abandoning the law of excluded middle we would not be preserving the law in *any* notation.

For the past five minutes I have been expressing a generous attitude toward analyticity that may seem out of character. In fact my reservations over analyticity are the same as ever, and they concern the tracing of any demarcation, even a vague and approximate one, across the domain of sentences in general. The crude criterion in *Roots of Reference*, based on word learning, is no help; we don't in *general* know how we learned a word, nor what truths were learned in the process. Nor do we have any reason to expect uniformity in this regard from speaker to speaker, and there is no reason to care. Elementary logic and the bachelor example *are* clear enough cases, but there is no going on from there.

It is the wrong kind of question. When in relativity theory momentum is found to be not quite proportional to velocity, despite its original definition as mass times velocity, there is no flurry over redefinition or contradiction in terms, and I don't think there should be. The definition served its purpose in introducing a word for subsequent use, and the word was

thereafter ours to use in the evolving theory, with no lingering commitments. Definition is episodic. Mostly in natural science we are not even favored with definitions, much less bound by them. New terms are just introduced by partial descriptions: electrons, neutrinos, quarks.

In short, I recognize the notion of analyticity in its obvious and useful but epistemologically insignificant applications. The needs that Carnap felt for the notion in connection with mathematical truth are better met through holism. Beyond its manifest cases I find analyticity less help than hindrance. It begets an uncritical notion of meaning, or synonymy, that can induce a false sense of understanding. For it is clear that analyticity and synonymy are interdefinable.

Well, domestically interdefinable. Two expressions are synonymous iff their biconditional or equation is analytic. But I should mention that synonymy and therefore meaning are even worse off than analyticity when we transcend a single language. For if the two expressions to be equated belong to different languages, their biconditional or equation is far from analytic; it is incoherent, belonging to no language.

I mentioned my letter of 1943, where I took issue with Carnap both on analyticity and on ontological commitment to abstract objects, the two issues being linked. This same pairing of the two issues recurs in the antepenultimate paragraph of "Two Dogmas." Carnap's separation of questions of existence into questions of fact and questions of framework was a separation of the synthetic and the analytic. Collapse this epistemological duality and you collapse the ontological duality. Sticks, stones, sets, and numbers all become, for me, denizens of the world on an equal footing. Values of variables.

So also for the contrast noted in the remaining two paragraphs of "Two Dogmas": the contrast supposed by Carnap and C. I. Lewis between the factual and the pragmatic. 'In repudiating such a boundary,' I wrote, 'I espouse a more thorough pragmatism.' This passage had unforeseen consequences. I suspect it is responsible for my being widely classified as a pragmatist. I don't object, except that I am not clear on what it takes to qualify as a pragmatist. I was merely taking the word from Carnap and handing it back: in whatever sense the framework for science is pragmatic, so is the rest of science.

Let us now stand off, look at the two dogmas, and consider what the point is in repudiating each of them. Repudiation of the first dogma, analyticity, is insistence on empirical criteria for semantic concepts: for

synonymy, meaning. Language is learned and taught by observing and correcting verbal behavior in observable circumstances. There is nothing in linguistic meaning that is not thus determined. John Dewey made this point long ago. What I did, not in “Two Dogmas” but in *Word and Object*, was to press that point for its negative implications regarding the notion of meaning.

As for the second dogma, reductionism, its repudiation is holism, seconding Duhem. The reductionism that is properly at stake here is moderate reductionism. Extreme reductionism, the notion that every scientific sentence should have a full translation in sense-datum language, is by now a straw man. Moderate reductionism was the lingering dogma, the notion that each scientific sentence has its own separate empirical content. Moderate *holism* is its denial. It says that a scientific sentence cannot in general be expected to imply empirical consequences by itself. A bigger cluster is usually needed.

“Two Dogmas” is occasionally quoted for my depiction of

[t]he totality of our so-called knowledge or beliefs . . . [as] a man-made fabric which impinges on experience only along its edges.

Maybe this sparked Joe Ullian’s title for the little manual that he and I coauthored twenty years later: *The Web of Belief*. Clearly my metaphor needed unpacking, and that was largely my concern in the ten years between “Two Dogmas” and *Word and Object*.

Becoming more consciously and explicitly naturalistic, I stiffened up my flabby reference to ‘experience’ by turning to our physical interface with the external world: the physical impacts of rays and molecules upon our sensory surfaces. For purposes of a systematic account, I needed to marshal these scattered impacts as unit aggregates of some sort. Exteroceptors offered a neat solution. Impacts of rays and particles are irrelevant except as they trigger receptors, and happily it is only a question of triggering, with no question of more or less. So I identified one’s input from the external world, on any given occasion, with one’s global *neural input* on that occasion: hence with the temporally ordered set of all one’s triggered exteroceptors during that brief moment.

Intricate brain processes, which neurologists are illuminating bit by bit, intervene between this neural input and perception. Exciting though that physiological research is, I was able to bypass it for my purposes by leaping

to the resulting relation of *perceptual similarity* between global neural inputs. Inputs that are grossly dissimilar intrinsically—that is, in respect of what receptors are triggered—can issue in similar perceptual effects after the brain's swift and unconscious work of selecting and correlating. Perceptual similarity comes in degrees, and admits of a coarse behavioral criterion in the reinforcement and extinction of responses.

The grouping of global neural inputs by this relation of perceptual similarity affords a sufficient theoretical basis, I have argued, for distinguishing the sensory modalities and the various sensory qualities and *Gestalten*. A striking parallel emerges with Carnap's *Logischer Aufbau der Welt*, in his derivations from similarities of global experiences.

I should say that in this sketch I am somewhat updating *Word and Object* in the light of later writings.

The sentences near the periphery of the fabric—in the metaphor of “Two Dogmas”—came in *Word and Object* to be called observation sentences. Their connection with experience was explained as association with ranges of perceptually similar neural inputs. Primitively the association was by conditioning. Further associations are forged in the fullness of time by retroaction from one's developing theory of the world.

The connections *into* the fabric, between observation sentences and theoretical sentences, are forged by shared vocabulary. For the observation sentences are couched, not in sense-datum language, but directly in the thing language, as Carnap called it. Thing words are among the words earliest learned, though learned at first only as one-word observation sentences or as segments of monolithic observation sentences. They are there awaiting gradual reification of their designata and incorporation into theoretical contexts.

Thanks to this sharing of vocabulary by observation sentences and theoretical sentences, logical relations hold between them. The pertinent logical relation, we saw, is logical implication of observation categoricals by clusters of theoretical sentences. This is, of course, the merest caricature of scientific method. It invites analysis of actual chunks of serious science, to see just what some of the typical chunks with critical semantic mass might be, and how the chains of logical implication from them to observation categoricals might run.

All this still leaves the heuristics of hypothesis untouched: that is, the technology of framing hypotheses worth testing. This is the domain of the

maxim of minimum mutilation, and of Occam's razor. It is the domain also of standard deviation, probable error, and whatever else goes into sophisticated statistical method.

Deeper insights into the nature of scientific inference and explanation may some day be gained in neurology, coupled perhaps with computer simulation, as hinted by the new developments in so-called connectionist models; I think of Paul Churchland. The nature and nurture of science remains a fertile field of inquiry, varied and inviting.

Structure and Nature

1992

We are familiar with three adequate but incompatible ways of modeling number theory in set theory, or the theory of classes; and there are infinitely many more. We bandy our numbers without caring which classes we are bandying from among this wealth of alternatives. We are just content that we are operating somewhere within the ontology of classes to which we have committed ourselves anyway for other purposes. This situation, long since remarked upon by Paul Benacerraf¹ and others, receives formal expression in Ramsey sentences.² Such a sentence merely affirms that *there are* classes obeying the desired laws—arithmetical laws in this case—and then proceeds to apply such unspecified classes to whatever arithmetical business we had in mind.

That is the structuralist treatment of number. It is just a way of eliminating an idle question—“What is number?”—and a gratuitous decision among indifferent alternatives. Ontological economy is not its point; the same old classes are presupposed.

David Lewis³ has ventured the next step: a structuralist treatment of classes themselves. Now just as structuralism in number theory depended on the various ways of reducing numbers to classes, so structuralism in class theory depends on reducing classes in turn to—what? For Lewis, as for most of us, I should like to think, there are only classes and concrete individuals. So his structuralism is indeed ontologically significant. It

1. “Mathematical Truth,” *The Journal of Philosophy* 70, no. 19 (November 8, 1973): 661–679.

2. Frank Ramsey, *The Foundations of Mathematics* (New York: Routledge, 1931), 212–236.

3. *Parts of Classes* (New York: Blackwell, 1991).

professes a reduction of classes to individuals—hence out-and-out nominalism.

Arithmetical structuralism, as expressed in Ramsey sentences, depended on there being classes in which the required arithmetical structures could be defined and realized, though we were freed of choosing among alternatives. Similarly for Lewis's structuralism of classes; it depends, for its truth, upon there being individuals in which the desired classial structures can be defined and realized. Thus he needs a theory of individuals.

What he uses is mereology, or the theory of part and whole, with atoms. By ingenious constructions he carries the day, contingently on there being enough atoms, and granted George Boolos's⁴ view of second-order logic as just a classless logic of plurals. If there are not enough atoms, then the higher reaches of set theory turn out meaningful but false. How much of axiomatic set theory is true comes to depend on the size of the concrete universe: on how many atoms there are. If there are only finitely many, then so much the worse for classical set theory even at its weakest. Truth in mathematics comes to depend on the richness of nature. Lewis appreciates all this.

Uncertain as to the richness of nature, we can still pursue higher set theory in a conjectural spirit, finding out how it would be if nature were up to it. In any event, Lewis's construction renders the whole of set theory *meaningful* from his nominalistic point of view, however much of it is thereby rendered true and how much false. And, ironically enough, the scientific utility of applied mathematics is independent of all that.

Structuralism for classes, hence for all abstract objects, is undeniably congenial. They are things that are known anyway only by their structural role in cognitive discourse; never by ostension. By ostension and extrapolation we do learn what individuals qualify as members of the class of cats, but that class itself is just an abstract entity to which each cat bears a cryptic epsilon relation. Your class of cats and mine can still be different things, and your membership relation and mine can be different, if it makes any sense to say so, though we see eye to eye on every cat. Such is the appeal of structuralism regarding abstract objects.

Much though I admire Lewis's reduction, however, it is not for me. My own line is a yet more sweeping structuralism, applying to concrete and abstract objects indiscriminately. I base it, paradoxical as this may seem, on

4. "Nominalism Platonism," *Philosophical Review* 94, no. 3 (July 1985): 327–344.

a naturalistic approach to epistemology. Natural science tells us that our ongoing cognitive access to the world around us is limited to meager channels. There is the triggering of our sensory receptors by the impact of molecules and light rays. Also there is the difference in muscular effort sensed in walking up or down hill. What more? Even the notion of a cat, let alone a class or number, is a human artifact, rooted in innate predisposition and cultural tradition. The very notion of an object at all, concrete or abstract, is a human contribution, a feature of our inherited apparatus for organizing the amorphous welter of neural input.

Language, the vehicle of science, is linked to our neural input by neural mechanisms of association or conditioning. Early in our learning of language we learn to associate certain expressions with certain ranges of neural input. I call these expressions *observation sentences*. Grammatically, some are indeed sentences, e.g., 'It's raining', and some are nouns or adjectives, e.g., 'Cat' or 'Milk' or 'White'. To begin with, they are just expressions associated holophrastically with ranges of neural input. They are as if to say, with William James, "Hello, thingumbob again"—except that even 'thingumbob' hints more of objectual reference than I could wish. The observation sentence names nothing, to begin with, neither neural input nor external object. But it is learned from adults who have learned the ways of reference, so it is indeed an expression, such as 'It's raining' or 'Cat' or 'Milk' or 'White', that is destined for eventual integration into a system of objectual reference on the child's own part as he matures. This growing system of terms and purported objects is our science of nature, our organization of the welter of neural input.

Reification and implication are the key principles by which that organizing proceeds. We master implication in the course of learning to use the logical particles, such as 'not', 'and', 'or', 'if', 'every', 'some'. For instance, our learning to use 'not' and 'or' consists in part in learning to assent to ' q ' whenever we have assented to ' $\text{not } p$ ' and ' p or q '. Our learning of 'every' consists in part in learning to assent to ' a is a G ' once we have assented to ' a is an F ' and 'Every F is a G '.

It is in implications of this last sort that reification makes its contribution, in our having reified the object a and the other F s and G s. The reification of bodies comes in stages in one's acquisition of language, each successive stage being more clearly and emphatically an affirmation of existence. The last stage is where the body is recognized as identical over time, despite long absences and interim modifications. Such reification

presupposes an elaborate schematism of space, time, and conjectural hidden careers or trajectories on the part of causally interacting bodies. Such identifications across time are a major factor in knitting implications across the growing fabric of scientific hypotheses.

At more sophisticated stages in the development of language and science, implications are enhanced by positing further objects, no longer observable; thus subvisible particles, also numbers and other classes.

The implications thus forged are what relate our evolving scientific theory to our heterogeneous neural input and hence ultimately to the external world that our scientific theory purports to describe. This fabric of interlocking sentences hooks on to the neural input at the observation sentences. Some of these, we saw, were learned by primitive conditioning to ranges of neural input. Many further observation sentences, more sophisticated ones, are scientific sentences retroactively associated with observation through the intervention of theory; thus the chemist sees at a glance that there was copper in the solution, and perceives at a sniff that sulphur dioxide escaped. Now the way implication links theory to neural input is that a bundle of scientific hypotheses is found to imply that the fulfillment of one of two given observation sentences should always ensure fulfillment of the other. This dictates an experiment: he brings about fulfillment of the one and checks for the other. Failing the other, the bundle of hypotheses is discredited; one or another of them must be revoked.

Such, in schematic caricature, is surely the evidential relation of scientific theory to the triggering of our neuroreceptors by the external world. I have ventured more detail elsewhere,⁵ and much more wants venturing. Perhaps a fair bit of science might be found on fuller analysis to be beyond the reach of evidence altogether. But this sketch will suffice for what I want to say about structuralism.

The point I now want to make is one that over the years I have repeatedly made in terms of what I call *proxy functions*. The point is that if we transform the range of objects of our science in any one-to-one fashion, by reinterpreting our terms and predicates as applying to the new objects instead of the old ones, the entire evidential support of our science will remain undisturbed. The reason is twofold. First, implication hinges only on

5. *The Roots of Reference* (Peru, IL: Open Court, 1974); and *Pursuit of Truth* (Cambridge, MA: Harvard University Press, 1990).

logical structure and is independent of what the objects, the values of the variables, may be. Second, the association of observation sentences with ranges of neural input is holophrastic. It is independent of reifications, independent of whatever objects the observation sentences or their parts may be taken to refer to as terms.

The conclusion is that there can be no evidence for one ontology as over against another, so long anyway as we can express a one-to-one correlation between them. Save the structure and you save all. Certainly we are dependent on a familiar ontology of middle-sized bodies for the *inception* of reification, on the part both of the individual and of the race; but once we have an ontology, we can change it with impunity. For abstract objects this is unsurprising, and quite in the spirit of Ramsey, Lewis, and Benacerraf. For familiar bodies it is less intuitive. But we must bear in mind that an observation sentence, say, 'That's a rabbit', retains all its original visual associations; it is only the purported object that is gratuitously displaced.

This global ontological structuralism may seem abruptly at odds with realism, let alone naturalism. It would seem even to undermine the ground on which I rested it: my talk of impacts of light rays and molecules on nerve endings. Are these rays, molecules, and nerve endings themselves not disqualified now as mere figments of an empty structure?

Naturalism itself is what saves the situation. Naturalism looks only to natural science, however fallible, for an account of what there is and what what there is does. Science ventures its tentative answers in man-made concepts, perforce, couched in man-made language, but we can ask no better. The very notion of object, or of one and many, is indeed as parochially human as the parts of speech; to ask what reality is *really* like, however, apart from human categories, is self-stultifying. It is like asking how long the Nile really is, apart from parochial matters of miles or meters. Positivists were right in branding such metaphysics as meaningless.

But early positivists were wrong if and when they concluded that the world is not really composed of atoms or whatever. The world is as natural science says it is, insofar as natural science is right; and our judgment as to whether it is right, tentative always, is answerable to the experimental testing of predictions.

We saw that reference could be wildly reinterpreted without violence to evidence. We see now that that is just part of a wider picture. Presumably yet more extravagant departures, resistant even to sentence-by-sentence

interpretation into our own science, could conform equally well to all possible observations. If we were to encounter such a case and somehow recognize it as such, we might even proceed to master it and then switch back and forth for richer perspectives on reality. But naturalism would still counsel us that reality is to be grasped only through a man-made conceptual scheme, albeit any of various.

My global structuralism should not, therefore, be seen as a structuralist ontology. To see it thus would be to rise above naturalism and revert to the sin of transcendental metaphysics. My tentative ontology continues to consist of quarks and their compounds, also classes of such things, classes of such classes, and so on, pending evidence to the contrary. My global structuralism is a naturalistic thesis about the mundane human activity, within our world of quarks, of devising theories of quarks and the like in the light of physical impacts on our physical surfaces.

Commensurability and the Alien Mind

1992

Variety is the spice of life. Variety in language and culture is the spice of cosmopolitan life. *Vivent les différences*. But let me venture some homely reflections on the nature and limits of such differences.

I begin, as is my wont, with the field linguist who is investigating a remote tribe and eliciting and codifying their strange language. He ends up with a manual of translation. What qualifies it as successful? That it determines an English equivalent for every native sentence? An empty answer; what counts as equivalent is the whole question. The real test of a manual of translation is simply successful communication, and the test of successful communication is smoothness of dialogue and success in negotiation.

I hold that two manuals of translation could be wholly successful and yet give conflicting translations. This is not merely to say that they could give different translations; phrasing and choices of words can vary without conflicting. The conflict that I picture is rather as follows. Each manual translates a given dialogue coherently, but the two manuals produce incoherence when used alternately sentence by sentence.

This would not happen as long as the sentences were mere reports of observable events, but it is always possible when what is translated is appreciably theory-laden.

This thesis of indeterminacy of translation is by no means a theory of untranslatability. There are good translations and bad, and the two conflicting manuals imagined are both good. However, there are also plenty of cases of untranslatable sentences, and they are commonplace even within our own language. A sentence about neutrinos admits of no translation into the English of 1900. Terms for newly postulated particles are adequately explained, but usually not to the point of providing a phrase that

can simply supplant the word, nor even of providing paraphrases that can supplant the contexts of the word. Even if such full definition were devised, it would have to contain other theoretical words whose reduction to prior language would present the same problem.

We see from this example that failure of translatability is humdrum, and that it is not necessarily an obstacle to explanation and successful communication. On the other hand, there are doubtless plenty of cases, between radically different cultures, where failures of translation spell incomprehension in varying degrees.

A notion that has had some currency among philosophers of science is incommensurability of physical theories old and new. *Mass* for Newton and for Einstein had quite different meanings, they tell us, each intelligible only within the total physical theory in which it figures. The sticking point here is the weasel word *meaning*, a word that I avoided when I appealed rather to success in dialogue and negotiation.

Surely two theories cannot be utterly incommensurable and still qualify as science. The distinctive feature of science, after all, is its responsiveness to checkpoints in predicted observation; and the sentences that directly report observation are the very sentences that are most clearly amenable to determinate translation if the language is translatable at all. At worst, then, scientific theories are commensurable through their checkpoints.

In Praise of Observation Sentences

1993

The classical question of epistemology was handed down to us through a succession of British empiricists. The story is that we are given a barrage of sense data, or *sensa*, or *sensibilia*, or *qualia*, and somehow we eventually come out with a theory of the external world, and a theory, moreover, that meets empirical demands by fostering prediction of subsequent *qualia*. The epistemological question, then, is how we manage this and why it works.

Some philosophers of science have dismissed the question, in recent decades, viewing the very notion of an epistemological given as ill-conceived. I think of N. R. Hanson, Paul Feyerabend, Thomas Kuhn. In their view, pure observation is a chimera. Observation is always theory-laden.

I am with them in jettisoning traditional epistemology, but they have overreacted, neglecting significant distinctions. I shall argue for an intermediate position. I agree that the notion of a phenomenalistic conceptual scheme, ready and waiting for the positing of a physical world, is perverse. What are in fact noticed and remembered are usually physical things and events. These we capture in words and retain in memory, forgetting most sensory aspects of the scene. Physical conceptualization and nomenclature are the very tools of the systematic organizing of experience, the very tools of memory itself.

It was perhaps such considerations as these which prompted the shift from phenomenalism to physicalism. The physicalist pictured the observer as projecting his theory of the external world not from an internal domain of *qualia*, but from observations of the world itself. The observations were reported in protocol sentences, as they called them in Vienna, and what was then seen as the new epistemologist's job was analysis of how the pro-

tol sentences support our science. Discussions ensued among Rudolf Carnap, Otto Neurath, and Moritz Schlick as to what sentences to rate as protocol sentences.

Subsequent physicalists were impatient with even this. They shelved the whole notion of epistemological starting point, and therewith the last remnant of epistemology. What estranged them was perhaps the shakiness of the new foundation: it was not clear what sort of sentence should qualify as protocol, nor why.

But it *becomes* clear, if one is both single-mindedly physicalist and single-mindedly epistemological. Our channel of continuing information about the world is the impact of molecules and light rays on our sensory receptors; just this and some kinaesthetic incidentals. The protocol sentences should be the sentences most closely linked causally to this neural intake: most closely linked not in respect of subject matter, but physically, physiologically, neurally. They should be sentences like 'It's cold', 'It's raining', 'That's milk', 'That's a dog', to which we have learned to assent unreflectively on the spot if we are queried when certain associated sensory receptors are triggered.

The likely objection to such sentences, as vehicles of evidence for our knowledge of the external world, is that they already *assume* such knowledge: knowledge of rain, milk, dogs. But the answer is that they need not. Observation sentences—as I call them—can be conditioned outright to distinctive ranges of sensory intake, or as physicalists let us say *neural* intake. The child can be conditioned simply to assert or assent to the sentence under some distinctive stimulation, and he can come to learn only later that part of the sentence is a term denoting bodies or substances in an articulated external world.

Each of us learned some of his observation sentences in early childhood, and each of us learned most of them later. We learn some of them from other ones by analogy, recombining their parts. We learn to form compounds of simple ones, using grammatical particles. As adults we learn many more through the mediation of sophisticated theory. Thus, take the sentence 'There is some copper in the solution.' We understand it by construction from its separate words, but it *becomes* an observation sentence for a chemist who has learned to spot the presence of copper by a glance at a solution. What qualifies sentences of both sorts as observational, for a given individual, is just his readiness to assent outright on the strength of

appropriate neural intake, irrespective of what he may have been engaged in at the time.

We must recognize degrees of observationality. Assertion of the sentence or assent to it may be more or less delayed or hesitant. There may even be afterthoughts: 'Oh, it isn't a dog after all'. This sort of self-correction intrudes at the sophisticated stage where the child has come to appreciate component terms of observation sentences in their referential capacity. It is infection of observation by theory; the anti-epistemologists have a point here. But there are pure cases, and they prevail at the early stages of language acquisition. Observation sentences in this pristine purity are the child's port of entry to cognitive language, for it is just these which he can acquire without the aid of previously acquired language.

Observation sentences are *occasion* sentences, true or false from occasion to occasion; sometimes it is raining, sometimes not. And strictly speaking, we see, occasion sentences vary in degree of observationality. But it will be convenient to talk simply of observation sentences, meaning very or sufficiently observational.

What I have said by way of definition of observation sentences is only half the story, the subjective or solipsistic half. We must also impose an intersubjective condition; for the sentences are learned from other speakers who are sharing the observations. Moreover, they are to be the vehicle of evidence for objective science, intersubjectively attested. The further requirement for our definition, then, is that assent to the sentence and dissent from it must command agreement of all competent witnesses. 'Competent' here means membership in the chosen community, whether just English-speaking chemists or all speakers of the language.

The observation sentence is Janus-faced. It faces outward to the corroborating witnesses and inward to the speaker. It faces outward to its subject matter and inward to the range of neural intake that is keyed to trigger it. As a response to neural intake, the sentence is holophrastic: the neural intake is keyed to the sentence as a monolithic whole, no matter whether the sentence was first acquired by simple ostension or by excursion through theory in the manner of the chemistry example. In its outward orientation to its subject matter, on the other hand, the sentence figures not holophrastically but piecemeal, word by word. Piecemeal is how the sentence relates to scientific theory, where its words recur in new combinations and contexts.

Holophrastically is how observation sentences served as the child's entering wedge into cognitive language. Holophrastically they are the human counterpart of animal cries. Birds and apes have repertoires of distinctive calls and cries for alerting one another to distinctive dangers and opportunities. Each cry has been associated, by instinct or conditioning, to some range of neural intake. But the great difference between us and the apes in this regard is the unlimited variety of our observation sentences and the facility with which we compound further ones from them by 'and', 'or', 'with', 'beneath', and the like.

Let us now pause to pick up some dividends. I mentioned that a reason for misgivings over observation sentences was that they are theory-laden. I have conceded that observation grades off into theory. But the more basic misgivings on this score can be resolved by appreciating the difference between taking observation sentences holophrastically and taking them piecemeal. Piecemeal they are utterly theory-laden. Their terms recur in the theory to denote objects the very conception of which is pure theory. Holophrastically, meanwhile, the observation sentences are anchored to sensory neural intake irrespective of their theoretical subject matter.

The anti-epistemologists to whom I just now responded were preceded by other physicalists, Hans Reichenbach and Ernest Nagel among them, who did recognize observation language but sensed a problem of bridging the gap between it and theoretical language. They bridged it with bridge laws, or *Zuordnungsdefinitionen*. From our present point of view this, again, is no problem. Bridge laws are not called for. Theory is in logical contact with observation sentences through the sharing of terms. Our Janus-faced observation sentence faces neural intake when taken holophrastically and faces theory when taken term by term.

Granted, a word that we first acquire as a fragment of a holophrastic observation sentence may undergo some semantic creep after being extricated from its observational context. Witness 'whale' and 'fish', or again 'weight' and 'mass'. Very well, this is something the scientist adjusts for in relating his theory to his observations.

The clearing up of these perplexities brings out the importance of focusing on observation sentences rather than observation terms. Neurath and Carnap were on the right track here, in focusing on protocol sentences. But terms predominated in Carnap's further writings.

We have seen observation sentences in a variety of vital roles. They were probably the origin of language, if we think again of the cries of the apes.

Second, they are the infant's entry to language. For much the same reason, they are the radical translator's way into the jungle language. And fourth, there is what mainly concerns us on the present occasion: their role as vehicles of evidence for our knowledge of the external world, and without benefit of bridge laws.

Some of my readers have wondered how expressions that are merely keyed to our neural intake, by conditioning or in less direct ways, could be said to convey evidence about the world. This is the wrong picture. We are not aware of our neural intake, nor do we deduce anything from it. What we *have* learned to do is to assert or assent to some observation sentences in *reaction* to certain ranges of neural intake. It is such sentences, then, thus elicited, that serve as experimental checkpoints for theories about the world. Negative check points.

It takes a pair of them, ϕ and ψ , combined into a generalization, 'Whenever ϕ , ψ '; 'When it snows, it's cold'; 'Where there's smoke, there's fire'. I call this an observation categorical. First we find that our theory, or some chunk of it, logically implies some observation categorical, 'Whenever ϕ , ψ '; then we test the theory experimentally by realizing the observable condition ϕ and checking for ψ . If ψ does not result, the observation categorical is false and then so is the theory that implied it. I think this little caricature brings out the essential relation between scientific theory and its evidence.

I can now dispute another negative tenet of our anti-epistemologists. They hold that radically different theories in natural science are incommensurable. Theoretical terms owe their meanings to their theories, it is argued, and so retain no meaning for purposes of comparing the theories. Observation sentences, however, again to the rescue. They are common coin, shared reference-points for the two theories. They are free of the indeterminacies that beset translation of theoretical sentences, for they can be learned holophrastically by ostension, as in childhood and as in the linguist's first break into the jungle language. Comparing the responses of the two theories to these shared checkpoints, then, should afford traces of commensurability insofar as the two theories are under empirical control at all. Here then is a fifth credit to observation sentences.

One further thought now, and a more visionary one, on physicalist epistemology. The old phenomenalist epistemology inspired a project of rational reconstruction: the derivation of natural knowledge from sense data. The physicalist stance invites an analogous but more realistic rational re-

construction. We might tackle unsophisticated science, say natural history and the rudiments of Newtonian mechanics, and try explicitly relating at least a skeleton of it to some implied observation categoricals, invoking all the auxiliary common-sense trivialities that normally go without saying. The first big hurdle would be our space-time matrix, which is needed for identifying intermittently observed bodies across time.

Pressed eventually into higher science, such studies might reveal strands of science that are “put-up jobs,” in A. S. Eddington’s sense, but still contributory to the structure. Such studies might reveal other strands that contribute nothing and could be dropped. Unexpected simplifications and clarifications of scientific theory might emerge. Epistemology might become a tool of the natural scientist’s trade.

And now to a sixth notable role of observation sentences. They are the primitive source of the idioms of belief and other prepositional attitudes. I see, as the initial phase, the mother’s ascription of perceptions to the child. When she teaches the child the observation sentence ‘It’s raining,’ or ‘Milk,’ or applauds him when he volunteers the sentence, she has had to empathize; she has had to imagine herself in his place, facing the way he is facing, and, imagining herself thus, she must feel moved to assent to the observation sentence herself. I see this as her perceiving that the child is perceiving that it is raining or that milk is there. In transmitting the art of language from generation to generation, mothers must be able, in effect if not in words, to ascribe perceptions to their children. The ascription consists in this vicarious assent to observation sentences from the child’s vantage point.

I see the verbal ascription of perceptions as the primitive idiom of prepositional attitude. Its content clauses are observation sentences. In the fullness of time this idiom came to be extended to admit nonobservational sentences as content clauses. Eventually, by analogy, other idioms of prepositional attitude developed alongside: thus belief, indirect quotation, hope, regret, and the rest.

Recognizing observation sentences is important in yet a seventh way: their holophrastic role bears significantly on the epistemology of ontology. Their association with neural intakes, being holophrastic, is unaffected by any reassigning of objects to the terms involved. But also the logical relations of implication that connect scientific theory with observation categoricals are unaffected by one-to-one reinterpretation of terms; all that

matters to logical structure is identity and difference. We conclude that the sensory evidence for science is indifferent to what things science says there are, so long as identity and diversity among them is preserved. Reference to things in space-time, for instance, could be reconstrued as reference to the sets of number quadruples comprising the space-time coordinates of the things. So far as evidence goes, objects figure only as neutral nodes in the logical structure of our total theory of the world.

This same insight can be gained by reflecting on Alfred Tarski's definition of truth, as Donald Davidson has remarked, and the recursive definition of satisfaction that underlies it. If a sentence turns out true under Tarski's constructions, and then we reinterpret all its predicates by a one-to-one reassignment of objects, it will still turn out true.

Applied to abstract objects—numbers, say, or classes—this is not surprising. Take the class of past English monarchs. We may agree on its members, but might not your *class* of them and mine be different objects? Perhaps *membership* in your sense is *nonmembership* in mine, and then your class of past English monarchs is my class of everything else.

It is with ordinary bodies, the sticks and stones, that our downgrading of ontology is hard to take. And no wonder; we were born body-minded. Natural selection implanted this bias in our innate standards of perceptual similarity, enhancing the salience of bodily contours and configurations. Bodies are our prototype for all reification, all existence. We could never, either as a race or as growing children, have worked our way up to objective reference without reifying bodies and substances, or something pretty much to that effect. But we now see, reflecting on the avenues of scientific evidence, that one could *retrospectively* swap even the bodies for arbitrary other objects without violence to the sentences affirmed in the theory or to the evidence for them, if there were any point in doing so.

Even if we did so, the observation sentences would retain their old sensory associations. 'There's a rabbit' would still carry its fluffy, hopping connotations. What would change is the denotation of the term 'rabbit' *within* the sentence. The sentence retains its sensory associations but gets construed as an indirect or deferred reference to something else: the rabbit's proxy.

And let us bear in mind that even our common-sense view of bodies was much more theoretical and sophisticated than what natural selection thrust upon us. The big difference is our reidentification of a body on its

reappearance. Looking the same is neither sufficient nor necessary. The identification depends rather on our theory of space, time, and causality, and the elaborate trajectories of absent bodies that we conjecture within that overall theory.

So far as evidence goes, then, our ontology is neutral. Nor let us imagine beyond it some inaccessible reality. The very terms ‘thing’ and ‘exist’ and ‘real’, after all, make no sense apart from human conceptualization. Asking after the thing in itself, apart from human conceptualization, is like asking how long the Nile really is, apart from our parochial miles or kilometers.

Appendix on Neural Intake

Since the sensory receptors at the subject’s surface are physical objects on a par with more familiar things farther out, why regard them as the interface between the subject and the external world?

Interface is an artifice. Causal continuity is the fact of the matter; the causal chain from Mama or the rabbit to utterance of the observation sentence ‘Mama’ or ‘Lo, a rabbit’. Psychologists fix upon one or another point on this causal chain and call it the stimulus, to which the response is conditioned. Neural intake is the place on the chain which I have chosen for that role. It is there that I tag the chain when I want to say to what the observation sentence is conditioned.

Davidson prefers to tag the chain at its *terminus a quo*: Mama or the rabbit, the observed subject matter.¹ My reason for tagging them rather at the observer’s surface is the awkward heterogeneity of the subject matter. Usually it is not so nicely segregated as Mama or rabbits. ‘It’s raining’ and ‘It’s cold’ are more to the point, and I am reluctant to settle for *situations* as points of reference. They are of a piece with facts and propositions,

1. On page 73 of his contribution “Meaning, Truth, and Evidence” to Barrett and Gibson’s *Perspectives on Quine* (Cambridge, MA: Blackwell, 1990), Davidson calls his position a distal theory of meaning and mine a proximal one. Actually, my position in semantics is as distal as his. My observation sentences treat of the distal world, and they are rock-bottom language for child and field linguist alike. My identification of stimulus with neural intake is irrelevant to that.

The phrase ‘stimulus meaning’, in other writings of mine, is clearly the villain of the piece. It is no part of the intersubjective business of semantics; I mean it rather in physicalistic parody of the mentalistic notion of private meaning. It denotes the set (nearly enough) of perceptually similar neural intakes that suffice to prompt the subject’s assent to the sentence.

whereas nerve endings afford a clearly individuated, homogeneous domain.

Granted, this domain is still redundant for the purpose. The receptors that are triggered on an occasion that prompts a given observation sentence are mostly irrelevant to it. How about locating the subject's neural intake farther in, at appropriate places within his brain?

Much is being learned about those appropriate places in the brain. David Hubel and Torsten Wiesel have shown that visual impulses from the cells of the retina are not just transmitted as a sheaf of parallel impulses, as one imagined, to produce an isomorphic arrangement in the brain. Rather, overall traits of the scene are selected and communicated as wholes. One bank of brain cells responds exclusively to scenes in which there are some conspicuous diagonals from upper right to lower left. Other banks of cells specialize in other broad traits. Various traits thus abstracted get superimposed to kick off the appropriate response.

It seems that our responsiveness to these various visual features, so seamlessly integrated into our smooth visual experience, has evolved helter-skelter down the ages, each feature making its separate contribution to survival. A striking example is cited by Daniel Dennett in his monumental new book *Consciousness Explained*:² we are innately responsive to bilateral symmetry. In the wild, such symmetry is associated mainly with the face of a fellow creature looking squarely at us. I think of William Blake:

Tiger, tiger, burning bright
In the forests of the night,
What immortal hand or eye
Could frame thy fearful symmetry?

That symmetric presentation is an emergency calling for quick aggression or flight.

Dennett gives an illuminating overview of latter-day cognitive neurology and psychology. It emerges that the various features of the world that are reacted to by various elements in the brain never get a simultaneous synthesis. Motor output has already been innervated by one aspect of your specious present while the intake of another aspect is still in process. The information brought in on one nerve tract may induce fleeting expectations that corrupt the information brought by another nerve tract millisec-

2. Boston: Little Brown, 1992.

onds later—all within the same specious present. Perception itself becomes, in this way, theory-laden. No wonder a self-sufficient phenomenal language is hopeless.

Our concern was whether to locate the subject's neural intake deep inside his brain instead of at his sensory surface. Now we see that the pertinent neural structures and activities within the organism engender disunity and heterogeneity much as did the things and events in the organism's environment. It is *just* at the surface receptors that the great skein or sheaf of wildly multifarious causal chains from environment to behavior admits of a neat overall cross section. Each causal chain intersects the sensory surface at a sensory receptor, and the subject's sensory receptors are fixed in position, limited in number, and substantially alike. In fact the pertinent feature of each is just yes or no: triggered or not.

Such, then, is my practical reason for pegging stimulation at the subject's surface receptors rather than farther in or farther out.

I conveniently include the totality of activated receptors in the stimulus on each occasion, though only a small portion are salient enough to matter. Anatomically very diverse stimuli thus prove perceptually similar. On this score, a reassuring symmetry emerges between stimulus and response. Parallel to stimulus as neural intake, identifiable with the totality of activated receptors, the response may be taken as motor output, identifiable with the totality of activated fibers in the motor muscles, including those of speech. Then, just as perceptual similarity is the relation of effective equivalence among the redundantly inclusive stimuli, so *functional* similarity is the relation of effective equivalence among the redundantly inclusive responses.

For a striking comparison, finally, let me dip into a late phase of the traditional epistemology. In *Der logische Aufbau der Welt*³ (1928), Carnap developed a theory of sense qualities to serve as the epistemological basis of our knowledge of the external world. His basic elements were not sense qualities, but global momentary experiences. Each was the subject's experience over some brief moment. The basic relation between one of these and another was *part similarity*; nearly enough; it involved also a memory factor. From part similarity of global experiences, then, together with the apparatus of logic and set theory, he constructed the substantial beginnings

3. *The Logical Structure of the World and Pseudoproblems in Philosophy* (Berkeley: California University Press, 1967).

of a full language of sense qualities. It was all translatable into just his two basic predicates of part similarity and global experience plus the signs of logic and membership.

Observe now the striking analogy between stimulus as global neural intake and Carnap's ground elements, his global experiences. Moreover, his relation of part similarity of global experiences carries over as our perceptual similarity of global neural intakes.

It is uncanny how much better Carnap's thoughts are suited to the physicalistic, naturalistic stance than they were to his own phenomenalist stance. His "rational reconstruction," as he called it, was a sort of fiction: he was conjecturing how one could, in principle, proceed consciously from nondescript global experiences to the articulated world of qualia, using a relation conceived for the purpose plus the resources of logic and mathematics. Transplanted, on the other hand, his moves take on empirical and causal inevitability. Neural intake is physical and indisputable. Perceptual similarity is the basis of all learning, all habit formation, and it is testable in people and other animals by the reinforcement and extinction of conditioned responses.

Truth

1994

1. Vehicles of Truth

What are true or false, it will be widely agreed, are propositions. But it would not be so widely agreed were it not for ambiguity of ‘proposition’. Some understand the word as referring to sentences meeting certain specifications. Others understand it as referring rather to the meanings of such sentences. What looked like wide agreement thus resolves into two schools of thought: for the first school the vehicles of truth and falsity are the sentence, and for the second they are the meanings of the sentences.

A weakness of this second position is the tenuousness of the notion of sentence meanings. The tenuousness reaches the breaking point if one is persuaded of my thesis of the indeterminacy of translation, which I shall not go into here. Even apart from that thesis, it seems perverse to bypass the visible or audible sentences and to center upon sentence meanings as truth vehicles; for it is only by recurring to the sentence that we can say which sentence meaning we have in mind.

There was indeed a motive for pressing to the sentence meanings. Many sentences in the same or different languages are deemed to be alike in meaning, and distinctions among them are indifferent to truth; so one narrowed the field by ascribing truth rather to the meanings. This motive would be excellent if the notion of sentence meaning were not so elusive. But as matters stand we fare better by treating directly of sentences. These we can get our teeth into.

There was also a second motive, equal and opposite to the first, for pressing on to the sentence meanings; namely, that one and the same sentence can be true on some occasions and false on others. Thus ‘The Pope will visit Boston’ was true but turned false after his last visit. ‘I have a headache’ is true or false depending on who says it and when. Ambiguity or

vagueness of terms, also, can cause the truth value of a sentence to depend in part on the speaker's intention.

Propositions, thought of as sentence meanings, were the meanings exclusively of sentences of a firmer sort, not subject to such vacillations; what we may call *propositional* sentences. The obvious response, then, is that the propositional sentences themselves can serve as the truth vehicles. Just think of 'I', 'you', 'he', 'she', 'here', and 'there', as supplanted by names and addresses or other identifying particulars as needed. Think of tenses as dropped; we can use dates, the predicate 'earlier than', and the like as needed. Think of ambiguities and vaguenesses as resolved by paraphrase—not absolutely, but enough to immobilize the truth value of the particular sentence. The truth values need not be known, but they must be stable.

The attitude is one that is familiar in the teaching of logic. When we take illustrative sentences from everyday language and paraphrase them into the notation of truth functions and quantifiers, we think of the reference of demonstratives and personal pronouns as fixed—albeit tacitly—and we never dream of reading ' $\exists x$ ' as 'there was' or 'there will be something x '.

Declarative sentences thus refined—propositional sentences—are what I shall regard as truth vehicles in ensuing pages, for the most part. On the whole it is the convenient line for theoretical purposes. We must recognize, though, that it bypasses most of what counts in daily discourse as true or false, since our utterances are not for the most part thus refined. The truth vehicles directly related to behavior are not sentences as repeatable linguistic forms but rather the individual acts of uttering them. These are for the most part univocal in truth value without benefit of paraphrase. There are just occasional failures, perhaps because some name turns out to be empty or because some vague term turns out to be indeterminate just where it matters for the utterance in question. Such utterances may be dismissed as neither true nor false.

So much by way of coming to terms with the realities of verbal behavior. Let us now return to the more conveniently manageable domain of prepositional sentences, whose truth or falsity, known or unknown, is unchanging.

2. Truth as Disquotation

Such being what admit of truth, then, wherein does their truth consist? They qualify as true, one is told, by corresponding to reality. But correspondence word by word will not do; it invites the idle clattering of real-

ity with a bizarre host of fancied objects, just for the sake of correspondence. A neater plan is to posit *facts*, as correspondents of true sentences as wholes; but this still is a put-up job. Objects in abundance, concrete and abstract, are indeed needed for an account of the world; but facts contribute nothing beyond their specious support of a correspondence theory.

Yet there is some underlying validity to the correspondence theory of truth, as Tarski has taught us. Instead of saying that

‘Snow is white’ is true if and only if it is a fact that snow is white

we can simply delete ‘it is a fact that’ as vacuous, and therewith facts themselves:

‘Snow is white’ is true if and only if snow is white.

To ascribe truth to the sentence is to ascribe whiteness to snow; such is the correspondence, in this example. Ascription of truth just cancels the quotation marks. Truth is disquotation.

So the truth predicate is superfluous when ascribed to any given sentence; you could just utter the sentence. But it is needed for sentences that are not given. Thus we may want to say that everything someone said on some occasion was true, or that all consequences of true theories are true. Such contexts, when analyzed logically, exhibit the truth predicate in application not to a quotation but to a pronoun, or bound variable.

The truth predicate proves invaluable when we want to generalize along a dimension that cannot be swept out by a general term. The easy sort of generalization is illustrated by generalizing on the term ‘Socrates’ in ‘Socrates is mortal’; the sentence generalizes to ‘All men are mortal’. The general term ‘man’ has served to sweep out the desired dimension of generality. The harder sort of generalization is illustrated by generalizing on the clause ‘time flies’ in ‘If time flies then time flies’. We want to say that this compound continues true when the clause is supplanted by any other; and we can do no better than to say just that in so many words, including the word ‘true’. We say ‘All sentences of the form “If p then p ” are true’. We could not generalize as in ‘All men are mortal’, because ‘time flies’ is not, like ‘Socrates’, a name of one of a range of objects (men) over which to generalize. We cleared this obstacle by *semantic ascent*: by ascending to a level where there were indeed objects over which to generalize, namely linguistic objects, sentences.

Semantic ascent serves also outside of logic. When Einstein propounded

relativity, disrupting our basic conceptions of distance and time, it was hard to assess it without leaning on our basic conceptions and thus begging the question. But by semantic ascent one could compare the new and old theories as symbolic structures, and so appreciate that the new theory organized the pertinent data more simply than the old. Simplicity of symbolic structures can be appreciated independently of those basic conceptions.

As already hinted by the correspondence theory, the truth predicate is an indispensable intermediary between words and the world. What is true is the sentence, but its truth consists in the world's being as the sentence says. Hence the use of the truth predicate in accommodating semantic ascent.

The disquotational account of truth does not define the truth predicate—not in the strict sense of 'definition'; for definition in the strict sense tells how to eliminate the defined expression from every desired context in favor of previously established notation. But in a looser sense the disquotational account does define truth. It tells us what it is for any sentence to be true, and it tells us this in terms just as clear to us as the sentence in question itself. We understand what it is for the sentence 'Snow is white' to be true as clearly as we understand what it is for snow to be white. Evidently one who puzzles over the adjective 'true' should puzzle rather over the sentences to which he ascribes it. 'True' is transparent.

For propositional sentences the disquotational account of truth is neat, we see, and simple. It is readily extended, moreover, to the work-a-day world of individual utterances; thus an utterance of 'I have a headache' is true if and only if the utterer has a headache while uttering it.

3. Paradox

It seems paradoxical that the truth predicate, for all its transparency, should prove useful to the point of indispensability. In the matter of paradox, moreover, this is scarcely the beginning. Truth is notoriously enmeshed in paradox, to the point of out-and-out antinomy.

An ancient form of the antinomy of truth is the Paradox of the Liar: 'I am lying', or 'This sentence is not true'. A looser and fancier version was the paradox of Epimenides the Cretan, who said that all Cretans were liars. The underlying antinomy can be purified for logical purposes to read thus:

- (1) 'yields a falsehood when appended to its own quotation' yields a falsehood when appended to its own quotation.

Executing the instructions in (1), we append the nine-word expression to its quotation. The result is (1) itself. Thus (1) says that (1) itself is a falsehood. It is thus tantamount to 'I am lying', but more clean-cut. It hinges only on the innocuous operations of quoting and appending and the notion of falsehood, which reduces to an innocent 'not' and *true*. The truth predicate is clearly the trouble spot. The inevitable conclusion is that the truth predicate, for all its transparency and seeming triviality, is incoherent unless somehow restricted.

For further explicitness a technical turn of phrase will be convenient. The truth predicate will be said to *disquote* a sentence *S* if the form:

——— is true if and only if ———

comes out true when *S* is named in the first blank and written in the second. Thus what the disquotational account of truth says is that the truth predicate disquotes every prepositional sentence. But the lesson of the antinomy is that if a language has at its disposal the innocent notations for treating of quoting and appending, and also the notations of elementary logic, then it cannot contain also a truth predicate that disquotes all its own propositional sentences—on pain of inconsistency. Its truth predicate, or its best approximation to one, must be incompletely disquotational. Specifically, it must not disquote all the sentences that contain it. That was the trouble with (1). And of course it must not disquote all the sentences containing terms by which that predicate could be paraphrased. This, except for its special orientation to the ingredients of (1), is substantially what has come to be known as Tarski's Theorem. He has proved harder things.

The truth predicate loses little in general utility thereby, for it can still disquote all the prepositional sentences that do not themselves contain it or other expressions to the same effect. And even these excluded applications can be accommodated by a hierarchy of truth predicates. The hierarchy begins with a predicate 'true₀', which disquotes all sentences that contain no truth predicate or equivalent devices. A predicate 'true₁', next, disquotes all sentences that contain no truth predicate or equivalent devices beyond 'true₀'. And so on up. It is a hierarchy of progressively more nearly perfect truth predicates. The plan dates back in a way to the early phase of Russell's theory of types, by which he meant to obstruct the Paradox of the Liar among others.

4. Tarski's Construction

We saw that disquotation is loosely definitive of truth. We may now be thankful for the looseness, seeing as we do that definability of truth for a language within the language would be an embarrassment. And thus it was that Tarski undertook the perilous adventure of defining it for the language within the language, as nearly as possible, if only to see what minimum obstacle saved the situation. This was not his order of presentation, but it comes out the same.

The language chosen for the construction contains the logical notations for quantification and the truth functions and the set-theoretic notation ' $x \in y$ ' for membership. It contains also a finite lexicon, as large as you please, of predicates for natural science and daily life. Finally, it contains the means in effect of quoting and appending, as in (1); that is, it can specify each of its single signs and it can express the concatenation of expressions.

Truth pertains to closed sentences, that is, sentences without free variables. Its analogue for open sentences is the two-place predicate of *satisfaction*. An assignment of objects to variables *satisfies* a sentence if the sentence is true for those values of its free variables.

What sort of object is an assignment of objects to variables? It is simply a function, or one-many relation, relating one and only one object to each variable—that is, to each letter, ' w ', ' x ', ' y ', ' z ', ' w ', etc. A relation, in turn, is a set, or class, of *ordered pairs*. Ways are well known of defining the notation ' $\langle x, y \rangle$ ' of ordered pairs contextually by means of epsilon and the logical particles.

Once satisfaction is defined, truth comes easily; for a closed sentence, having no free variables, is vacuously satisfied by all assignments or none according as it is true or false. We can simply define

(2) ' y is true' as ' $\forall x(\text{Assignment } x \rightarrow x \text{ satisfies } y)$ '.

So Tarski's big job is to define satisfaction. First he defines it for *atomic* sentences, each of which consists of just a predicate adjoined to one or more variables. For instance, an assignment satisfies the atomic sentence ' $x \in y$ ' if and only if what is assigned to the letter ' x ' is a member of what is assigned to the letter ' y '. Correspondingly for each of the other predicates in the lexicon. An assignment satisfies an alternation of sentences, next, if and only if it satisfies one or both of them; it satisfies their conjunction if and only if it satisfies both; and it satisfies a negation if and only if it does

not satisfy the sentence that is negated. Finally, an assignment satisfies an existential quantification ' $\exists x(\dots x \dots)$ ' if and only if some assignment, matching that one except perhaps for what it assigns to ' x ', satisfies ' $\dots x \dots$ '.

Such is Tarski's recursive or inductive definition of satisfaction. It explains satisfaction of atomic sentences outright, and it explains satisfaction of sentences of each higher grade of complexity in terms of satisfaction of their components. Universal quantification is passed over because it is expressible in terms of existential quantification and negation in familiar fashion.

5. Paradox Skirted

Clearly all the clauses of this inductive definition can be formulated within the formal language itself, except for the word 'satisfies' that is being defined. Thus we have apparently defined satisfaction for the language within the language. Invoking (2), then, we have done the same for truth. This was supposed to spell contradiction.

We could even get contradiction directly from satisfaction, without the detour through (2), 'truth', and (1). We have merely to ask whether assignment of the sentence 'not (x satisfies x)' to the variable ' x ' satisfies the sentence 'not (x satisfies x)' itself. Such is Grelling's so-called Heterological Paradox.

What saves the situation is that the definition of satisfaction is inductive rather than direct. The inductive definition explains satisfaction of each specific sentence, but it does not provide a translation of ' x satisfies y ' with variable ' y '. Consequently it does not support the truth definition (2) for variable ' y '; it just explains truth of each specific closed sentence. It leaves the truth predicate in the same state in which the disquotational account left it; namely, fully explained in application to each specific sentence of the given language but not in application to a variable.

It was a near miss, and I turn now to a nearer one. Treating relations again as classes of ordered pairs, we can write ' $\langle x, y \rangle \in z$ ' to mean that x bears the relation z to y . Now imagine the above inductive definition of satisfaction written out in our formal language, with the variable ' z ' always in place of 'satisfies' and so ' $\langle x, y \rangle \in z$ ' in place of ' x satisfies y '. Let the whole inductive definition, thus edited, be abbreviated as ' Φz '. It fixes z as the satisfaction relation. Evidently we arrive thus at a direct definition:

$$(3) \exists z(\Phi z \cdot \langle x, y \rangle \in z)$$

of ‘ x satisfies y ’ strictly within the formal language itself. Doesn’t this spell contradiction?

No. The catch this time is that there might not be any relation z such that Φz . Indeed there better not be, on pain, we see, of contradiction. The two-place predicate ‘satisfies’ remains well defined in its inductive way, but a grasp of the predicate and how to use it carries no assurance of the existence of a corresponding abstract object, a corresponding set of ordered pairs. And, failing such a pair set, (3) fails to translate ‘ x satisfies y ’. Though the satisfaction predicate is well explained even within the formal language by the recursion, it does not get reduced to the prior notation of that language. Satisfaction, and truth along with it, retain the status that truth already enjoyed under the disquotatation account: clear intelligibility without full eliminability.¹

6. Interlocked Hierarchies

The inductive definition fully explains what it is for an assignment to satisfy a sentence. That there is such a satisfaction relation, then, or pair set, is sheer common sense. The paradoxes of set theory, however—Russell’s, Burali-Forti’s, Cantor’s—have overruled the common-sense notion that clear membership conditions assure the existence of a class, a set. All of those were paradoxes ultimately of the membership predicate ‘ $\in$ ’; what is striking about the present case is just that we find set theory responding also to paradoxes of truth and satisfaction.

Since the paradoxes of the membership predicate barred naive set theory, which would countenance a set for every formulable membership condition, set theories have come only in varied strengths, with no strongest, on pain of paradox. This multiplicity of set theories can be seen as a multiplicity of varieties of membership, each determining its peculiar range of sets. There is an analogy here to the hierarchy of truth predicates contemplated in §3. Now what our recent reasoning shows, regarding a pair set z such that Φz , is just that the particular variety of membership expressed by the ‘ $\in$ ’ of the formal language there under consideration is one that generates a set theory devoid of a pair set z such that Φz . A similar formal language with a differently interpreted epsilon, more prolific in sets,

1. The foregoing analysis is adapted from my *Philosophy of Logic*, 35–46. A somewhat different analysis, in my 1952 paper “On an Application of Tarski’s Theory of Truth,” is called for when the set theory is of the kind that admits both sets and ultimate classes.

might well countenance a pair set (in *its* sense) precisely matching the satisfaction predicate of that previous formal language but not matching its own satisfaction predicate. Satisfaction for the first language will thus be directly definable by (3) in the second language, not just inductively; and similarly for truth. The analogy between a hierarchy of epsilons and a hierarchy of truth predicates thus turns out to be by no means mere. They are of a piece.

In his early version of his theory of types, mentioned at the end of §3, Russell sought to block the paradoxes both of truth and of membership by decreeing a single complex hierarchy of predicates. The scheme was vague and cumbersome. It was subsequently sharpened and simplified for purposes of set theory by dispensing with the truth aspect as extraneous to set theory. And now we see a new interlocking of the membership hierarchy with the truth hierarchy, along clean-cut lines and for clear but subtle reasons that could not be foreseen in Russell's day.

We can picture the hierarchy of membership predicates and the hierarchy of truth predicates as embraced within a single inclusive language. If we do, we must beware of assuming that the truths at each level constitute a set, a class; for that could enable us to define a single comprehensive truth predicate after all, by quantifying over those classes.

Alternatively, we can picture a hierarchy of languages, each with its unique membership predicate and a unique truth predicate for the next lower language. This approach has an important mathematical application in establishing relative strengths of formal systems. One system is proved to be stronger than another by interpreting its predicates in such a way as to be able to define, within it, the truth predicate of the other system.

7. Excluded Middle

Pilate was probably not the first to ask what truth is, and he was by no means the last. Those who ask it seek something deeper than disquotation, which was the valid residue of the correspondence theory of truth (§2). Yet there is surely no impugning the disquotation account; no disputing that 'Snow is white' is true if and only if snow is white. Moreover, it is a full account: it explicates the truth or falsity of every clear sentence, with all the clarity that is enjoyed by the sentence itself. It is even a more than full account: it imposes a requirement on the truth predicate that is too strong for any predicate within the language conceived—on pain of contradiction (§3).

Let us look into some of the seemingly deviant notions about truth. One such, traceable to Aristotle, is that a prediction is neither true nor false until events have occurred that causally determine it. Theologians have favored the doctrine. If contingent predictions were true now, they reason, the events would be determined now by God's knowledge, and hence would not be contingent. The consequent determinism, it is felt, would leave no place for man's moral responsibility.

This doctrine, for all its bizarreness, is no repudiation of the disquotational account of truth. If it is not yet *true* that there will be a sea-fight tomorrow—to take Aristotle's example—then it is a mistake to say now that there will *be* a sea-fight tomorrow; for as of now the contingent sentence is neither true nor false. The logic, granted, is deviant: the law of excluded middle is suspended pending causal determination. But the disquotational character of truth remains.

Bizarreness remains too. There is the abandonment of the law of excluded middle. Worse, the propositionalizing of tensed sentences (§1) is barred when they are contingent predictions. Happily, however, the theological argument underlying this desperate move is inconclusive on two points. One is the assumption of an omniscient God. The other is the notion that universal determinism precludes freedom of action. We are free and responsible, it can be argued, in that we act as we choose to; whether our choices are determined by prior causes is beside the point.

Some other apparent challenges to the law of excluded middle are, in part, not what they seem. Let it be clear, to begin with, that ignorance of the truth or falsity of a sentence is par for the course and quite in keeping with its being true or false. Further, it commonly happens that a sentence can be rendered prepositional in divergent ways, reflecting a speaker's intentions in different situations. Here it is rather the respective utterances that are true or false, together with their full unambiguous prepositional elaborations if we care to elaborate them. The original ambiguous sentence is indeed then neither true nor false, but this need not be seen as a breach of the law of excluded middle; it is better seen as a propositionally incomplete expression that has still to be filled out in one or another way. This line was not available in the case of the theologians' strictures on contingent predictions, because the predictions were meant still to become true or false once they stopped being contingent.

There is another and stronger case that likewise threatens the law of excluded middle. It is where a purported name or singular description fails to designate anything. When a sentence contains such a term, one possible

line is, as before, to drop the sentence from consideration; treat it as meaningless. This line is awkward when we regiment our sentences for logical purposes, for the existence of the object may be an open question—as in the case of Camelot or Prester John or the outermost satellite of Pluto. It is quite in order for the truth value of a sentence to remain an open question, but it is inconvenient to leave the very meaningfulness of a sentence forever unsettled.

One might accordingly relinquish the law of excluded middle and opt rather for a three-valued logic, recognizing a limbo between truth and falsity as a third truth value. What then comes to hinge on existence of Camelot, or whatever, is truth value rather than meaningfulness, and that is quite in order. But a price is paid in the increased cumbersomeness of three-valued logic. Alongside ‘not’, which sends truths into falsehoods, falsehoods into truths, and now limbo into limbo, there would be a truth function that sends truths into limbo, limbo into falsehoods, and falsehoods into truths; also three more such one-place truth functions, playing out the combinations—as contrasted with a single one, negation, in two-valued logic. When we move out to two-place truth functions (conjunction, alternation, and their derivatives), proliferation runs amok. It can still be handled, but there is an evident premium on our simple streamlined two-valued logic.

We can adhere to the latter, in the face anyway of the threat of empty singular terms, by simply dispensing with singular terms. They can all be paraphrased into predicates and variables. ‘Camelot is fair’ becomes ‘ $\exists x(x$ is Camelot and x is fair)’. It does not go into limbo; it simply goes false if it is false that $\exists x(x$ is Camelot). The predicate ‘is Camelot’ is seen on a par with ‘is fair’, as a predicate irreducibly. The expedient is inspired by Russell’s treatment of singular descriptions.

8. Truth Versus Warranted Belief

There are recurrent references to a coherence theory of truth, or a pragmatist theory of truth. The question that motivates this quest beyond disquotation can perhaps be phrased thus: if to call a sentence true is simply to affirm it, then how can we tell whether to affirm it?

The lazy answer is ‘That all depends on what the sentence is. In the case of “Snow is white” you just look at snow and check the color.’

The more sympathetic answer is a general analysis of the grounds of

warranted belief, hence scientific method. The most primitive propositional sentences of science, I would suggest, are ones that correlate observables: sentences of the form 'All so-and-so's are thus and so' or 'Wherever this, that' or 'Whenever this, that' where so and so, thus and so, this and that are directly recognizable to the satisfaction of all competent witnesses. These are what I call *observation categoricals*. Examples are 'Crows are black', 'Turtles have shells', 'Where there's smoke there's flame', 'When the sun comes up the birds sing'. An observation categorical is refuted by observing a so and so that is not thus and so; by observing this without that; smoke without flame.

Scientific theory above the primitive level of observation categoricals owes its empirical content to the observation categoricals that it implies. Here we see the essence of the experimental method. Some portion of scientific theory, a bit of chemistry perhaps, is found to imply that whenever a particular experimental situation is set up, such and such should happen. What is implied is an observation categorical: 'Whenever this, that.' The experimental arrangement has to be observable, to the satisfaction of the experimenter and his colleagues, and the predicted result has to be observable.

Our more casual beliefs about the world, quite apart from deliberate experimental science, likewise have their empirical content in the observation categoricals that they imply. Our most trivial disappointment—'I had fully expected to find it there'—can be analyzed as the refutation of some tacit observation categorical that was implied by our beliefs.

When the scientist sets up an experiment, he is bent on testing some one hypothesis, newly devised or perhaps newly questioned. The observation categorical that the experiment enacts is not, however, in most cases logically implied by the targeted hypothesis in isolation. A backlog of auxiliary beliefs is almost always needed to clinch the implication. Some of them will be tacit truisms of common sense, some will be theoretical, some perhaps even mathematical.

Such is the important insight of *holism*, stressed by Pierre Duhem. It is congenial to proponents of a vaguely so-called coherence theory of truth, but it is better seen as a theory of evidence, or of warranted belief. Seen either way, it seems uncongenial to a line currently urged by Michael Dummett.

Dummett questions the law of excluded middle (§7) on epistemological grounds. The attack was mounted in mathematics by L. E. J. Brouwer early

in this century, and Dummett adopts the attitude toward science in general. His rough idea is to reckon a sentence of natural science neither true nor false if no procedure is known for making a strong empirical case for its truth or falsity.

Holistic considerations make it doubtful what sentences to exclude. Clear candidates for retention are the observation categoricals, but other sentences share empirical content in varying degrees by implying observation categoricals only jointly. It seems vain to seek an invidious distinction between sentences eligible for truth or falsity and sentences in limbo, unless we either draw that boundary at the observation categoricals themselves or else draw it at the far extreme to exclude just those sentences that never imbibe any empirical content by participating in the joint implying of observation categoricals. The expanse between these empty sentences and the observation categoricals embraces the whole of serious scientific theory and most of our responsible beliefs of a more casual kind as well.

Truth is one thing, warranted belief another. Clarity can be gained and the sweet simplicity of two-valued logic can be enjoyed by heeding the distinction. Sentences utterly beyond the reach even of shared empirical content do remain unwelcome, but they can in large part be banished outright from our scientific language, if grammatical and lexical constraints to the purpose can be drawn up without prohibitively complex gerrymandering.

9. Truth in Mathematics

Even mathematical truths share, we saw (§8), in the empirical meaning of sciences where they are applied. They do so by participating with other sentences in jointly implying observation categoricals. But note that this cannot be said of logical truths. Any sentence already logically implies any logical truth, and thus gains no further implying power by being conjoined with it. Logical truths differ from mathematical ones on this point simply because I am taking logical implication in a narrow sense of 'logic'—the logic of truth functions, quantification, and identity.

What now of those parts of mathematics that share no empirical meaning, because of never getting applied in natural science? What of the higher reaches of set theory? We see them as meaningful because they are couched in the same grammar and vocabulary that generate the applied parts of mathematics. We are just sparing ourselves the unnatural gerrymandering

of grammar that would be needed to exclude them. On our two-valued approach they then qualify as true or false, albeit inscrutably.

They are not wholly inscrutable. The main axioms of set theory are generalities operative already in the applicable part of the domain. Further sentences such as the continuum hypothesis and axiom of choice, which are independent of those axioms, can still be submitted to the considerations of simplicity, economy, and naturalness that contribute to the molding of scientific theories generally. Such considerations support Gödel's axiom of constructibility, ' $V = L$ '.² It inactivates the more gratuitous flights of higher set theory, and incidentally it implies the axiom of choice and the continuum hypothesis. More sweeping economies have been envisioned by Hermann Weyl, Paul Lorenzen, Erret Bishop, and currently Hao Wang and Solomon Feferman, who would establish that all the mathematical needs of science can be supplied on the meager basis of what has come to be known as predicative set theory.³ Such gains are of a piece with the simplifications and economies that are hailed as progress within natural science itself. It is a matter of tightening and streamlining our global system of the world.

10. Equivalent Theories

A multitude of unlike scientific theories can be counted on to be empirically equivalent, in the sense of implying all the same observation categoricals. This is obvious, since the mere set of all observation categoricals implied by a given theory is itself an empirically equivalent theory, however cumbersome and unimaginative.

Theories can differ utterly in their objects, over which their variables of quantification range, and still be empirically equivalent. There is a simple argument from what I call proxy functions.⁴ Given any rule for assigning a new and distinct object fx to each object x of a given theory, we can immediately get an empirically equivalent theory whose objects are the new ones. We have merely to reinterpret the predicates correspondingly; 'dog' become ' f of a dog'. The two theories are readily seen to be empirically

2. See my *Set Theory and Its Logic*, 2nd ed., 234–238.

3. See *Quiddities*, 34–36.

4. See *Theories and Things*, 19–22.

equivalent; moreover, they are convertible, one into the other, simply by applying the proxy function or its converse everywhere. We hardly seem warranted in calling them two theories; they are two ways of expressing one and the same theory. It is interesting, then, that a theory can thus vary its ontology.

Effort and paper have been wasted, by me among others, over what to count as sameness of theory and what to count as mere equivalence. It is a question of words; we can stop speaking of theories and just speak of theory formulations. I shall still write simply 'theory', but you may understand it as 'theory formulation' if you will.

Theories (theory formulations) can be logically incompatible and still be empirically equivalent. A familiar example is Riemannian and Euclidean geometry as applied to the surface of a sphere. Riemannian geometry says that straight lines always meet. Euclidean geometry says that some do and some do not, and in particular that there are none on a sphere. The conflict is resolved by reinterpreting 'straight line' in the Riemannian glossary as 'great circle'.

The next example, due to Poincaré, is less trivial. We have on the one hand our commonsense conception of infinite space and rigid bodies that move freely without shrinking or stretching, and on the other hand the conception of a finite spherical space in which those bodies shrink uniformly as they move away from center. Both conceptions can be reconciled with all possible observations; they are empirically equivalent. Yet they differ, this time, more deeply than in the mere choice of words. The theory with the finite space makes crucial use of a theoretical term that admits of no counterpart in the theory with the infinite space—namely, 'center of space'.

Imagine now two global theories of the world, one of them our own, such that we are persuaded of their empirical equivalence, but we see no way of systematically converting one into the other by reinterpretation sentence by sentence, as we did in the examples of the proxy function and the sphere. There are three cases to consider.

Case 1: The other theory is logically compatible with our own and is expressed directly in our own terms. It differs from ours only in that it implies some theoretical sentences that ours leaves unsettled, and vice versa. Yet the implied observation categoricals are the same for both; such is their empirical equivalence. This case presents no problem. We would simply

accept the other theory and incorporate it into our own as an enrichment, answering many theoretical questions that ours left open.

Case 2: Again the other theory is logically compatible with ours, but it hinges on some theoretical terms not reducible to ours.

Case 3: The two theories are logically incompatible. Donald Davidson showed me that this case can be reduced to Case 2 by the following maneuver. Take any sentence *S* that the one theory implies and the other denies. Since the theories are empirically equivalent, *S* must hinge on some theoretical term that is not firmly pinned down to observable criteria. We may then exploit its empirical slack by treating that term as two terms, distinctively spelled in the two theories. *S* thus gives way to two mutually independent sentences *S* and *S'*. Continuing thus, we can make the two theories logically compatible.

11. Irresolvable Rivalry

So we may limit our attention to Case 2. We are imagining a global system of the world empirically equivalent to our own and logically compatible with ours. It may seem that as staunch empiricists we should reckon both theories as true. Still, this line is unattractive if the other theory is less simple and natural than ours; and indeed there is no limit to how grotesquely cumbersome a theory might be and still be empirically equivalent to an elegant one. We do better, in such a case, to take advantage of the presence of irreducibly alien terms. We can simply bar them from our language as meaningless. After all, they have not contributed to a satisfactory empirical theory, any more than 'phlogiston' did, or indeed 'fate', 'grace', 'nirvana', 'mana'. We thus consign the theory to the limbo of nonsentences.

We have here an interesting encroachment of coherence considerations upon standards of truth. Simplicity and naturalness, by whatever standards, are making the difference between truth and meaninglessness.

If, on the other hand, the rival theory is every bit as neat and natural as our own, our empiricist scruples reawaken. Should we then incorporate that theory into our own, as in Case 1? No, this would ill accord with the scientists' quest for simplicity and economy; for the irreducible new terms imported with the annexed theory have added no new coverage of observables. The two theories were already empirically equivalent to each other, and hence to their conjunction. The two theories were streamlined and

neck-and-neck, but the tandem theory is loaded beyond necessity: loaded with all the sentences containing the new terms.

Such considerations would not oppose enriching our original theory with any novel findings of the other theory that do not use the novel terms. It would be a matter of welcoming information from a presumed dependable outside source, much as supplementary truths of number theory are got by excursions through analysis, or as the four-color-map question was settled by an elaborate computer program.

As wholes, however, with their untranslatable key terms, the two theories are perhaps best seen as two true but partly incommensurable descriptions of the real world. A resilient scientist would still be free to oscillate between them, for the sake of added perspective from which to triangulate on his problems.

If, unknown to him, the two theories imply only true observation categoricals, and admit of no improvement on the score of simplicity, then empiricism can make no sense of further demands for their correspondence to the real world. It is a unique world, still, and a describable one, but groping at it with our human terms we grasp it variously. I think of the disparate ways of getting at the diameter of an impenetrable sphere: we may pinion the sphere in calipers, or we may girdle it with a tape measure and divide by π , but there is no getting inside.

In receiving both theories as true we would be thinking of truth as defined for the sentences of an inclusive language, in which both theories were expressed. The language would thus include those terms of either theory that had no place in the other. For the inclusive language to be self-sufficient, we should elect the first of the two options at the end of §6, thus recognizing a hierarchy of truth predicates and membership predicates within the language. The variables of the inclusive language would range over all the objects of both theories, if the theories differ in their ontologies. We may then think of the range of variables within each theory as expressly limited by a distinctive predicate.

I have been picturing two theories, but there could be more. The sky is the limit, since we are here embarked on fantasy. Even a single well-defined, close-knit, comprehensive theory of nature is something still to aspire to—let alone two empirically equivalent and equally elegant ones. And even the discovery of their equivalence would be no small matter, failing reducibility of one to the other sentence by sentence (though the equivalence does shine through in Poincaré's simple example, §10).

The fantasy of irresolubly rival systems of the world is a thought experiment out beyond where linguistic usage has been crystallized by use. Hence my vacillation, in recent years, over what line to take on it. Still we have had to cope with it, as a challenge to our concept of truth.

Acknowledgments

In §11 my position on inelegant rival theories was influenced by Dagfinn Føllesdal, and my position on elegant ones by Donald Davidson. The thought in the last paragraph is due to Burton Dreben, whose suggestions have been helpful also at many other points.

References

- Dummett, Michael. 1978. *Truth and Other Enigmas*. Cambridge, MA: Harvard University Press.
- Gödel, Kurt. 1940. *The Consistency of the Continuum Hypothesis*. Princeton: Princeton University Press.
- Poincaré, Henri. 1929. *Science and Hypothesis*, ch. IV. New York: Science Press.
- Quine, W. V. 1952. "On an Application of Tarski's Theory of Truth." *Proc. Nat Acad. Sci.* 38:430–433. Reprinted in Quine, *Selected Logic Papers* (New York: Random House, 1965).
- . *Set Theory and Its Logic*. Cambridge, MA: Harvard University Press. 2nd edition, 1969.
- . 1970. *Philosophy of Logic*. Englewood Cliffs, NJ: Prentice Hall. 2nd edition, Cambridge, MA: Harvard University Press, 1986.
- . 1981. *Theories and Things*. Cambridge, MA: Harvard University Press.
- . 1987. *Quiddities*. Cambridge, MA: Harvard University Press.
- Russell, Bertrand. 1908. "Mathematical Logic as Based on the Theory of Types." *American Journal of Mathematics* 30:222–262. Reprinted in Jean van Heijenoort, ed., *From Frege to Gödel* (Cambridge, MA: Harvard University Press, 1967).
- Tarski, Alfred. 1956. "The Concept of Truth in Formalized Languages." In A. Tarski, *Logic Semantics, Metamathematics* (Oxford: Clarendon). 2nd edition, Indianapolis: Hackett, 1983.

Promoting Extensionality

1994

Logic is the technology of deduction. Its business is to show how, from a given set of premisses, to derive conclusions that can be depended on to be true if the premisses are true. Part of the technique, even in antiquity, was regimentation. Regimentation figured mainly, at first, in matching up clauses of the premisses.¹ Often a clause in one of the premisses admits of recasting, without violence to the speaker's or the writer's purposes, in such a way as to match it with a clause of another premiss. Each new duplication thus created is vital to the deductive potential.

It is in symbolic logic that regimentation takes on modern efficiency: that of prefabrication and mass production. A few basic and ubiquitous idioms are singled out, each of which can accommodate a vast range of clauses when these are suitably paraphrased. These few forms stand open in their symbolic rigidity waiting for content to be poured in at the user's pleasure. The fewer the forms, the greater the incidence of eventual duplication among the clauses of a set of premisses and hence the richer the deductive yield. Such is the premium on economy of primitive notation.

Symbolic logic in its neoclassical form is predicate logic. In its austere form, it imposes just two idioms, or grammatical constructions: a truth function and the universal quantifier, together with variables for referring back to the quantifier. The mold is built up of these devices in iteration. The content to be poured into the mold consists of predicates of one or more places.

In practice it is convenient and customary to relax this stark economy. Symbols for five truth functions are admitted, though all can be reduced to

1. I was alerted to this central, but largely neglected, aspect by Sherry 1991.

one by paraphrase. Existential quantification is added, though reducible to universal and negation. From the unlimited stock of predicates to be drawn on for content, one in particular is appropriated as peculiarly logical: the two-place predicate '=' of identity. Finally, the gates are opened to other ingredients besides predicates as admissible content; namely, constant singular terms as well as functors such as 'plus', 'times', and 'log' for forming complex singular terms. Each of these could be reduced to a corresponding predicate, with help of truth functions, quantification, and identity.

The matching up of clauses is no longer the main point of regimentation. Equally important is the streamlining of canons of deduction. In earlier times this scarcely mattered; the rules of the syllogism added little to the unimplemented light of pure reason. With the rise of symbolic logic, however, algorithms and complete proof procedures in rigorous formalization assumed vital importance, both for the philosophical understanding that they have afforded and for their practical utility. The burgeoning computer industry testifies to the latter.

On the philosophical side, the regimentation embodied in predicate logic has also brought illumination quite apart from the technology of deduction. It imposes a new and simple syntax on our whole language, insofar as our logic is to apply. Stripped down to the austere economy that I first described for predicate logic, our simple new syntax is as follows. The parts of speech are: (1) the truth-functional connective, (2) the universal quantifier, (3) variables, and (4) atomic predicates of one and more places. The syntactic constructions are: (1) application of a predicate to the appropriate number of variables to form a sentence; (2) prefixure of a quantifier, with its variable, to a sentence; and (3) joining sentences by the truth-functional connective and then adjusting parentheses.

I hesitate to claim that this syntax, so trim and clear, can accommodate in translation all cognitive discourse. I can say, however, that no theory is fully clear to me unless I can see how this syntax would accommodate it. In particular, all of pure classical mathematics can be thus accommodated. This is putting it mildly. The work of Whitehead and Russell and their predecessors and successors shows that the described syntax together with a single two-place predicate by way of extra-logical vocabulary, namely the ' $\in$ ' of class membership, suffices in principle for it all. Even '=' is not needed; it can be paraphrased in terms of ' $\in$ '.

What makes for the surpassing clarity of theories couched in this syntax

is their *extensionality*. This means, in part, that if within a true sentence we supplant a true component sentence by another truth, or a false component by a falsehood, the containing sentence will remain true. Further, if we supplant a component open sentence by another that is fulfilled by just the same values of the variables, the containing sentence will still remain true.

The useful idioms of propositional attitude—‘*x* believes that *p*’, ‘*x* hopes that *p*’, ‘*x* says that *p*’, and the rest—are in conspicuous violation of extensionality. We may well believe that *p* and not that *q*, though both be true.

What is worse, even scandalous, is that these idioms violate the substitutivity of identity: the putting of equals for equals. How can something be true and false of the same thing under different names?

Yet these idioms are useful to the point of indispensability. Moreover, I think they are rooted in the earliest stages of language. I picture the earliest idiom of propositional attitude as ‘*x* perceives that *p*’, where ‘*p*’ stands for an observation sentence such as ‘It’s raining’, ‘That’s milk’, ‘That’s a dog’. When the mother is monitoring the child’s utterance of such a sentence, she has to empathize with him. She imagines herself in his place, facing in the same direction, and then checks whether she, thus oriented, feels moved to volunteer the sentence herself. In short, she checks, however inarticulately, whether the child really perceives that it’s raining, that it’s milk, that it’s a dog. This much in the way of an idiom of propositional attitude, all unspoken, is essential to the very handing down of language from generation to generation; for observation sentences are the child’s entering wedge to language. And then, down the ages, the idiom ‘*x* perceives that *p*’ was extended to non-observational sentences, and, by analogy, the other idioms of propositional attitude emerged.

Gottlob Frege, confronted with the paradoxical failure of substitutivity of identity, concluded that in those idioms the recalcitrant terms have changed their reference and taken to referring to what would normally be their meanings, or senses, rather than their normal objects. A better solution is suggested by the mother’s relation to the child in monitoring his observation sentence; namely, empathy. When someone ascribes a propositional attitude to someone, he impersonates that person to some degree. The subordinate clause of the construction is uttered from the subject’s point of view, somewhat as if from the subject’s mouth. No wonder substitutivity of identity fails; the subject, poor fellow, didn’t know the things

were identical. Likewise for failure of extensionality: the subject would have been unprepared to interchange the two coextensive clauses in question, simply because he didn't know they were coextensive.

Along with the failure of extensionality and the failure of substitutivity of identity, in clauses of prepositional attitude, there is a third failure: failure of quantifying in. A quantifier outside the clause cannot bind a variable inside the clause, at least not without raising problems of interpretation. The quantification

(1) $\exists x$ (Ralph believes that x is a spy)

raises an ontological problem of the value of ' x ', since Ralph trusts Bernard J. Ortcutt but suspects a man in a brown hat who is in fact Ortcutt. The difficulty, again, is just discrepancy between the real world, to which the outlying ' $\exists x$ ' relates, and the empathized world—Ralph's—in which the recurrence of ' x ' is trapped.

Clarity is gained by banishing propositions, as objects of belief and other prepositional attitudes, and settling somewhat unnaturally for the sentences themselves, the mere strings of letters or phonemes. This, after all, was where it all began, according to my conjecture: in the observation sentences that the mother felt moved to affirm when empathizing the child's attitude. Let us construe the conjunction 'that' of the prepositional-attitude construction simply as a quotation mark.

It does take an obvious bit of adjusting. 'He believes that he is Napoleon' becomes

He believes 'I am Napoleon',

since the believed sentence is conceived as in the subject's mouth. In the case of indirect discourse, moreover, which is an idiom of prepositional attitude, confusion with direct quotation would have to be avoided. We might agree to use 'say' for the indirect and 'utter' for the direct.

In taking the objects of the attitudes as sentences, I do not presume that the creature in the attitude speaks the language of the sentence, or any language. The sentence is in the language of the ascriber of the attitude, but he frames it from what he takes to be the subject's point of view.

The failure of extensionality, the failure of substitutivity of identity, and the failure of quantifying in all become natural and obvious when the objects of the attitudes are seen as sentences, set off in quotation. The quota-

tion simply designates the depicted string of letters or phonemes, unrelated syntactically to the outlying context.

We may seem to have paid a price for this clarification. We may seem to have taken leave of the clear and simple syntax of predicate logic by admitting a potential infinity of sprawling quotations as unanalyzed singular terms. However, not so. Quotations can be reduced to the syntax of predicate logic by spelling them out. What we need are just names of the several signs or letters of our alphabet, together with a term functor—analogue to ‘plus’—to express the concatenating of one letter or string of letters with another to form a longer string. Given this much, we can define any quotation by spelling it out letter by letter. When at last we are bent on reducing everything to the *minimal* syntax of predicate logic, we reduce the letter names and the term functor to predicates, with help of ‘=’, as noted earlier. All in all, first by quotation and then by spelling, we digest the prepositional attitudes syntactically and logically. Extensionality prevails, as well as the substitutivity of identity and the syntax of predicate logic.

Semantically, however, the propositional attitudes remain low grade. The objective criteria for whether x believes that p , regrets that p , etc., are heterogeneous, varying radically with ‘ p ’, and, as often as not, they are indecisive. This is something to put up with, pending progress of some sort or other; but the logical and the syntactical barriers are gone.

In all this, however, I have been taking the prepositional attitudes only *de dicto*. That is to say, I have been construing the content clause as the ascriber’s attempt to speak strictly from the subject’s point of view. The other interpretation, *de re*, permits the ascriber’s own voice to intrude on his simulation of the subject’s voice. If we recognize the *de re* alternative, the sentence

(2) Ralph believes that Ortcutt is a spy

becomes ambiguous. *De dicto* it is false, since Ralph denies the sentence ‘Ortcutt is a spy’, but *de re* it is true, because the subject in the brown hat is Ortcutt.

More precisely, what may be taken *de re* or *de dicto* is not the whole ascription of the propositional attitude, nor yet the content clause, but only one or another place within that clause: one or another place occupied by a singular term or a variable—thus the place occupied by ‘Ortcutt’ in (2). A term in *de re* position, or what I have called *referential* position, is subject

to substitutivity of identity after all, and a variable in that position can be quantified from outside. Referential positions are where the ascriber's world has intruded.

If the content clause of an ascription of a prepositional attitude contains referential positions, therefore, the content clause can no longer be represented in quotation. Accordingly, in 1956 I proposed analyzing such cases by means rather of quoted predicates, thus freeing the referential positions from quotation.² (2) becomes:

Ralph believes 'is a spy' of Ortcutt,

when taken *de re*. The referential position is thus exposed to substitutivity and also to quantification,

$\exists x$ (Ralph believes 'is a spy' of x),

which captures (1).

It seems important to make sense thus of (1) and to contrast it with the trivial

Ralph believes ' $\exists x$ (x is a spy)'

(1) suggests that some trait or circumstance, some description, singles out the suspect in Ralph's mind. But this will not serve, as Robert Sleigh showed in his example of the shortest spy.³ There is probably a unique shortest, and Ralph surely believes that the shortest spy is a spy; but this does not make for the political significance of (1). Nor should it matter that there might be more than one shortest spy.

I now despair of a coherent theory of propositional attitudes *de re*, despite appearances.⁴ When security agents are told (1), they are only given a lead and not a coherent message. Following the lead, they interrogate Ralph and record some straight *de dicto* beliefs, all nicely within the bounds of our extensional scientific framework.

So I feel we have seen extensionalism safely through the perils of the prepositional attitudes. But threats remain from other quarters, notably the modal logic of necessity and possibility.

2. Quine 1956.

3. Sleigh 1968.

4. But see Burdick 1982. He accommodates them extensionally by citing additional open sentences as tacit parameters.

Modal logic differs from the propositional attitudes in that all term positions are referential, or potentially so. Quantification into modal contexts is unrestrained, but substitutivity of identity in modal contexts works for some terms and not for others. It is a question now not of position, not of where, but of what. Hence Dagfinn Føllesdal's genuine singular terms, or Saul Kripke's rigid designators. These are the terms that obey substitutivity of identity even in modal contexts. These are the terms, also, that support inference by existential generalization, even in modal contexts; other terms do not. As we might say in a modal spirit, these are the terms that name their objects necessarily. They name them on the core of essential traits, not accidental ones.

Modal logic fails of extensionality, and up to a point it fails of substitutivity of identity. Must I curb and compromise my extensionalism after all? My answer this time is that I have not been sold on the notion of necessity, the distinction between necessity and contingency, which is what modal logic in its standard interpretation and primary motivation is about.

The adverb 'necessarily' in its everyday use is another matter. Typically it serves merely to mark a statement that one's interlocutor is presumed to agree with, in contrast to whatever matters are under discussion or investigation at the time. This use is clear and convenient, but it has no place in logically regimented science, any more than other occasion-dependent locutions such as 'I', 'you', 'here', 'now', and the tenses of verbs.

The pioneer ventures in modal logic undertaken a century ago by Hugh MacColl and more economically in 1918 by C. I. Lewis were motivated by dissatisfaction with the truth-functional conditional. They wanted a stronger connective: 'necessarily if'. What was lurking in the background was the subjunctive conditional, and there is no denying that it pervades scientific thinking. It is implicit in the design of every experiment: 'If this and this were set up, such and such would occur'. However, the universally quantified truth-functional conditional covers such cases. The experimenter's conjecture is a general one: 'Whatever this and this obtain, such and such occurs'. Then he tests this general conjecture by setting this and this up and watching for such and such.

The contrafactual case of the subjunctive conditional, however, must be treated as elliptical and occasion-dependent. A universally quantified truth-functional conditional is again implicit, I think, but with a complex antecedent some of whose clauses are left tacit, to be divined from the con-

text and circumstances. It is along such lines that I would make sense of subjunctive conditionals, contrafactual and otherwise, without compromising extensionality.

Disposition terms such as 'soluble'—are naturally associated with the contrafactual conditional; thus, 'If this were in water, it would dissolve'. My view of disposition terms, however, is that they do not ascribe a special sort of property, a potentiality, but that they just name ordinary properties in a special way, namely, by alluding to a fairly dependable and convenient symptom or test. Solubility in water is a physical property analyzable, like others, in microphysical terms, but an easy and generally reliable test of it is immersion in water to see if the stuff dissolves. Again, fragility is explicable microphysically, but a fairly reliable test is to drop the object and see if it breaks.

The relation of cause and effect has been seen as a challenge to extensionality, but again I think I see a reconciliation. I see the relation as relating events, and I see events as physical objects in my broad sense. Each event, each physical object, is just the total content of some portion of space-time. If someone whistled a certain tune all the way to the bus, and only then, I am prepared to identify his whistling that tune with his walking to the bus; they are the same event. So, if his walking caused his arrival at the bus stop, his whistling caused it. This is odd usage, but harmless, for I do not generalize on it. Causal laws relate classes of events, as usual, and whistling a tune does not usually cause an arrival.

Finally, how does probability fare under extensionality? Where the probability is statistical, we can settle for a statement of the statistical basis itself as the ratio of sizes of classes of cases. Where the probability is subjective, i.e., the degree of expectation, we are up against issues of behavioral or physiological analysis of mental phenomena, irrespective of extensionality. The neat behavioral measure of a subjective probability is the minimum acceptable odds at a wager.

Enigmatic as subjective probability may be, the crucially enigmatic locus of probability is quantum mechanics. It seems that something like probability invests the behavior of elementary particles irreducibly. This problem I leave to the scientists, eagerly awaiting further progress on their part and enlightenment on mine. There has long been discussion of revisions even of logic for the simplification and clarification of quantum theory, and I can conceive that extensionality might not remain immune. I can only hope, and leave it to the future to speak for itself.

References

- Burdick, Howard. 1982. "A Logical Form for the Propositional Attitudes." *Synthese* 52:185–230.
- Quine, W. V. 1956. "Quantifiers and Propositional Attitudes." *Journal of Philosophy* 53:177–187.
- Sherry, David. 1991. "The Conspicuous Role of Paraphrase." *History and Philosophy of Logic* 12:151–66.
- Sleigh, Robert. 1968. "On a Proposed System of Epistemic Logic." *Noûs* 2:391–389.

Indeterminacy without Tears

1994

Misgivings regarding the idea of meaning, and indeed the underlying idea of idea, led me to ponder radical translation. By this I mean production of a manual of translation of a hitherto unknown language on the basis of field work, unaided by bilinguals or by kindred languages. Pondering it is a way of isolating the ultimate evidence for linguistic meaning, namely, observable native behavior, verbal and otherwise, in observable circumstances. The test of a good result, a good manual, is fluent dialogue and successful negotiation.

The manual will of course countenance many translations of one and the same sentence, for within our own language we have various equivalent ways of saying things. However, over and above this, I have hazarded the following indeterminacy thesis: two manuals could meet the dialogue and negotiation tests impeccably and yet disagree as to the admissible translations of some native sentences. They would produce translations in our language that were not equivalent to each other. I based the conjecture on the limitations of possible evidence in the native's verbal behavior and its observable circumstances, immediate and remote. The philosophical point of my conjecture is the untenability of the notion of propositions, or sentence meanings.

In stating the thesis just now I invoked a relation of equivalence, applied to sentences within our own language. I must eliminate this, for it is too close to the notion of sentence meanings that I am challenging. The following, then, is an aseptic restatement of the imagined conflict between the two impeccable manuals. We have a recording of a long native discourse, and we have smooth, clear English translations of it from the two

manuals; but if we mesh the translations, alternating them sentence by sentence, the result is incoherent.

This thesis does not impugn translation. There are good manuals of translation and bad ones. In particular both manuals of my thought experiment are impeccable, so their conflict would obtrude on no user of either unless he happened to collide with a user of the other in construing one of the discrepant sentences. Certainly the thesis does not support anti-rational relativism or cast doubt on the reality of the world as science increasingly reveals it. Experimental method is still the way to knowledge, and predicted observation is its touchstone.

Finally a word of counter-caution: these reassuring remarks must not be taken to suggest that every language is wholly translatable into every other, determinately or not. Some sentences of our current English are untranslatable even into the English of 1900, because of such neologisms as 'quark' and 'neutrino'. For note that these have never been defined, in the strict sense; they cannot be paraphrased outright into previously available expressions, nor can their contexts. They are just explained, and well enough so that we can use them. Similarly for exotic languages: they can all be learned, but translation is not the only way in.

Assuming Objects

1994

I am going to talk about talking about things. I'll be talking about talking about bodies, substances, particles, numbers, classes—whatever there is to talk about. I'll examine the notion of object: what it comes to, how it arises, and how it contributes to our systematic understanding of nature.

In trying to get to the bottom of things, or objects, let us look back to the beginning of language and a bit beyond: to the cries of the apes. Apes have a repertoire of distinct signals for distinct purposes. One signal warns the fellow apes of the approach of a lion or leopard, another warns of an eagle, another perhaps reports the sighting of some fruit trees while the troop is ranging through the forest. Each distinctive cry has become keyed to some distinctive range of stimulations, whether by instinct or conditioning. In each of these examples the source of the stimulation is a body: a beast, a bird, a fruit. This is prophetic, for reification begins with bodies. Things are bodies first and foremost. But reification is not yet.

Natural selection has heightened our responsiveness and that of other animals to the sight and smell of bodies. There is survival value in promptly discerning something to eat, or something to avoid being eaten by. What sets bodies apart from other stimuli? Typically a body contrasts with its visual surroundings in color and in movement or parallax, and typically it is fairly chunky and compact. If it is animate and seen full face, it is bilaterally symmetric. If we make contact, it resists pressure. It is merely such traits, at first, that distinguish bodies from the glow of the evening sky, the feel of a cool breeze, or other details of the passing show.

The apes' cries are not language. What is distinctive about language is its critical mass, recursively generating sentences without limit, in chain reaction. But the apes' cries, few and simple, do foreshadow language. I imag-

ine that language developed from them. Certainly the child's first linguistic acquisitions of a cognitive kind are on a par with the ape's cries: each is an expression that he has learned to associate directly with some range of stimulation. It is the only way into language. Likely early acquisitions are 'Mama', 'Doggy', 'Milk', '(It's) cold', and after a while 'It's getting dark', 'It's raining'. I call them observation sentences.

Some of them—'Mama', 'Doggy', 'Milk'—are nouns, we see, denoting objects, and the child learned them from adults for whom they were just that. I shall call them observation terms, but for the child they must be seen still as observation sentences, keyed like the rest to ranges of neural intake. They simply report episodes in the passing show, on a par with 'It's cold' or 'It's raining'.

Our observation sentences far outrun the apes' cries in their multiplicity and in the wealth of distinctions they draw. The child learns to compound new ones from old by use of connectives: 'and', 'or', 'after', 'beside', etc. One of these primitive grammatical constructions, in particular, is significant as a first step toward reification of bodies. Applied to the simple observation sentences 'Black', or 'That's black', and 'Dog', or 'That's a dog', it forms the compound 'Black dog', or 'The dog is black'. As an observation sentence it can be phrased either way, since terms are not yet recognized as denoting. I call this primitive grammatical construction *observational predication*, in anticipation of the mature predication into which it will evolve.

Although 'Black dog' and 'The dog is black' are not to be distinguished at this stage, observational predication is quite another matter from mere conjunction, 'Black and dog'. The conjunction describes any scene in which black and dog are both salient, whereas 'Black dog' requires that they be situated together, the black patch engulfing the canine patch. The predication expresses the compact clustering of visual qualities that is characteristic of a body.

What I picture as the linguistic development next beyond such observation sentences is a generalized expression of expectation. All learning, all habit formation, depends on an innate propensity to expect similar events to be followed by events that are similar to each other. It is primitive induction. Conditioning depends on it: if one action has an unpleasant effect, we abstain from a similar action lest it have a similar effect.

This expression of expectation, as I picture it, joins two observation sentences to express the general expectation that whenever the one observation sentence holds, the other will be fulfilled as well. Examples: 'When it

snows, it's cold'; 'Where there's smoke, there's fire'; 'When the sun rises, the birds sing'; 'When lightning, thunder'. They are our first faltering scientific laws. I call them *observation categoricals*.

In the evolution of language, and also in the child's learning of it, the leap from ordinary observation sentences to observation categoricals was a giant one. I would like to know how it came about and how it is learned. It was a vital development, for observation categoricals are the direct expression of inductive expectation, which underlies all learning.

The survival value of apes' cries, and of our ordinary observation sentences, lay in vicarious observation; we learn things that only someone else is in a position to observe. Now the observation categoricals bring us much more. They bring us vicarious habituation, vicarious induction. One gets the benefit of generalized expectations built up over the years by some veteran observer or even by his informant long dead. Observation categoricals can be handed down.

Even at this stage there is no denotation, no reference to bodies or other objects, to my way of reckoning. The observation categorical just asserts concomitance or close succession of separately specified phenomena. The child's observation terms 'Mama' and 'Doggy' at this stage still merely register repeatable features of the passing show, on a par with 'Cold' and 'Thunder'. Any difference here is only qualitative, not ontological.

What am I holding out for? Just the traditional sort of categorical, 'All acorns are bitter'. We can almost get it with the observation categorical 'Whenever there is an acorn, there is a bitter acorn', but not quite. The latter is compatible with there being sweet acorns as long as they keep close company with bitter ones. To get 'All acorns are bitter' we must change the second 'there' in our observation categorical to the pronoun 'it': 'Whenever there is an acorn, *it* is a bitter acorn'; 'All acorns are bitter'.

The pronoun is an added link between the component observation sentences 'Acorn', or 'Lo, an acorn', and 'Bitter', or 'Lo, bitter'. It posits common carriers of the two traits, acornhood and bitter. The carriers are acorns, bodies. I see this pronominal construction as the root of objective reference.

An observation categorical of this strengthened sort, 'Whenever . . . , it . . .', I call a *focal* observation categorical. The earlier ones are *free* observation categoricals. The focal one amounts to ' $\forall x(\text{Acorn } x \rightarrow \text{Bitter } x)$ ', and in reckoning it as the inception of reification I am reiterating my old dictum that to be is to be the value of a variable.

Reification can be seen as a leap from what Peter Geach calls pronouns of laziness to essential pronouns. In the context 'I bumped my head and it hurts' the pronoun 'it' is one of laziness. It could be supplanted by its grammatical antecedent, 'my head': 'I bumped my head and my head hurts'. In the focal categorical 'Whenever there is an acorn, it is bitter', on the other hand, the pronoun is essential. Supplanting it by its grammatical antecedent, 'an acorn', delivers a weaker message: 'Whenever there is an acorn, an acorn is bitter'. This is the free observation categorical again, compatible with sweet acorns. 'All acorns are bitter' requires the focal categorical, and hence the essential pronoun, and hence reification.

Multiple presentation—several acorns at a time—is at the bottom of it all. Multiple presentation itself can be taken in stride without all that, for we can learn 'Apple', 'Two apples', 'Three apples' each outright by direct ostension, and likewise 'Acorn', 'Two acorns', 'Three acorns', and then produce new cases by analogy. But multiple presentation of a bitter and a possibly non-bitter acorn is what made our free categorical fall short of 'All acorns are bitter'. Multiple presentation, again, is what calls for indefinite singular terms—'an acorn', 'any acorn', 'each acorn', 'every acorn', 'some acorn'. Solitary presentations could be managed with 'the acorn' or simply 'acorn', on a par with 'thunder' or 'sunrise'. Indefinite singular terms, in turn, are the grammatical antecedents that call for essential pronouns, which finally are the quantified variables of logic and the logical vehicles of objective reference. We read ' $\exists x$ ' and ' $\forall x$ ', after all, as '*something*' and '*everything* is such that'.

To the lay mind the paradigms of objective reference are definite singular terms; proper names like 'George' and 'Boston' and descriptions like 'this acorn' and 'the author of *Waverley*'. This is the wrong place to look. At the observational level these terms behave on a par with words like 'thunder' and 'cool', and there is no deep distinction between when we are reifying and when not. Reification comes unequivocally with multiplicity and the indefinite singular terms and essential pronouns, or variables.

The focal categorical stands to the free categorical much as the observational predication 'Bitter acorn' stands to the simple conjunction 'Bitter and acorn'. The focal categorical and the predication inject the bitter into the acorns, while the free categorical and the conjunction merely have them sharing a specious present. The focal categorical generalizes the predication, while the free categorical generalizes, in a way, the conjunction.

In origin, however, the focal categorical may just be a direct offshoot of

the free categorical by an easy conflation. Thus consider again our two categoricals, free and focal, about acorns. The free one required only that some of the acorns present on any occasion be bitter, while the focal one required them all to be bitter. But in practice these demands are equivalent, because we can always narrow our sights to any single acorn, creating an occasion where it must be bitter by either categorical. This expedient would work for just about any observable bodies, and the transition from free to focal could have passed unnoticed. As for the essential pronoun, in a language like ours it could already have been there as a pronoun of laziness, ready for inadvertent enlistment.

I find this aetiology amusing, in pinning the reification of bodies, the very dawn of the notion of object, on a confusion. We shall come to a further case of this sort. They are frequent in biological evolution; a disused organ or accidental growth becomes diverted in a later generation, by natural selection, to a vital new use. Boons, like people, are to be valued for themselves and not for their origins. Serendipity and opportunism are laws of nature. I had a reason for preferring the bizarre example 'All acorns are bitter' to the well-worn 'All ravens are black'. 'Raven', when serving someone as a primitive observation sentence, is keyed to a range of visual stimuli that are perceptually similar for him. Would an albino raven be similar enough to qualify? Probably not, whereupon 'All ravens are black' collapses into vacuous analyticity. The distinction between analytic and synthetic truths is reasonably clear when applied thus to a given speaker's observation categoricals, although dubious when applied to theoretical sentences across the board. So I resorted to 'All acorns are bitter', which as an observation categorical is certainly not analytic, since acorns are recognized visually without regard to taste.

Incidentally my example is not true; there are sweet species of oak. But as an example of observation categorical it is all the more typical for being false.

It may be well here to sort out the subjective and objective aspects of an observation categorical for a given speaker. Just what environmental situations would qualify as fulfilling either of its clauses, for him, depends on his usage of the expressions, and this depends on his private standards of perceptual similarity. These factors are embodied in the present fairly stable but evolving state of pertinent portions of his neural network. It is this lingering physical state that determines what his verdict on an observation sentence *would* be, under specified environmental conditions, regardless of

whether those conditions are ever going to be realized. The dispositional 'would' here alludes to a present physical state of nerves whereof one symptom is the subject's ready assent to the sentence, or dissent, when he is appropriately tested.

Reification (to resume the thread) was man's main resource in connecting loose ends of experience to produce a structured system of the world. Linguistically what is linking them is the 'it' that links the two clauses of a focal categorical, or, more vividly, the repetition of 'x' in ' $\forall x(Fx \rightarrow Gx)$ '. But thus far I have been speculating only on early reification, which is weak in the time dimension. It still makes no sense of saying that this raven is the one we saw yesterday, or that it is not. I am still imagining a stage of language that is limited to the specious present; to short-term memories and expectations. If at this stage of our development we recognize a particular body after a long absence, it is only by habituation to its distinctive observable traits. Distinguishing a body from an utterly similar one at a later time is what calls for the next big development. It requires acquisition of our elaborate schematism of space and time and the unobserved trajectories of bodies within it. A silver cup now viewed may be exactly similar to one viewed years ago, without being the same one; and the original one may have tarnished and changed deceptively over the same period. Settling the question of identity would involve research and speculation about the movement of the cup or cups in our absence.

Our framework of space and time and our individuation of bodies across time were intimately interdependent developments. We determine sameness of place from one time to another by qualitative sameness of fixed bodies from the one time to the other. But also, conversely, we determine the fixity of a body holistically by a tacit maxim of minimum mobility on the part of the overall environment. Trajectories of bodies through time, finally, determine and are determined by causal laws that comprise our theory of the world.

Besides depending in general on conjectured trajectories and causal chains, the question of identity of a body from one time to another is generally meaningless apart from the choice of a governing predicate. Viewed four-dimensionally, a body is the content of a portion of space-time; and some spatial or temporal portions of the body may qualify as bodies in turn. The tadpole is a body and the supervening quadruped is another body, and together they add up to a third body, a lifelong animal, a frog.

By the time language has reached this sophisticated stage, its stock of

predicates will have far outrun the observation sentences. New terms, irreducible to observation, are generated by the growing scientific theory. The relative clause emerges, which encapsulates in a single complex predicate anything we can say about an object. By substituting relative clauses for the predicate letters 'F' and 'G' in our old focal categorical structure 'All F are G', we get the full power of the universally quantified conditional: 'For all x , if x is such and such then x is thus and so'.

Reification continued beyond bodies and substances, adding new sorts of objects and new quantified conditionals to tighten the logical structure of science. Atoms tentatively entered the ontology already in antiquity, in easy analogy to the primordial bodies. Abstract objects were reified even earlier, it would seem, though there was nothing so suggestive in the way of analogy. Perhaps the first abstract objects to be assumed were properties, thanks again to a serendipitous confusion: a conflation again of essential pronouns with pronouns of laziness. Granted, the ancient Greeks were not speaking our language, but theirs was near enough.

Here is the scenario. A zoologist describes some peculiarity in the life style of a strange invertebrate, and then adds "it is true as well of the horse-shoe crab." His 'it' is a pronoun of laziness, saving him the trouble of repeating himself. But let him and others conflate it with an essential pronoun, and we have them dreaming up a second-order predicate such as 'property' or 'attribute' for a new brand of object, quantified over as values of variables.

Again a happy confusion, if confusion it was. Science would be hopelessly crippled without abstract objects. Even so earthy science as natural history would suffer without reification of species. A species name, say 'dog', does not commit us to species as objects, but we are committed to them when we say that species are seldom cross-fertile. We are quantifying over them. And when we move into higher levels of the harder sciences, numbers and other abstract objects bid fair to steal the show. Mathematics subsists on them, and serious hard science without serious mathematics is hard to imagine.

Even so, the pioneer abstracts objects, which I take properties to have been, are *entia non grata* in my book. There is no entity without identity, and the identity of properties is ill defined. Properties are sometimes distinguished even though they are properties of entirely the same things; and there are no clear standards for so doing.

However, the utility that made properties such a boon can be retained

by deciding to equate properties that are true of all the same things, and to continue to exploit them under another name: classes.

Once the ice was broken, further abstract objects flooded in. Numbers, most conspicuously. Saying that there are nine planets does not yet reify the number nine, though in an explicitly regimented theory it may be seen to reify the class of the planets. But we have reified numbers when we state laws of computation, or even ask whether nine is prime.

To reify something is to recognize it as a value of a variable or pronoun, and a repeatable value, even if individually unspecified. We have seen it at work forging sameness of reference between clauses. That is what it is for. The function of entity is recurrence in discourse. It began by forging joint reference on the part of the clauses of a focal observation categorical. Next it joined forces with the budding schematism of space and time, to forge identity of an object from one place and time to another. Here the contribution was reciprocal: recurrences of enduring bodies linked up the place-times as well as vice versa.

But there remains a third question of sameness of reference. Besides sameness of reference in the two clauses of a focal observation categorical and sameness of reference across time, there is intersubjective sameness. Bodies present no evident problem here: intersubjective agreement is established in primary cases by ostension, nearly enough, and indirectly in other cases by reduction to those primary cases through causal chains established or conjectured in our growing science. I shall come back to this.

When we come to abstract objects, however, there is no such recourse. Who is to say whether what you refer to as the number nine is the same thing as what I refer to by that phrase? We both say that it is the size of the class of the planets,¹ but who is to say that your two-place predicate 'size of' relates the same things as mine does? For that matter, there is no saying whether your class of the planets is the same as mine, even granted that we agree on each planet. Your class of the planets has the same members, in your sense of 'member', as mine does in my sense of 'member', but there the matter is left hanging. I submit that intersubjective sameness of reference makes no sense as applied to abstract objects. Practically it comes down to successful dialogue.

Much the same is true even of concrete objects when we move beyond the reach of ostension and consider theoretical ones such as elementary

1. Quine wrote this before Pluto lost its planetary status.

particles. To revive an old example of mine, picture two physicists pondering a crisis in particle physics. Each of them proposes a new particle. One proposes a particle without rest mass, and the other a particle with rest mass. Both of them apply the same new word, 'neutrino'. Query: are they disagreeing about the mass of neutrinos, or are they positing different particles under the same name? Clearly it is an empty question.

Regarding gross bodies, on the other hand, I said a few minutes I ago that intersubjective sameness presents no *evident* problem; also that I would come back. Consider now the scientific philosopher who is seeking a close-knit and economical system of the world. He identifies bodies with the portions of space-time that they occupy. Similarly for substances and portions of them. Sensible qualities of the bodies and substances simply carry over to these four-dimensional regions. A body is now a sinuous filament of space-time, long perhaps in the time dimension, colorful to the eye and resistant to the touch. We would say not that bodies have given way to space-time, but that they have been identified with or interpreted as space-time.

Still no problem. Next, however, our philosopher identifies a point of space-time with the quadruple of numbers that are its coordinates in some arbitrary frame of reference. It is a matter of economy, since he needs the numbers in his science anyway. A body then becomes a class of quadruples of numbers—what I shall call a *number table*. So likewise for any substance, however scattered and diffuse. Our sensory associations now carry over to these abstract number tables.

This still need not be thought of as an extravagant reinterpretation of terms that formerly denoted familiar things. We learned names and predicates for the things by ostension, and can continue to do so. As always, the pointing just serves to enhance the salience of the features of our sensory input that are to be associated with the object, albeit a number table. We need not know any numbers.

This example came of two steps that we can suppose motivated by considerations of ontological economy: identification of physical objects with spatiotemporal regions and identification of these with number tables. However, serious motivation is beside the point. Thus we might reinterpret every reference to a physical object arbitrarily as a reference rather to its cosmic complement: the rest of the physical universe. The old names and predicates would be introduced by ostension as usual, but it would be deferred ostension: pointing to what was not part of the intended object.

Sensory associations would carry over similarity. The word 'rabbit' would now denote, not each rabbit, but the cosmic complement of each, and the predicate 'furry' would now denote, not each furry thing, but the cosmic complement of each. Saying that rabbits are furry would thus be reinterpreted as saying that complements-of-rabbits are complements-of-furry things, with 'complements-of-rabbits' and 'complements-of-furry' seen as atomic predicates. The two sentences are obviously equivalent.

'Cosmic complement of' and 'number table of' express what I call *proxy* functions: one-to-one reinterpretations of objective reference. They leave the truth values of sentences undisturbed, as these two examples did. It is a matter of reconstruing all terms and predicates as denoting, or true of, the *proxies* of what they had denoted. A term that had named an object x now names the proxy of x , and a predicate that had been true of x is now true of the proxy of x . No big deal; we are proxying both sides of the predication, and it cancels out. We appreciated the triviality where the proxy was the cosmic complement.

I have stated the matter by reference to one-place predicates, but the many-place predicates must of course be reinterpreted place by place in like fashion.

So we have found that a set of sentences can be reinterpreted in any one-to-one way, in respect of the things referred to, without falsifying any of the sentences. Such is ontological relativity, as I have called it, or the indeterminacy of reference. What is relative about it is that the ontology to be ascribed to the speakers of an alien language comes to be relative to the chosen manual of translation, from among empirically equivalent manuals.

There are just two caveats. One is that for dependability the reinterpretation should be one-to-one. Mothers are poor proxies, for a man need not be a doctor in order that his mother be the mother of a doctor. The other is that the chosen reinterpretation must be applied throughout the language, or so much of it as is to be taken into consideration for the purposes in hand.

More particulars are in order regarding this ontologically startling triviality. When we reconstrue physical objects as their cosmic complements, what happens to their classes? They get automatically reconstrued through the mandatory reinterpretation of the two-place membership predicate ' $\in$ ' as applied to physical objects.

For another example we might reconstrue classes themselves across the

board, supplanting each by its logical complement and leaving individuals alone. Again the requirement of reinterpreting predicates, in this case the membership predicates again, preserves the truth values of all sentences.

A familiar and less trivial example at the abstract level is the use of Gödel numbers for strings of signs; one need not care which of the two media one is dealing with. The indeterminacy of reference is evident also from Tarski's celebrated definition of truth in terms of what the terms of the language denote. Those familiar with Tarski's construction will appreciate that the sentences it accounts true will continue to be accounted true under any one-to-one transformation of the denoted objects.

For abstract objects, F. P. Ramsey was already illustrating the indeterminacy of reference sixty-odd years ago with what we have come to call Ramsey sentences. The underlying thought is that abstract objects serve natural science only through formal laws that we can set down. Certain of those laws might serve one scientific context, and further ones might serve another. The laws needed for either context might not determine those abstract objects uniquely and no matter. All that matters is that there be some abstract objects or other fulfilling those laws. For instance there are no end of ways of interpreting natural numbers in set theory, each of which yields classical arithmetic. Ramsey's idea, then, in the case of natural numbers, is to waive the choice of interpretation and just write an existential quantification to the effect that there is a sequence of abstract objects fulfilling such and such arithmetical laws and functioning thus and so in the desired application.

Ramsey's treatment thus brings out indeterminacy of reference not by reinterpretations, but by waiving the choice of interpretation. However, the idea does not extend, in general, to concrete objects. The trouble is that each Ramsey sentence needs to be existentially quantified afresh, and consequently there is no assurance of sameness of object from sentence to sentence. In the case of abstract objects, this does not matter; they can be happily dismissed after each application and introduced anew for the next. For concrete objects, on the other hand—gross bodies, particularly—endurance and identity over time are essential to the very fabric of our spatio-temporal account of things. Thus it is that the structuralism proclaimed by Ramsey was a structuralism of abstract objects, and he claimed no more. On the other hand my simple argument for indeterminacy of reference applies to objects indiscriminately.

I conclude from it that what matters for any objects, concrete or ab-

stract, is not what they are but what they contribute to our overall theory of the world as neutral nodes in its logical structure. This conclusion should not be viewed as devastating. Reification is human handiwork, or human mindwork. Its utility for science lies in forging logical bonds throughout our manmade theory of the world. This service is structural and hence invariant under one-to-one transformations.

As Frege appreciated, and Bentham hazily before him, it is whole sentences that mean; sentences or strings of sentences. Their truth is our objective, and their observable consequences are our checkpoints. It is a commonplace of our scientific theory itself that our data about the world reach us only through the triggering of our nerve endings. This is not a counsel of subjectivism, thanks to our occasional checkpoints in observation. We can still ask what there is in the world, but it is a question couched in our homemade concepts, and we can ask no better than to be answered in kind. We can be told, rightly, that there are sticks, stones, neutrinos, galaxies, numbers, classes. The answer is fallible, and subject to challenge at the checkpoints.

Naturalism; Or, Living within One's Means

for Henri Lauener's sixtieth birthday

1995

Abstract

Naturalism holds that there is no higher access to truth than empirically testable hypotheses. Still it does not repudiate untestable hypotheses. They fill out interstices of theory and lead to further hypotheses that are testable.

A hypothesis is tested by deducing, from it and a background of accepted theory, some observation categorical that does not follow from the background alone. This categorical, a generalized conditional compounded of two observation sentences, admits in turn of a primitive experimental test.

The observation sentences themselves, like ape cries and bird calls, are in holophrastic association with ranges of neural intake. Denotation of determinate objects figures neither in this association nor in deducing the categorical from the scientific hypotheses. Hence the indeterminacy of reference; ontology is purely auxiliary to the structure of theory. Truth, however, is seen still as transcendent at least in this sense: we say of a superseded scientific theory not that it ceased to be true, but that it is found to have been false.

Names of philosophical positions are a necessary evil. They are necessary because we need to refer to a stated position or doctrine from time to time, and it would be tiresome to keep restating it. They are evil in that they come to be conceived as designating schools of thought, objects of loyalty from within and objects of obloquy from without, and hence obstacles, within and without, to the pursuit of truth.

In identifying the philosophical position that I call naturalism, then, I shall just be describing my own position, without prejudice to possibly divergent uses of the term. In *Theories and Things* I wrote that naturalism is

“the recognition that it is within science itself, and not in some prior philosophy, that reality is to be identified and described”; again that it is “abandonment of the goal of a first philosophy prior to natural science” (pp. 21, 67). These characterizations convey the right mood, but they would fare poorly in a debate. How much qualifies as “science itself” and not “some prior philosophy”?

In science itself I certainly want to include the farthest flights of physics and cosmology, as well as experimental psychology, history, and the social sciences. Also mathematics, insofar at least as it is applied, for it is indispensable to natural science. What then am I excluding as “some prior philosophy,” and why? Descartes’ dualism between mind and body is called metaphysics, but it could as well be reckoned as science, however false. He even had a causal theory of the interaction of mind and body through the pineal gland. If I saw indirect explanatory benefit in positing sensibilia, possibilia, spirits, a Creator, I would joyfully accord them scientific status too, on a par with such avowedly scientific posits as quarks and black holes. What then *have* I banned under the name of prior philosophy?

Demarcation is not my purpose. My point in the characterizations of naturalism that I quoted is just that the *most* we can reasonably seek in support of an inventory and description of reality is testability of its observable consequences in the time-honored hypothetico-deductive way—whereof more anon. Naturalism need not cast aspersions on irresponsible metaphysics, however deserved, much less on soft sciences or on the speculative reaches of the hard ones, except insofar as a firmer basis is claimed for them than the experimental method itself.

Where naturalistic renunciation shows itself most clearly and significantly is in naturalistic epistemology. Various epistemologists, from Descartes to Carnap, had sought a foundation for natural science in mental entities, the flux of raw sense data. It was as if we might first fashion a self-sufficient and infallible lore of sense data, innocent of reference to physical things, and then build our theory of the external world somehow on that finished foundation. The naturalistic epistemologist dismisses this dream of prior sense-datum language, arguing that the positing of physical things is itself our indispensable tool for organizing and remembering what is otherwise, in James’ words, a “blooming, buzzing confusion.”

To account for knowledge of an external thing or event, accordingly, the naturalistic epistemologist looks rather to the external thing or event itself and the causal chain of stimulation from it to one’s brain. In a paradigm

case, light rays are reflected from the object to one's retina, activating a patch of nerve endings, each of which initiates a neural impulse to one or another center of the brain. Through intricate processes within the brain, finally, and abetted by imitation of other people or by instruction, a child comes in time to utter or assent to some rudimentary sentence at the end of such a causal chain. I call it an observation sentence. Examples are 'It's cold', 'It's raining', '(That's) milk', '(That's a) dog'.

Customarily the experimental psychologist chooses one or another object or event, from somewhere along such a causal chain, to represent the chain, and this he calls the stimulus. Usually it is an event of his own devising. In one experiment it will be a flash or a buzz in the subject's vicinity, and in another it will be an ice cube or a shock at the subject's surface. For our more general purposes, not linked to any particular experiment, an economical strategy in defining the stimulus is to intercept the causal chains just at the subject's surface. Nothing is lost, for it is only from that point inward that the chains contribute to the subject's knowledge of the external world.

Indeed, even what reaches the subject's surface is relevant only if it triggers neural receptors. So we might for our purposes simply identify the subject's stimulus, over a given brief moment, with the temporally ordered set of sensory receptors triggered in that moment.

Still further economy might be sought by intercepting the causal chains rather at a deeper level—somewhere within the brain; for even the surface receptors that are triggered on any given occasion are largely without relevant effect on the subject's behavior. However, our knowledge of these deeper levels is still too sketchy. Moreover, as research increasingly penetrates these depths, we become aware of complexity and heterogeneity radically at variance with the neat simplicity at the surface. Each receptor, after all, admits of just two clean-cut states: triggered or not.

Moreover, the behaviorally irrelevant triggerings in a global stimulus can be defined out anyway, in due course, by appeal to perceptual similarity of stimuli. The receptors whose firing is *salient* in a given stimulus are the ones that it shares with all perceptually similar stimuli. Perceptual similarity itself can be measured, for a given individual, by reinforcement and extinction of responses.

So it seems best for present purposes to construe the subject's stimulus on a given occasion simply as his global neural intake on that occasion. But I shall refer to it only as neural intake, not stimulus, for other notions of

stimulus are wanted in other studies, particularly where different subjects are to get the same stimulus. Neural intake is private, for subjects do not share receptors.

Perceptual similarity, then, is a relation between a subject's neural intakes. Though testable, it is a private affair; the intakes are *his*, and are perceptually more or less similar for *him*. Perceptual similarity is the basis of all learning, all habit formation, all expectation by induction from past experience; for we are innately disposed to expect similar events to have sequels that are similar to each other.

The association of observation sentences with neural intakes is many-many. Any one of a range of perceptually fairly similar intakes may prompt the subject's assent to any one of a range of semantically kindred sentences. But in contrast to the privacy of neural intakes, and the privacy of their perceptual similarity, observation sentences and their semantics are a public matter, since the child has to learn these from her elders. Her learning then depends indeed both on the public currency of the observation sentences and on a preestablished harmony of people's private scales of perceptual similarity. The harmony is formal, in this sense: if a witness finds the first of three scenes less similar to the second than to the third, another witness is apt to do likewise. This approximate harmony is preestablished in a shared gene pool. Different people's feelings still might not match, whatever that might mean.

This much is a naturalistic analogue or counterpart of the traditional epistemologist's phenomenalist foundation in sense data. However, it pretends to plausibility in psychology, in genetics, and even in prehistory. Observation sentences have their antecedents in birdcalls and in the signal cries of the apes.

Building on this naturalistic foundation, then, in parallel to the old epistemologist's proposed construction of science on a foundation of sense data, the naturalist would venture a psychologically and historically plausible sketch of the individual's acquisition of science and perhaps the evolution of science down the ages, with an eye primarily to the logic of evidence. I will spare you most of that, for I have gone into it in *Word and Object* and better in *The Roots of Reference, Pursuit of Truth*, and elsewhere. There are just a couple of aspects that I want to remind you of.

One is reification, or the positing of objects. Observation sentences commonly contain words that refer to objects when used in mature discourse, but the infant first acquires such a sentence only as a seamless

whole, conditioned—like the signal cry of the ape—to an appropriate range of global neural intakes. But there is a harbinger of reification already in our innate propensity, and that of other animals, to confer salience on those components of a neural intake that transmit corporeal patches of the visual field. It is what Donald Campbell calls our innate reification of bodies, but I construe reification rather in degrees. Special ways of compounding observation sentences mark further steps in the reification of bodies, and the job is complete only when the speaker has mastered past and future tense and knows about the unseen but continuous translation of an identical body through space between observations. It is only then that she makes sense of a body's being the same body from one observation to another despite intervening changes in appearance.

At that point the reification of bodies is full fledged. Reification of less conspicuous objects, notably abstract ones such as numbers and classes, takes further explaining, and admits of it. A crucial step there, as I see it, is mastery of relative clauses and pronouns.

This burgeoning language of science is a direct extension of the faltering language of observation. Segments of observation sentences carry over and become—some of them—terms for objects. Conversely, sentences learned only later by grammatical synthesis from a sophisticated vocabulary can come to qualify as observation sentences as well. For, what I take as definitive of observation sentences is just this pair of conditions: first, the speaker must be disposed to assent to the sentence or dissent from it outright on making the appropriate observation, irrespective of his interrupted line of thought if any, and second, the verdict must command the agreement of any witnesses from the appropriate language community. This second requirement, intersubjectivity, is needed in order that the child be able to learn observation sentences from his elders; and those sentences, some of them, are his indispensable entering wedge in acquiring cognitive language. Intersubjectivity of observation sentences is likewise essential at the other end, to assure objectivity of science.

The sharing of vocabulary by observation sentences and sentences of science was necessary not only for the emergence of scientific language; it is necessary also as a channel for the empirical testing of scientific hypotheses. The primordial hypotheses are what I call observation categoricals, compounded of pairs of observation sentences: thus 'When it snows, it's cold'. To check such a hypothesis experimentally, we contrive to put ourselves in a situation where the first component, 'It's snowing', is observ-

ably fulfilled, and then we check for fulfillment of the second component. If it is fulfilled, the categorical remains standing until further notice. If it is not fulfilled, the categorical is refuted once for all.

I see this as the key to the empirical testing also of more sophisticated hypotheses. We conjoin the hypothesis in question to a set of already previously accepted statements, sufficient together to imply some observation categorical that was not implied by the previous set alone. Then we check the observation categorical.

The appeal to logical implication here presents no problem. The basic laws of logic are internalized in learning the use of the logical particles. For instance, the child learns by observation and parental correction that it is misuse of the conjunction “and” to affirm an “and” compound and then deny one of the components. The child has thus internalized one simple logical implication, namely that an “and” compound implies its components, on pain of simply getting a word wrong. Correspondingly for other basic implications, up to and including the laws of quantifiers and identity. Insofar I am with Lauener in recognizing analyticity.

Scientists of course do not trace all these links of implication from hypothesis to observation categorical. It would mean filling in all the logically requisite supporting statements, most of which are so familiar to him or so trivial as to go without saying. In practice, moreover, many tacit premisses often express mere statistical trends or probabilities, which he will take in stride unless unexpected results prompt him to reconsider.

Still the deduction and checking of observation categoricals is the essence, surely, of the experimental method, the hypothetico-deductive method, the method, in Popper’s words, of conjecture and refutation. It brings out that prediction of observable events is the ultimate test of scientific theory.

I speak of test, not purpose. The purpose of science is to be sought rather in intellectual curiosity and technology. In our prehistoric beginnings, however, the purpose of the first glimmerings of scientific theory *was* presumably prediction, insofar as purpose can be despiritualized into natural selection and survival value. This takes us back to our innate sense or standard of perceptual similarity, and the innate expectation that similars will have mutually similar sequels. In short, primitive induction.

Prediction is verbalized expectation. Conditional expectation, when correct, has survival value. Natural selection has accordingly favored innate standards of perceptual similarity that have harmonized with trends in our

environment. Natural science, finally, is conditional expectation hypertrophied.

I said that prediction is not the main purpose of science, but only the test. It is a negative test at that, a test by refutation. As a further disavowal let me add, contrary to positivism, that a sentence does not even need to be testable in order to qualify as a respectable sentence of science. A sentence is testable, in my liberal or holistic sense, if adding it to previously accepted sentences clinches an observation categorical that was not implied by those previous sentences alone; but much good science is untestable even in this liberal sense. We believe many things because they fit in smoothly by analogy, or they symmetrize and simplify the overall design. Surely much history and social science is of this sort, and some hard science. Moreover, such acceptations are not idle fancy; their proliferation generates, every here and there, a hypothesis that can indeed be tested. Surely this is the major source of testable hypotheses and the growth of science.

The naturalization of epistemology, as I have been sketching it, is both a limitation and a liberation. The old quest for a foundation for natural science, firmer than science itself, is abandoned: that much is the limitation. The liberation is free access to the resources of natural science, without fear of circularity. The naturalistic epistemologist settles for what he can learn about the strategy, logic, and mechanics by which our elaborate theory of the physical world is in fact projected, or might be, or should be, from just that amorphous neural intake.

Is this sort of thing still philosophy? Naturalism brings a salutary blurring of such boundaries. Naturalistic philosophy is continuous with natural science. It undertakes to clarify, organize, and simplify the broadest and most basic concepts, and to analyze scientific method and evidence within the framework of science itself. The boundary between naturalistic philosophy and the rest of science is just a vague matter of degree.

Naturalism is naturally associated with physicalism, or materialism. I do not equate them, as witness my earlier remark on Cartesian dualism. I do embrace physicalism as a scientific position, but I could be dissuaded of it on future scientific grounds without being dissuaded of naturalism. Quantum mechanics today, indeed, in its neoclassical or Copenhagen interpretation, has a distinctly mentalistic ring.

My naturalism has evidently been boiling down to the claim that in our pursuit of truth about the world we cannot do better than our traditional scientific procedure, the hypothetico-deductive method. A rebuttal sug-

gests itself here: surely mathematics. The obvious defense against that rebuttal is to say that mathematical truths are not about the world. But this is not a defense of my choosing. In my view applied mathematics *is* about the world.

Thus consider again a case where we are testing a scientific hypothesis by conjoining it to some already accepted statements and deducing an observation categorical. Likely as not, some of those already accepted statements are purely mathematical. This is how pure mathematics gets applied. Whatever empirical content those already accepted statements can claim, then, from being needed in implying the observation categorical, is imbibed in particular by the mathematical ones.

Thus it is that I am inclined to blur the boundary between mathematics and natural science, no less than the boundary between philosophy and natural science. If it is protested that proved mathematical truths are not subject to subsequent refutation, my answer is that we safeguard them by choosing to revoke non-mathematical statements instead, in cases where a set of statements has been found conjointly to imply a false observation categorical. Reasons can be adduced for doing so; but enough.

That leaves open the vast proliferations of mathematics that there is no thought or prospect of applying. I see these domains as integral to our overall theory of reality only on sufferance: they are expressed in the same syntax and lexicon as applicable mathematics, and to exclude them as meaningless by *ad hoc* gerrymandering of our syntax would be thankless at best. So it is left to us to try to assess these sentences also as true or false, if we care to. Many are settled by the same laws that settle applicable mathematics. For the rest, I would settle them as far as practicable by considerations of economy, on a par with the decisions we make in natural science when trying to frame empirical hypotheses worthy of experimental testing.

Traditional epistemology was in part normative in intent. Naturalistic epistemology, in contrast, is viewed by Henri Lauener and others as purely descriptive. I disagree. Just as traditional epistemology on its speculative side gets naturalized into science, or next of kin, so on its normative side it gets naturalized into technology, the technology of scientizing.

What might be offered first of all as a norm of naturalized epistemology is *prediction of observation* as a test of a hypothesis. I think of this as more than a norm: as the name of the game. Science cannot all be tested, and the softer the science the sparser the tests; but when it is tested, the test is pre-

diction of observation. Moreover, naturalism has no special claims on the principle, which is rather the crux of empiricism.

What are more distinctively naturalistic and technological are norms based on scientific findings. Thus science has pretty well established—subject to future disestablishment, as always—that our information about distant events and other people reaches us only through impact of rays and particles on our sensory receptors. A normative corollary is that we should be wary of astrologers, palmists, and other soothsayers. Think twice about E.S.P.

For a richer array of norms, vague in various degrees, we may look to the heuristics of hypothesis: how to think up a hypothesis worth testing. This is where considerations of conservatism and simplicity come in, and, at a more technical level, probability theory and statistics. In practice those technical matters spill over also, as I remarked, to complicate the hypothetico-deductive method itself.

I said at the beginning of this paper that according to naturalism it is within science itself and not some prior philosophy that reality is to be identified. Farther along in a more narrowly scientific spirit, I speculated on how we round out our recognition of objects *as* objects, bit by bit, with our acquisition of language and science. These matters call now for some more broadly philosophical reflections.

Let us recall, to begin with, that the association of observation sentences with neural intake is holophrastic. What objects the component words may designate in other contexts is irrelevant to the association. This is obviously so if the observation sentence is to be acquired as a first step in language learning; but the association is equally direct and holophrastic in its operation even if the sentence was acquired through synthesis of its words, and gained its immediacy only through subsequent familiarization.

Moreover, the specifics of designation and denotation are not only indifferent to the association of observation sentences to neural intake; they are indifferent also to the implication of observation categoricals by scientific theory. It is logical implication; and logic, unlike set theory and the rest of mathematics, responds to no traits of objects beyond sameness and difference. So we must conclude that objects of any sort figure only as neutral nodes in the structure of scientific theory, so far as empirical evidence is concerned. We can arbitrarily change the values of our variables, the designata of our names, and the denotata of our predicates without dis-

turbing the evidence, so long anyway as the new objects are explicitly correlated one-to-one with the old. Such is the indeterminacy of reference, as I have come to call it.

At first it is perhaps alarming. We are left with no basis, it would seem, for judging whether we are talking about familiar things or some arbitrary proxies. The shock subsides, however, when we reflect on a homely example or two. Thus think of a body in the scientific framework of space and time. Insofar as you specify the precise sinuous filament of four-dimensional space-time that the body takes up in the course of its career, you have fixed the object uniquely. We could go farther and *identify* the object, a chipmunk perhaps, with its portion of space-time, thus saying that it is tiny at its early end and bigger at its late end. The move is artificial, but actually it confers a bit of economy, if we are going to have the space-time anyway. Subjective connotations of brownness, softness, swift and erratic movement, and the rest simply carry over. Surely all matters of evidence remain undisturbed. We are even prepared to say that it was what a body was all along, an appropriately filled-in portion of space-time as over against empty ones.

Next we might identify space-time regions in turn with the sets of quadruples of numbers that determine them in some arbitrarily adopted frame of coordinates. We can transfer sensory connotations now to this abstract mathematical object, and still there is no violence to scientific evidence. To speak intuitively, nothing really happened.

Thus we can come to terms somewhat with the indeterminacy of reference, as applied to bodies and other sensible substances, by just letting the sensory connotations of the observation sentences carry over from the old objects to their proxies.

In the case of abstract objects such as numbers, devoid of sensory connotations, the indeterminacy of reference is already familiar. It is seen in Frege's so-called Caesar problem: the number five may be Julius Caesar. We happily use numbers without caring whether they be taken according to the Frege-Russell constructions or Ackermann's or von Neumann's. The point was dramatized long ago by F. P. Ramsey with his expedient of Ramsey sentences, as they have come to be called. Instead of invoking the abstract objects specifically, when certain of their properties are needed in an argument, the Ramsey sentence just says that there *are* objects with the properties, and then invokes the objects by variables without further identification. This expedient only works for abstract objects, however, used as

auxiliaries here and there without regard to whether they remain the same objects from one context to another.

The indeterminacy of reference can be seen again in its full generality, as Davidson once remarked, by an examination of Tarski's classical truth definition. If a sentence comes out true under that definition, it continues to do so when objects are reassigned to its predicates in any one-to-one way.

These reflections on ontology are a salutary reminder that the ultimate data of science are limited to our neural intake, and that the very notion of object, concrete or abstract, is of our own making, along with the rest of natural science and mathematics. It is our overwhelmingly ingenious apparatus for systematizing, predicting, and partially controlling our intake, and we may take pride.

This conventionalist view of ontology appeals, I expect, to Henri Lauener. He in his pragmatism even settles for a plurality of scientific specialties, each with its working ontology, and no dream of an overarching, unifying fact of the matter.

Naturalism itself is noncommittal on this question of unity of science. Naturalism just sees it as a question within science itself, albeit a question more remote from observational checkpoints than the most speculative questions of the hard and soft sciences ordinarily so called.

Naturalism can still respect the drive, on the part of some of us, for a unified, all-purpose ontology. The drive is typical of the scientific temper, and of a piece with the drive for simplicity that shapes scientific hypotheses generally. Physicalism is its familiar manifestation, and physicalism is bound to have had important side effects in the framing of more special hypotheses in various branches of science; for physicalism puts a premium on hypotheses favorable to closer integration with physics itself. We have here a conspicuous case of what I touched on earlier: scientific hypotheses which, though not themselves testable, help to elicit others that are.

In any event, we are now seeing ontology as more utterly a human option than we used to. We are drawn to Lauener's pragmatism. Must we then conclude that true reality is beyond our ken? No, that would be to forsake naturalism. Rather, the notion of reality is itself part of the apparatus; and sticks, stones, atoms, quarks, numbers, and classes all are utterly real denizens of an ultimate real world, except insofar as our present science may prove false on further testing.

What then is naturalism's line on truth and falsity themselves? The truth

predicate raises no problem in its normal daily use as an instrument of what I have called semantic ascent. Tarski's disquotational account accommodates it, so long as what are called true are sentences in our own language; and we then extend the predicate to sentences of other languages that we accept as translations of truths of our own. However, paradoxes arise when the truth predicate is applied to sentences that contain that very predicate or related ones; so we are called upon to recognize rather a hierarchy of truth predicates, each of which behaves properly only in application to sentences that do not contain that predicate itself or higher ones. It is a hierarchy of better and better truth predicates but no best. In practice, except in contexts such as these philosophical ones, occasions seldom arise for venturing above the first rung of the ladder. Truth *off* the hierarchy, absolute truth, would indeed be transcendent; bringing it down into scientific theory of the world engenders paradox. So naturalism has no place for that.

Still, our concept of truth strains at its naturalistic moorings in another way. We naturalists say that science is the highest path to truth, but still we do not say that everything on which scientists agree is true. Nor do we say that something that was true became false when scientists changed their minds. What we say is that they and we *thought* it was true, but it wasn't. We have scientists pursuing truth, not decreeing it. Truth thus stands forth as an ideal of pure reason, in Kant's apt phrase, and transcendent indeed. On this score I am again with Lauener.

C. S. Peirce tried to naturalize truth by identifying it with the limit that scientific progress approaches. This depends on optimistic assumptions, but if we reconstrue it as mere metaphor it does epitomize the scientists' persistent give and take of conjecture and refutation. Truth as goal remains the established usage of the term, and I acquiesce in it as just a vivid metaphor for our continued adjustment of our world picture to our neural intake. Metaphor is perhaps a handy category in which to accommodate transcendental concepts, from a naturalist point of view.

Progress on Two Fronts¹

1996

In these pages I update my views on two epistemological matters.

I. Preestablished Harmony

Correct translation, we used to say, preserves meaning. In *Word and Object*,² chary of mentalism and meaning, I represented correct translation of an observation sentence as preserving what I called its *stimulus meaning*. I explained its stimulus meaning as the range of stimulations, any one of which would prompt the subject to assent to the sentence. Each stimulation, in my sense of the term, was the near-simultaneous firing of some subset of the subject's neuroreceptors.

Not that I pictured the translator as neurologizing. He would just guess his translation of the native's observation sentence in the obvious way, picturing himself in the native's stance and considering what English sentence it would prompt of him. Eventually, he would build his translations of theoretic sentences upon those of observation sentences by what I called *analytical hypotheses*; still no talk of nerve endings. Talk of stimulus meaning and of analytical hypotheses was rather *my* business; my theory of the translator's activity. Stimulus meaning was what, theoretically speaking, correct translation of an observation sentence preserved.

This is uncomfortable theory, however. It calls for sameness of stimulus meaning of the native sentence for the native and the English sentence for

1. Thanks to discussions with Burton Dreben, this paper has been much improved through several drafts.

2. Cambridge, MA: MIT Press, 1960.

the translator, and hence a sharing of stimulations by native and translator. Well, they cannot share neuroreceptors, so we must settle rather for homology of receptors. Such homology is by no means to be expected, and anyway surely should not matter, as I remarked in a lecture five years later.³ I expressed the same discomfort again in 1974,⁴ citing Charles Darwin⁵ on the diversity of nerve nets from individual to individual. But I saw no steps to take.

The matter surfaced again at a conference with Donald Davidson, Burton Dreben, and Dagfinn Føllesdal at Stanford in 1986. Davidson favored treating translation purely in terms of the external objects of reference, bypassing consideration of neural receptors. Such is the familiar practice also of experimental psychologists, for whom the so-called stimulus is out there, shared by subject and experimenter. I recognized the adequacy of this object-oriented line in describing the procedures of translation, having described them thus myself, and its adequacy likewise for lexicography. There remains a problem for epistemology and neuropsychology, however, when we reflect with Darwin on the intersubjective diversity of nerve nets and receptors. How does the mere sameness of the distal cause, the jointly observed object, prevail over the diversity of the proximal segments of the causal chains, inside the two observers, and still issue in agreeing response? In short, *why* can the translator and the lexicographer blithely rest with the distal stimulus, as indeed they can?

Let me pinpoint the problem. A rabbit appears, the native says 'Gavagai', and the translator conjectures 'Rabbit'. On a later occasion they espy another rabbit, the translator says 'Gavagai', and the native concurs. The two occasions were *perceptually similar*⁶ for the native, by his subjective standards of perceptual similarity, and likewise for the translator by his independently testable subjective standards of perceptual similarity. Anatomic likeness of the native's receptors and those of the translator could have helped to account for this agreement, but that is out. What then does?

If I had thus pinpointed my uneasiness back then, I would have quickly resolved it. What we have is a preestablished *harmony* of standards of perceptual similarity, independent of intersubjective likeness of receptors or

3. "Propositional Objects," 1965, published in *Ontological Relativity and Other Essays* (New York: Columbia, 1969); see page 157.

4. *Roots of Reference* (La Salle, IL: Open Court, 1974), 23–24, 41.

5. *Origin of Species* (London: John Murray, 1859), 45–46.

6. *Roots of Reference*, 16–18; *From Stimulus to Science* (Cambridge, MA: Harvard University Press, 1995), 17–18.

sensations. Shades of G. W. Leibniz, thus, but without appeal to divine intervention. The harmony is explained by a yet deeper, but more faltering preestablished harmony between perceptual similarity and the environment. This, in turn, is accounted for by natural selection, as follows.

We have, to begin with, an *inductive* instinct: we tend to expect perceptually similar stimulations to have sequels that are similar to each other. This is the basis of expectation, habit formation, and learning. Successful expectation has always had survival value, notably in the elusion of predators and the capture of prey. Natural selection has accordingly favored innate standards of perceptual similarity which have tended to harmonize with trends in the environment. Hence the success, so much better than random, of our inductions and expectations. Derivatively, then, through our sharing of an ancestral gene pool, our innate standards of perceptual similarity harmonize also intersubjectively.

Natural selection is Darwin's solvent of metaphysics. It dissolved Aristotle's final cause, teleology, into efficient cause, and now Leibniz's preestablished harmony as well.

Harmony without interaction: that was the subtlety. We take its ubiquitous effects for granted, not thinking them through. It was only in *From Stimulus to Science* (pp. 20–21) that I got the matter into proper focus and explained it as here.

Note that the harmony was needed not only in respect of the rabbits, in my example, but also in respect of the utterance of 'Gavagai'. The sound of it was perceptually similar for the native and likewise for the translator when it was spoken by either of them. This applies equally, of course, to the child's learning of words from the mother.

The effect of the intersubjective harmony, we see, is that what the two observers agree on is the shared distal subject matter and not the unshared proximal stimulations. The latter, however, are what are related by perceptual similarity, though not intersubjectively. Might it be more straightforward to readjust the terminology and treat the external subject matter as the relata of perceptual similarity?

One reason not to is that it would put the whole vital causal structure of preestablished harmony back where it had been, out of sight and out of mind. Another reason is individuation. What is distally perceived may be an ill-defined event rather than a body, or, if a body, a body in an ill-defined range of perspectives, accommodating our two observers on the occasion in question and perhaps excluding others. I cut through all this by adhering to the subject's neural intakes as the relata of that subject's re-

lation of perceptual similarity, which then enjoys harmony, generous but probably not complete, with that of other subjects.

Ironically, I had described this harmony of perceptual similarity standards already in *Roots of Reference*, on the very page (23) where I then proceeded to bemoan my dependence on intersubjective matching of neural intakes. I even accounted for the harmony by natural selection, shared environment, and heredity, just as in the present pages and in *From Stimulus to Science*. My account differed only in one crucial detail. Today, I have the two subjects exposed to the same external event, whereas in *Roots of Reference* I was still grudgingly according them homologous neural intakes. It made a difference of twenty years. I left myself and others hanging with this enigmatic envoi: "I shall not pause over the lesson, but there is surely one there" (*op. cit.*, p. 23).

II. Theory-Laden Observation

In his early years, the only way a child can go wrong on an observation sentence is by affirming it in a situation that would not have commanded the adult's outright assent. Second thoughts are not yet relevant; they become so only at a later stage, when scientific theory has begun to interrelate observation sentences and generate conflict. The adult may affirm 'There's a rabbit' outright, and then rescind it when a later observation suggests that it was the neighbor's cat.

In *Word and Object* (pp. 42–44) I consequently recognized degrees of observationality, and treated the intrusion of theory or collateral information as dilution of observationality. But in later writings, I held observationality as absolute, based on immediacy of assent, and then I accommodated the intrusion of theory by contrasting the *holophrastic* conditioning of the observation sentence to neural intake with the *analytic* relations of the component words to the rest of language. The sentence figures holophrastically both in the infant's first acquisition of it and in the scientist's immediate assent to it when testing a theory.

My attention was turned anew to these matters by a recent letter from Lars Bergström adumbrating an empirical theory of truth. I ended up not adopting it, but it has prompted me to reflect further on the bipolarity of the holophrastic and analytic perspectives, as against the gradualism of observationality in *Word and Object*, or, as we might say, a gradualism of theoreticity.

In conclusion, I retained the absolute notion of an observation sentence as simply an occasion sentence that commands the subject's immediate assent, however fallible and revisable. Fallibility is then accommodated in a separate dimension, *theoreticity*, which invests observation sentences in varying degrees.

As before, all observation sentences are holophrastic in their association to stimulation. Thus it is that assent to them is immediate, unless the stimulation is off in the fringe of appropriate stimuli. Also, as before, it is in the analytic or word-by-word relations of the sentence to others that theory comes in, in varying degrees. Theoreticity is at a minimum in the observation sentence 'That's blue', and appreciable in 'That's a rabbit'. It may even be said to invest innate perceptual similarity, for theory is implicit in our innate readiness to integrate varied perspectives into an objective reality. A square on the floor, seen from various angles, projects on the retina now a square and now one and another trapezoid, but all these neural intakes are perceptually similar; we are said to "see the square."

I am thus won over to Thomas Kuhn and others who have insisted that observation is inseparable from theory. But my observation sentences stay on in their old definition and their old role as clauses in the observation categoricals that are the checkpoints of science.

This fits nicely with scientific practice. The experimenter settles for criteria that he can confidently adjudicate on the spot: confidently, not infallibly. He looks carefully, and if he is subsequently in doubt he repeats the experiment if he can. The finality of an experiment is historical, not logical.

The evidential burden borne by a highly theoretic observation sentence such as 'There was copper in it', when it is invoked at a checkpoint, can certainly be shifted to less theoretic observation sentences by deeper analysis. The ideal of reducing all checkpoints to minimally theoretic observation sentences like 'That's blue', however, is the old phenomenalistic reductionism of which I have long since despaired.

Observation categoricals, formed of observation sentences two by two, remain the checkpoints of scientific theory as before, and the experimental refutation of an implied categorical refutes the theory that implied it. But the refutation is not final, for the refutation of the categorical was not final; the experimenter may have misjudged the protasis of the categorical as fulfilled, or misjudged the apodasis as failed, such being the fallibility of observation. We just expect him to do the best he can.

The Flowering of Thought in Language

1997

Our first mental endowment was instinct. Then came thought, and later language. Thanks to language, thought then proceeded to flourish. Such was our phylogeny. Ontogeny, then, true to form, recapitulates the sequence in the development of each child. By instinct a child a few days old will show anxiety when an object moving steadily to a screen and behind it fails to emerge on schedule at the far edge. Also new-borns are said to respond emotionally to emotional facial expressions and even to imitate them.

A further innate endowment, and an elaborate one, is manifest in one's standards of *perceptual similarity*. This relation is basic to all learning, so some of it has to be innate. We are born with standards of perceptual similarity, which then develop and change in the course of experience.

I treat perceptual similarity as a relation between stimulations, or sensory intakes. The similarity is subjective, each individual having his own standards, but it is objectively testable for each individual by reinforcement and extinction of responses. It does not match up with *receptual* similarity, that is, mere identity or proximity of the nerve endings that were triggered on the two occasions. Perceptual similarity is illustrated rather by our readiness to equate the various perspectives of an object, prior even to probing it. It is a striking case of innate knowledge, however inarticulate.

We have a further instinct that parlays perceptual similarity into an ever widening fund of knowledge: the instinct of *induction*. This instinct is an innate disposition to expect perceptually similar stimulations to have sequels that are perceptually similar to one another. Flashes of lightning are similar to one another, and their sequels, thunderbolts, are likewise similar

to one another. This inductive instinct is the basis of expectation and hence of habit formation and learning.

So induction engenders expectation and hence belief, but how knowledge? how truth? Here Darwin lends a hand: the creative hand of natural selection. Successful expectation has had survival value down the ages in helping our forebears avoid their predators and capture their prey. Those who lived long enough to be forebears did so through predominantly correct expectation. Their genes carried standards of perceptual similarity that meshed pretty well with environmental trends. Thus it is that induction has tended to engender not just expectation but successful expectation, hence knowledge, or anyway true belief.

Such is our innate head-start, or part of it, in the business of thinking. It is a head-start that we share with other higher animals, whose standards of perceptual similarity can be elicited experimentally in the same way as ours. Up to this point, we and they were perhaps about equally endowed. Then, with the development of language, human thought proceeded to soar. In respect of cause and effect, language and thought have been intimately intertwined: language as vehicle of thought, thought as innovator of language. I despair of disentangling them. But I shall delineate a succession of advances in language and thought that may have brought us to the level of language and thought that we enjoy today.

Birds and apes do have audible signals that are precursors of speech.

Chimpanzees are said to have distinctive cries by which to warn one another of predators: one cry for lions and leopards, another for eagles. A third cry may report fruit at some treetop. The cries serve the basic function of language, namely a social sharing of information. The individual gets the benefit of his neighbour's perceptions.

The human infant's first acquisition in the way of cognitive language is comparable to the apes' cries: 'Mama' perhaps, or 'Milk' or 'Ball'. Whether the apes' cries are innate or learned is as may be, but the child's words are learned; English has not made its way into the genes.

The child learns the word by ostension: the ball is in focus when he is told the word 'Ball', and he makes the association. An imitative instinct is probably at work: imitation of the mother's word; for such an instinct is proverbial in man and ape. At recurrence of the ball, the child is prepared to volunteer the word or assent to it. He has learned to associate the word with a range of mutually fairly similar neural intakes, similar in featuring the ball. The similarity is perceptual similarity by the child's standards.

There is a richness here that escapes the casual eye. The recurrences that continue to elicit the word, or to elicit assent to it, from both the child and the mother, are similar for the mother and for the child, but similar according to their separate, private scales of perceptual similarity; no intersubjective perceptual similarity is posited or even defined. Yet the two scales keep pace. Failing such harmony, the second ball-sighting might have resembled the first for the child but not for the mother, or vice versa. This does happen, but not usually. The harmony is this: two shared occasions that are perceptually similar for one witness tend to be so for another.

Without this harmony there could be no appreciable communication. The harmony is required not only in the sightings of the ball, but also in the sound of the word. It must sound similar to the child and similar to the mother when spoken by the child and by the mother, and on one occasion and another, though no similarity is assumed between how it sounds to the child and to the mother.

The cause of the harmony becomes evident when we recall the cause of the perceptual similarity standards themselves; namely, natural selection, moulding them to suit the trends of the environment. Shared ancestry in shared environment tuned us alike, nearly enough, to our continuing intake.

We have the child learning 'Mama' or 'Milk' or 'Ball'. This much is on a par with the apes' cries. Each such word is associated with some range of perceptually more or less similar neural intakes. The child goes on learning further ones hand over fist, promptly outstripping the apes' meagre repertoire. Presently the child masters the use of prepositions, conjunctions, and the copula to combine simple expressions, and thus moves into mass production.

My examples have been nouns: 'Mama', 'Milk', 'Ball'. But the sentence 'It's cold' or 'It's raining' would do as well, and the three nouns can be thought of as sentences too: 'That's Mama', 'That's milk', 'That's a ball.' But they all are *occasion* sentences: true on some occasions, false on others. Sometimes it is raining, sometimes not. Besides this general trait, their distinctive trait is that each has become associated, for the speaker in question, with some range of perceptually more or less similar neural intakes. Any one of them would prompt his unhesitating assent. I call such an expression an *observation sentence* for the speaker in question.

We can imagine a stage in the prehistory of man when his cognitive lan-

guage was limited to such observation sentences. It was language only by courtesy, for it lacked the recursive power of true grammar that generates essentially new sentences without limit. The conjunctions, prepositions, and copula are productive of new observation sentences up to a point, but the added requirement of unhesitating assent in my definition of 'observation sentence' limits their length.

Thus far the advantage over the apes' meagre endowment is purely quantitative, but enormous. My fancied *Urmensch* can not only broadcast a fine-tuned roster of species that he sees approaching or ripening, he can also talk to himself about sealing up his shelter. 'A bit of clay here', he ponders. 'Water on clay. Flat stone in here'. Like the modern engineer, he also invokes arm and finger muscles in planning the moves, and then his vocabulary enables him to tell his son to help him find what he needs. Gestures fill in for a few verbs.

Even a bit of logic can seep in. In learning the conjunction 'and', he learns simply as a matter of usage not to affirm a sentence of the form '*p* and *q*' without being prepared to affirm '*q*'; and similar points of usage regarding 'or', 'not', and 'if'. Violations of logic would be like calling a lizard a snake: a misuse of words.

At some later point in the dim wastes of antediluvian time man achieved a more resounding linguistic breakthrough; and at some point our child recapitulates it, picking it up from us. The new device puts the yield of our instinct of induction into words. It expresses generalised expectation by combining observation sentences in pairs: thus 'When it's snowing, it's cold'; 'When lightning, thunder'; 'When the sun rises, birds sing.' I call the compounds *observation categoricals*.

The survival value of our old observation sentences and the apes' cries, as I remarked, is vicarious observation. The value of the observation categoricals, which on those observation sentences, is vicarious induction. They convey to us the inductive generalisations already achieved by our elders, and they can be handed down the generations.

This linguistic innovation, putting the output of our instinct of induction into words, marks the advent of *standing* sentences: sentences which, unlike our old observation sentences and other occasion sentences, are true or false once for all rather than on occasion. And not only that; they are generalisations stating, when true, scientific laws.

The observation categorical is a universal conditional of a sort, but its two clauses depend on no sharing of subject matter. To us in our sophisti-

cation, generality connotes quantification, and the 'When' in my examples suggests quantification over times; but all this is anachronistic in the present context. The intent of the primitive observation categorical is just to counsel the hearer to expect fulfilment of the antecedent clause to be accompanied or promptly followed by fulfilment of the consequent.

But what I picture as the next great prehistoric development is indeed the genuine universally quantified conditional, the classical categorical 'All ravens are black' or, phrased as an observation categorical, 'When anything is a raven, it is black.' In symbolic logic it is the universally quantified conditional ' $\forall x(\text{if } x \text{ is a raven then } x \text{ is black})$ '. I see this as the dawn of reification. It reifies whatever ravens there may be, and it is false if there are any white ones.

The humble pronoun 'it' in the second of the above three renderings is an *essential* pronoun: it cannot be replaced by its grammatical antecedent 'anything'. The essential pronoun, or the 'x' that is its mathematical version, is vital for thought and language because of the logical link that its cross-reference creates between its clause and that of its antecedent. The pure conditional 'If p then q ' is only a weak link, contributing little to the structure of scientific theory. The universal conditional ' $\forall x(\text{if } Fx \text{ then } Gx)$ ' is what structures our theory with its texture of causal laws.

I have pinned the idea of an object down to the essential pronoun or bound variable. Other idioms, intertranslatable with these, accomplish the same. An equivalent criterion applicable to other idioms is fulfilment of a predicate. I have favoured the former rendering because it rests on a more explicitly regimented, less mistakable notation.

The medieval nominalists took a wrong turning when they sought to de-reify universals by dismissing their names as *flatus vocis* or vocal breeze. You could likewise de-reify particulars if it were just a matter of dismissing their names as vocal breeze. Caesar could go the way of Pegasus. Where reification sets in, whether of universals or particulars, is where we treat of them by predicate or general term rather than individually by name. Our quantification over ravens reified them without their even bearing any names, and our quantification over numbers and other classes reifies the unspecifiable real numbers and other unspecifiable classes no less than the specifiable ones.

The full benefits of reification depended still on another vital achievement by early man, whether earlier or later than reification itself. It was the transcendence of the specious present: our ability to speak of the more re-

mote past and future. In our familiar language this is managed by the tenses of verbs. In a language with the structure of the predicate calculus it is managed by treating physical objects as spatiotemporally four-dimensional and then explicating the predicates 'earlier than' and 'part of' or predicates to that effect. This linguistic access to remote times extends the horizons of thought from the specious present to plus and minus infinity.

It is only after that much development of the child and the race that it becomes meaningful to speak of identity of a recurrent object after an appreciable absence. This is an achievement that we tend not to appreciate. The dog eagerly welcomes his master after a long absence, but this is just a qualitative identification like the child's continuing recognition of milk, ball, rain, thunder, and the mother. In the dog's case it is largely based on odour. We humans, and surely only we, learn to make sense of the question whether a red ball is the same red ball next week or another like it. Our decision or conjecture will depend on mastery of the sophisticated scheme of time, space, and the unseen trajectories of objects between appearances. This imposing development reflects a substantial theory of nature and a subtle language to embody it.

It is only after all this that man can be said to have fully reified physical objects; for it is only then that identity and distinctness of physical objects over time makes sense. This development was vitally important in tightening the structure of the four-dimensional world as we conceive it; for the continuing identity of a changing object is what integrates the time dimension, relating one time to another. Physical objects are time-binding, in Korzybski's happy phrase.

It was with the advent of the universal conditional that science could begin to develop its texture of causal laws, but only with the transcendence of the specious present that these laws could proliferate. From then on the continuity of bodies over time and the causal connections between the times have been explained or conjectured hand in hand.

The reification of abstract objects—numbers, notably and other classes—ensued on the analogy of physical objects, and took the same linguistic form: they are values of essential pronouns, or bound variables. This move has been indispensable in mathematics and hence in the hard sciences generally. The special service of time-binding, however, is not one to which the abstract objects contribute. Being changeless, they are above all that.

With science a going concern, the old observation categorical gets pressed into vital service as empirical check-point for scientific theory. The

primordial adverb 'When' with which I apologetically formulated it can now be taken frankly as quantifying over times, making the observation categorical a categorical in the classical sense of a universal conditional. But its two clauses are still, in this new use, observation sentences. However sophisticated in content, as I define them they must still command the qualified observer's unhesitating assent when the stimulation is clearly within the associated range. The assent is fallible and corrigible, but unhesitating. The checking of a theory then proceeds by logically deducing an observation categorical from the theory, and testing the categorical. The experimenter arranges fulfilment of its antecedent clause to his satisfaction and awaits fulfilment of the consequent.

I, You, and It: An Epistemological Triangle

1999

The triangle I have in mind is one that Donald Davidson has occasionally invoked. It has you at one vertex, some one of you. I am at another vertex, and at the third vertex there is some object. The object is reflecting light at both of us, but in somewhat different patterns owing to our difference in perspective. The light rays trigger our nerve endings, setting off one train of events in your nervous system and another in mine.

The object at the third vertex is a creature, unfamiliar to me. You tell me a name for it: *aardvark*. What went on in your nervous system and mine when we observed the *aardvark* differed in perspective and probably in more. We are differently wired. Our sensations may have differed too, if that makes sense. All we clearly shared was the distal cause of our neural events: the *aardvark*. Still I end up associating this same word with my stimulus, my neural intake, as you did with your different intake—numerically different certainly, and somewhat different in further ways. We thus differ in the proximal causes of our concordant use of the word, but we share the distal cause, the reference, farther out on our causal chains.

This is something worth wondering about. The causal chains from the *aardvark* to you and to me part company already at the *aardvark* itself and make their divergent ways into our unlike nerve nets, but we both end up calling the creature an *aardvark*. We shall find that what forges our link in the triangle, linking us, is a preestablished harmony.

I have now learned the word ‘*aardvark*’, thanks to you, and I proceed to apply it to similar creatures as they turn up. You do likewise, independently, and we find ourselves in agreement in our continuing use of the term. This suggests that we are both good judges of similarity, in some objective sense of the term. But wait: I know no objective sense of the term, apart from geometry.

Subjective similarity, yes. We can make good psychological sense, experimentally testable, of perceptual similarity of neural intakes for an individual. We can test it by reinforcing and extinguishing his responses. Stimulus *A* is perceptually more similar to stimulus *B* than to *C* if *A* evokes a response that was rewarded on the occasion of *B* but penalized on the occasion of *C*.

Back now to the aardvarks. The next time one turns up and you and I are there to observe it, I say 'There's an aardvark' and you agree. My neural intakes from the two aardvarks were perceptually similar by my lights, and your two were perceptually similar by yours. There is no presumption of perceptual similarity between your intakes and mine; intersubjective perceptual similarity is not even defined. What we have rather is a parallelism, a preestablished harmony, between your standards of perceptual similarity and mine. In general, if external events out there at the third vertex on two occasions produce neural intakes in both of us, and yours are perceptually similar for you, mine are apt to be perceptually similar for me.

Without this harmony, my perceptual similars might not have paired off so nicely with yours. My later intake might not have qualified as perceptually similar for me to my earlier one, though your later one did so qualify for you. Our scales of similarity might have diverged. Thanks to the harmony, they did not. Aardvarks look pretty much alike to me, and they look pretty much alike to you.

The preestablished harmony is needed to account for our meeting of minds not only on aardvarks, but also on what to call them: on the mellifluous Dutch disyllable 'aardvark' itself. The phonetic constancy of a word, from one utterance of it to another, is itself a product of the speakers' subjective standards of perceptual similarity. Thanks to the harmony, communication proceeds apace. Oh, we sound alike. Oh, who says so? Each of us, by his own standards of perceptual similarity, all of which are in harmony.

Perceptual similarity is vital not only for language, but for all learning, all habit formation, all expectation. It was its role in learning, or habit formation, that underlays the experimental test of perceptual similarity itself. Since it is presupposed in learning, moreover, perceptual similarity cannot itself have been learned, not all of it; some had to be innate, though it gets overlaid and changed as learning progresses.

So perceptual similarity standards are in part innate and are in preestablished harmony. Natural selection accounts nicely for both traits. The rele-

vance of similarity standards to survival becomes evident when we consider the role they play in expectation. We tend innately to expect similar events to have sequels that are in turn similar to each other. This is primitive induction. There is survival value in successful induction, successful expectation: it expedites our elusion of predators and our pursuit of prey. Natural selection, then, has favored similarity standards that mesh relatively well with the succession of natural events. This explains why our expectations have run a better than random score of success. It also explains the preestablished harmony: the standards are largely fixed in the genes of the race, the species.

What I call an observation sentence, for a given speaker, is any expression, such as 'aardvark', to which he has learned to assent in response to all neural intakes within some range of perceptually similar intakes. The expression may be a single word, as here, or it may be a grammatical sentence: 'It's cold', 'It's raining'. They are "about" the external world, not about sense data or stimulation. They are directed at the "it" vertex of the triangle, where reference lies.

An observation sentence is an *occasion* sentence, true on some occasions and false on others. An occasion sentence qualifies as an observation sentence for a speaker if he is disposed to assent to it outright when appropriately stimulated. It is an observation sentence for a society of speakers if it is an observation sentence for each member and if, further, all members who witness an occasion are disposed to agree in assenting or in not assenting.

The child's first cognitive utterances are observation sentences—thus 'Mama', 'Milk', 'Dog'—though it will be a while before the child gets a grip on the schematism of space, time, and recurrent enduring objects, to which these observation sentences as we know them are committed. These adult refinements make sense of sameness of a recurrent object and distinctness of an exactly similar object. For an infant the observation sentence 'Ball' would merely signal a recurrent state of affairs, at first, like 'Thunder' or 'It's cold'. Only later does it make sense to wonder whether it is the same ball next time around or another one like it. Only then is reification in full flower. But the reference is out in the world early and late, at the "it" vertex of the triangle. The child merely has to learn more about it. Meanwhile the range of neural intakes prompting the child's assent to the observation sentence 'Ball' can remain unchanged.

At the primitive or infantile stage the observation sentences are the hu-

man analogues of the apes' cries, the signals by which an ape alerts his kin to the approach of a predator or the prospect of some fruit at a treetop. Language is, itself perhaps a luxuriant outgrowth of the apes' cries. In any event the child promptly outstrips the apes, amassing new observation sentences hand over fist and learning ways of combining them with help of prepositions and conjunctions to form further observation sentences.

He then somehow caps this achievement with what I picture as his next great acquisition: the *observation categorical*. It is a generalized expression of expectation, the fruit of induction. It is a way of joining two observation sentences to express the general expectation that whenever the one observation sentence holds, the other will be fulfilled as well. Examples: 'When it snows, it's cold'; 'Where there's smoke, there's fire'; 'When the sun rises, the birds sing'; 'When lightning, thunder'. They are our first faltering scientific laws.

The survival value of the apes' cries, and of our ordinary observation sentences, lay in vicarious observation. One learns from one's kinsman, or kinsape, about something that one couldn't see from where one was sitting at the time. Now the observation categoricals bring us much more: they bring us vicarious habituation, vicarious induction. We get the benefit of generalized expectations built up over the years by some veteran observer or even by his informant, long dead. Observation categoricals can be handed down.

There is more to be said to the glory of the observation categoricals: they are the empirical checkpoints of scientific theory. If a theory implies an observation categorical, we can subject the theory to an experimental test by bringing about fulfillment of the observation sentence that is the protasis of the categorical and then watching for fulfillment of the apodasis. I see in this the distilled essence of the experimental method, and the only empirical check to which science is susceptible. Failure of the test brands a theory as false. Any appeal must rest on challenging the observations themselves, whereof more anon.

Language and science go on developing hand in hand, in the child and in the race. I already remarked on the big ontological leap achieved in conceiving of an enduring body, a body intermittently observed between unseen trajectories. Another big ontological leap was the reification of numbers and other abstract objects; it became clearly recognizable only with the advent of relative pronouns, or variables.

Observation sentences were at first acquired at mother's knee, and affirmed on appropriate occasions without peradventure. The growth of theory, however, calls upon us to think twice. 'That's a rabbit', affirmed under appropriate stimulation, had seemed final. But at a later stage, when we allow for the identity over time of an intermittently observed body, we may come to suspect that yesterday's supposed rabbit was just this toy that we now find in the garden.

We continue to accrete observation sentences throughout life. Some, like the child's early ones, are learned outright as wholes by direct conditioning to appropriate ranges of neural intake. Mostly such acquisitions will subsequently become infected with theory in varying degrees. Other observation sentences may be assembled from words learned in earlier contexts. Speakers vary in how they arrive at the same usage of an observation sentence.

There are sentences that are learned only through theory, but become observational later in the specialist's career; thus 'There was copper in it', said by the chemist after a glance at the solution, or 'There goes a hyperthyroid', said by a physician after a glimpse of a stranger's face.

What I have said thus far will have been pretty familiar to those of you who have been reading my writings. But there now comes a change, the effect of correspondence with Professor Bergström. He propounded what he called an empiricist theory of truth, which I found both appealing and disconcerting. In the end I did not adopt the theory, but I came to this decision only by sharpening my own views in a quarter to which I had hitherto given too little attention; the *recanting* of observation sentences.

The speaker assents outright to 'Rabbit' and retracts his assent later on discovering that what he had seen was a toy. In *Word and Object* (p. 44) I treated such intrusions of error as showing that the retracted sentence was not purely observational; theory had intruded. Theory had indeed, but I now see a better way of sorting matters out.

Immediacy of assent is still my criterion of observability, but emphatically and unequivocally so. The assent may be recanted, but its susceptibility to recantation becomes an independent dimension, *theoreticity*, for separate consideration.

The range of neural intakes to which the speaker's assent to the sentence is keyed will of course be vague along the edges. The speaker may hesitate over 'It's raining' in a fine mist, and over 'That's a swan' in the startling

presence of a black one. Vagueness of boundaries invests language at every turn, and I shall continue to take it in stride. But assent must be immediate when the stimulation is in the clean-cut range between vaguenesses. Immediate but fallible.

As always, the subject matter of the observation sentence is the real world, the “it” of the triangle, and it is the domain of fallible scientific theory. ‘It’s raining’, affirmed in full view of a rain-drenched window, is recanted on spotting the garden hose, and ‘Rabbit’ on spotting the toy. Such recantation reflects conflict of observations through theoretical connections between observation sentences. The degree of susceptibility to recantation measures how theoretic the observation sentence is. It is its *degree of theoreticity*.

An observation sentence that is perhaps minimally theoretic is ‘This looks blue’. I say ‘looks’ here, rather than ‘is’, to allow for the possibility that reflected light or environmental contrast may be affecting the color that the object would otherwise show. The reference of the sentence is still to an external patch or body, or is to become so with the flowering of reification.

This move brings me to terms with Kuhn, Hanson, and Feyerabend, who have objected to observation sentences that they are suffused with theory. They are indeed, I now agree, in all degrees, but their observational role remains. The immediacy of assent shows that in one way or another they have become conditioned as wholes to ranges of perceptually similar neural intakes and so serve as observation points in the experimental testing of hypotheses. The experimenter is still free, indeed advised, to examine the circumstances of his observation for possible experimental error, or even to try again.

Sophisticated observation sentences, such as the ones about copper and the hyperthyroid, are apt to be reducible to more primitive ones, delineating more directly sensory evidence. These will tend to be less theoretic by the stated criterion; that is, less susceptible to recantation. Some sophisticated ones, however, are not thus reducible. I think of the subtle traits that the wine expert learns to detect.

Such reduction, where possible, bolsters scientific theory; for the increased resistance to recantation of observation sentences increases their dependability in checking scientific theory.

A thoroughgoing reduction project of the kind, however, is surely a forlorn hope. It is utterly alien to what goes on and went on in the develop-

ment of language and science in the child and in the race. Observation sentences, already theoretic in varying degrees, are learned outright and helter-skelter by direct holophrastic conditioning. Further ones are synthesized along the way from bits of those at hand. It is not in general a matter of basing more theoretical ones on less theoretical ones. They vie with one another in a surging equilibrium of evidential claims. Such, in my sometime collaborator Joe Ullian's vivid phrase, is the web of belief.

Only limited reduction of overall theoreticity, then, is to be hoped for by reduction of some observation sentences to others. Some reduction may also be sought, however, in another way: by comparing several observation categoricals that are deducible from a theory as checkpoints. One of these may contain less theoretic observation sentences than others, and we could choose it for testing.

The last traces of analogy between my naturalized epistemology and the phenomenalist's dream of a reduction to sense data have now, it seems, been dissipated. Observationality becomes a simple behavioral matter of prompt response to stimulation, and overall reduction of the more theoretic to the less theoretic ceases to figure as an ideal.

My observation categoricals were always a schematic caricature of what goes on in testing scientific theory. They represented an adverse pair of observations as refuting a theory beyond redemption, whereas real life is less clean-cut. Now my recognition of theoreticity as a second dimension for observation sentences enhances the realism of my caricature by accommodating the inconclusiveness of an experimental verdict. The surging equilibrium becomes part of the picture.

Settling then for a modicum of irreducible theoreticity in the observation sentences, let us assess the consequent debilitation of the experimental evidence for science. A theory, let us suppose, has been found to imply an observation categorical. The experimenter has fulfilled the observation sentence that is the protasis of the categorical, to his satisfaction, and the predicted observation does not ensue. He concludes that the theory has failed the test. There are two places where the test itself can have gone wrong. Perhaps the experimental condition, the protasis, only seemed to have been fulfilled; something intruded to create the illusion. Or perhaps the predicted result did occur, but something intruded to obscure it.

At this point the morals to draw are two: a word of caution and a consoling reflection. The word of caution is 'Repeat your experiment'. It is a

superfluous caution, for the scientist would be eager to do just that—the more so since he was probably hoping to save his theory rather than refute it. The consoling reflection is that at worst the error is rejection of a tenable theory and not the retention of a false one. Ignorance is less pernicious than outright error.

I am indebted to Burton Dreben, as usual, for a critical reading of earlier drafts and helpful suggestions.

Three Networks: Similarity, Implication, and Membership

2000

This essay addresses the problem of how to account for our meeting of minds, for our being able to linguistically express agreement regarding external events despite wild dissimilarity of our nerve nets. An explanation is provided based on the instinct of induction, the instinct of similarity, and natural selection. There are three networks at play in the meeting of minds: *perceptual similarity*, the intersubjective harmony of similarity standards and thus the relation structuring the intake of perceptions; *implication*, the relation expressed by the universally quantified conditional and structuring our system of the world; and *class membership*, the relation structuring the domain of pure mathematics.

Language is for communicating with our fellows about matters of mutual concern. Its focus is not the private sensations that are our evidence for what is happening around us; focus is rather the things and events out where we all can jointly observe them. As I wrote on page one of *Word and Object*, “entification begins at arm’s length; . . . the points of condensation in the primordial conceptual scheme are things glimpsed, not glimpses.”

Our awareness of these external events, still, is due to causal chains leading from those external events to our sensory receptors and then continuing through our nervous systems. One would expect a pretty faithful homology between your nervous system and mine. How else could we account for the intersubjective agreement of our reports on external events that we jointly observe? Cameras similarly situated show similar agreement. Why shouldn’t we?

But no. The cameras are made alike; clearly we are not. Darwin reported in *The Origin of Species* that even two simple insects from the same swarm have widely dissimilar neural hookups. Let me quote more at length from *Word and Object*:

Different persons growing up in the same language are like different bushes trimmed and trained to take the shape of identical elephants. The anatomical details of twigs and branches will fulfill the elephantine form differently from bush to bush, but the overall outward results are alike.¹

So we are left with the problem of accounting for our meeting of minds. We keep finding ourselves in agreement regarding external events despite wild dissimilarity of our nerve nets. That is the problem. In *Word and Object* I coped lamely with it. I reflected that when I view a landscape, points in it do obviously trigger correspondingly placed receptors in my retina. If you and I view the landscape side by side, then each point in the landscape will trigger roughly corresponding receptors in your retina and mine. We might then be regaled with pretty homologous scenes, however dissimilar the intervening neural processing may be. I fondly hoped that somewhat analogous considerations might account also for intersubjective agreement in audition and other sense modalities.

This makeshift accommodation lingered in place for thirty-five years, from *Word and Object* in 1960 to *From Stimulus to Science* in 1995, when the fog lifted. The explanation of our meeting of minds is circuitous; through instincts and natural selection. I shall now try to sketch it without trying your patience.

One of the two instincts is that of induction. It is the tendency to expect any two similar perceptions to be followed respectively by two more perceptions that are in turn similar to each other. This tendency is the basis of conditioning, and hence of all learning, and hence cannot itself have been learned, so it must be instinctive.

Since it depends on some standard of similarity of perceptions, some such standard must be instinctive too. Our standard changes with experience, but has to begin in instinct in order to implement induction. So we have an instinct of similarity and an instinct of induction. So far so good; let me go on.

Our expectations by induction are overwhelmingly more successful than random guessing. Philosophers have wondered why. The answer is natural selection, which shapes our instinctive standards of similarity and therewith our inductive expectations. Natural selection has favored the survival of animals whose standards of perceptual similarity mesh well with trends in the environment.

1. W. V. Quine, *Word and Object* (Cambridge, MA: MIT Press, 1960), 8.

This, then, is the instinctual background of our meeting of minds. The rest of the explanation comes naturally. Since our ancestors' similarity standards were warped into harmony with a shared environment, one person's standards end up in harmony with another's. Thus we end up inheriting a harmony of similarity standards with our contemporaries.

All this is accounted for, we see, without assuming any direct similarity between your sensations and mine. It is a pre-established harmony, in Leibniz's phrase, but pre-established by natural selection.

Perceptual similarity, in its intersubjective harmony, is the life blood of language. Consider the mother teaching her infant the word 'dog'. The dog enters, the mother says the word, the child repeats it, and the mother manifests approval. Next day the dog enters again, the child says 'dog', and the mother again approves. The dog's two appearances were perceptually similar for the child and perceptually similar for the mother. The sounds of the word 'dog' as spoken by the mother on the first occasion and by the child on both occasions were all perceptually similar for the mother, and again all perceptually similar for the child. All this harmony is essential to this modest transaction. But there is no assumption of direct similarity between the child's perceptions and the mother's.

For simplicity I have been treating perceptual similarity as one-dimensional, a simple scale. Patricia Churchland has protested that the comparisons diverge into respects. However this may be, I think my present account holds in essentials. The more complex relation can surely be accommodated if it is needed.

The subject himself of course is not presumed to have any notion of his standards of perceptual similarity; it is a relation that the experimental psychologist can abstract from the subject's behavior under the reinforcement and extinction of responses. Speaking of his *standards* of similarity thus gives a wrong impression of self-consciousness. But I am at a loss for a better phrasing.

In daily discourse we all do indeed persistently declare things to be similar in various degrees, and a definition of that relation is notoriously lacking. It is an unconscious projection of perceptual similarity.

Perceptual similarity, we have been seeing, weaves the web of language in early childhood. What weaves the web of our increasingly scientific theory of nature, however, is rather *implication*, expressed by the universally quantified conditional ' $\forall x (Fx \supset Gx)$ ': 'Whenever this, that'; 'Wherever this, that'; 'If ever this, that'. This accommodates cause and effect. More-

over it marks the advent of the bound variable, and hence of reification; for to be is to be the value of a variable.

In speaking of the advent of reification I am doing logical reconstruction, not history. The small child learns words of animals and other things, and early man did likewise. But where to draw the line between the denotative and the syncategorematic use of a word is moot or arbitrary at the level of radical translation, and often even at home. In our own vernacular, objective reference seems most clearly intended in the use of the relative pronoun, and when we switch to symbolic logic it is unmistakable in the bound variable.

Implication, expressed by the universal conditional ' $\forall x (Fx \supset Gx)$ ', generates structure generously, establishing an instance of ' Gx ' for every established instance of ' Fx '. In particular it expresses causal connections, and accommodates the time dimension. My familiar example is 'If a cat eats spoiled fish and sickens, it thereafter avoids fish'. The pertinent value of the variable in this universal conditional is perforce an enduring creature, the same cat early and late. Such are the universal conditionals that reify enduring objects and accommodate temporal differences.

Deductive exploitation of one and other established universal conditionals takes us step by step downward to one or another observation categorical, we hope, which can then be subjected to experimental test. This is where science lines up for inspection. It is where the scientific network of implication, or universal conditionals, measures up to the perceptual network of similarity.

Such is the overwhelming utility of the universal conditional: it weaves the web of science, it binds it to its empirical evidence, and through its bound variable it determines our very ontology.

I would venture more: the universal conditional embodies the very *raison d'être* or survival value of ontology itself. What matters for the biological survival value of science is not what ontology it reveals, but what life-supporting and life-threatening events it conditionally predicts, ultimately in observation conditionals. For this purpose the ontology does not matter except numerically: any one-to-one correlates would serve as well. It is just a matter of isomorphism, as I have dramatized by proxy functions. I called it indeterminacy or inscrutability of reference.

So much for concrete objects. What of abstract ones? Where might occasion first arise for quantifying over abstract objects as values of variables? One thinks of biological taxonomy—species, genera, and their ilk. How-

ever, these can be treated as concrete objects if we construe physical objects generously as all occupants of space-time, however discontinuous. A species, then, can be arbitrarily identified with the mereological sum of its members, and a physical object then qualifies as a member of a species if it is an organism and a part of the species. Similarly for genera and other taxa.

A domain that does seem inescapably to call for quantification over abstract objects, however, is that of numbers. Counting does not require quantifying over numbers, but other uses of number do require it, and applied mathematics abounds in such uses. Numbers, for all their abstractness, must accordingly be accounted an integral part of our theory of the world.

Research in the logical foundations of mathematics, most voluminously by Whitehead and Russell, has shown that all of classical mathematics can be translated into pure set theory, and that set theory is translatable into elementary logic plus a single two-place predicate, that of class membership. It is a startling economy, but it leaves us with quantification over classes, which are abstract objects. So I see no hope of nominalism.

Just as perceptual similarity emerged as the relation structuring our intake of perceptions, and implication as expressed by the universal conditional was the relation structuring our system of the world, so we see class membership emerge as the relation structuring the domain of pure mathematics.

I take my resulting commitment to classes lightly, for all their abstractness. An exclusively concrete ontology would be welcome for its economy but little more, in view of the purely structural character of ontology's service to science. What I find more imperative is extensionality which condones reification of classes but challenges that of properties. The urgency of extensionality arises from the demand for individuation: no entity without identity. Thus it is that I reify classes but not properties; for we cannot satisfactorily distinguish two properties if they have the same extension.

I have no sense of loss as regards properties, but I was stopped by the idioms of propositional attitude; for these fail of extensionality but are indispensable—I have arrived at an accommodation here also, but must refer you to my latest thin book, *From Stimulus to Science*.

Confessions of a Confirmed Extensionalist

2001

I am neither an essentialist nor, so far as I know, an existentialist. But I am a confirmed extensionalist. Extensionalism is a policy I have clung to through thick, thin, and nearly seven decades of logicizing and philosophizing. I shall now define it, though I was heeding it before knowing the word or having the concept clearly in mind.

I shall call two closed sentences *coextensive* if they are both true or both false. Two predicates or general terms or open sentences are coextensive, of course, if they are true of just the same objects or sequences of objects. Two singular terms are coextensive if they designate the same object. And finally to the point: an expression is *extensional* if replacement of its component expressions by coextensive expressions always yields a coextensive whole. *Extensionalism* is a predilection for extensional theories.

In defining coextensiveness I lumped predicates, general terms, and open sentences together. They are what can be predicated of objects or sequences of objects, and in that capacity they all three come to the same thing. They are what the schematic predicate letters in quantification theory stand for. Open sentences are the most graphic of the three renderings. Two open sentences are coextensive if they have the same free variables and agree with each other in truth-value for all values of those variables.

The clarity and convenience conferred by extensionality are evident: free interchangeability of coextensive components *salva veritate*. When in particular those components are singular terms, indeed, their interchange-

ability would seem mandatory from any point of view; for this is simply the substitutivity of identity. Still, “Tom believes that Cicero denounced Catiline” and “Tom believes that Tully denounced Catiline” might be respectively true and false despite the identity of Cicero and Tully. We must come to terms with such cases either by compromising extensionalism or in some happier way, whereof more anon.

Meanwhile extensionalism faces a challenge from another quarter. Karel Lambert has argued¹ that an irreferential singular term such as “Pegasus” can disrupt extensionality. The predicates “flies if existent” and “flies and exists” are coextensive, since everything exists. But the sentence, “Pegasus flies if existent” is vacuously true, since Pegasus is not existent, whereas “Pegasus flies and exists” is false. So these two sentences are not extensional; their truth values are switched by switching coextensive predicates.

Happily this threat is thwarted by my practice, for lo these many decades, of treating all singular terms as singular descriptions—thus “Pegasus” as “(ιx)(x pegasizes).” We may think of the descriptions as defined contextually, following Russell. Under his definition, where “ F ” stands for any predicate, “ $F(\iota x)(x$ pegasizes)” becomes

$$\exists x(x \text{ and only } x \text{ pegasizes} \cdot Fx),$$

and this is false regardless of “ F .” So both of Lambert’s predicates, both “flies if existent” and “flies and exists,” must issue in falsehood when the subject is “Pegasus.” Then there is no breach of extensionality.

To reckon “Pegasus flies if existent” true, as Lambert does, we would have to analyze “Pegasus flies if existent” into “If Pegasus exists then Pegasus flies,” using two predicates. Then, treating “Pegasus flies and exists” correspondingly, we do find the sentences respectively true and false, as he claims. But our predicates are no longer the coextensive pair “flies if existent” and “flies and exists” that raised the problem. When singular description is evaporated into primitive notation, all is in order.

Lambert already recognized this, writing that my way with singular terms bypasses his challenge to extensionalism. This would indeed have been a good reason for that early move on my part, but actually my motive for it back then was just simplicity of foundations.

1. Karel Lambert, “Predication and Extensionality,” *Journal of Philosophical Logic* 3 (1974): 255–264.

The elimination of singular terms bears also, it might seem, on our question regarding the singular terms “Tully” and “Cicero.” But that is another story, and I shall continue to postpone it.

I have discussed extensionality thus far without mentioning classes. I hasten to do so, for classes are deemed the very paradigms of extensionality. Thus far I only defined extensionality of an *expression*: “Replacement of its parts by coextensives always yields a coextensive whole.” But classes are not expressions. They are objects, abstract objects. To bring them into the act I turn rather to the familiar expression for specifying a class: the class abstract “ $\{x: Fx\}$.” Being a singular term, “ $\{x: Fx\}$ ” may be thought of as defined as a singular description in the obvious way, namely as “ $(\iota y)((\forall x)(x \in y \equiv . Fx))$,” and then dissolved into Russell’s contextual definition of descriptions.

The expression “ $\{x: Fx\}$ ” is indeed extensional by my definition. For, if “ Fx ” and “ Gx ” stand for coextensive open sentences, then the singular terms “ $\{x: Fx\}$ ” and “ $\{x: Gx\}$ ” designate the same class, the same abstract thing, and such was my definition of coextensiveness of singular terms. If now we transfer the epithet “extensional” from the class abstract to the class itself, saying that classes are extensional just means that they are determined by their members.

The one difference between classes and properties, apart from metaphor and free association, is extensionality: a class is determined by its members. A property is not in general determined by its instances. I am told that among normal animals the property of having a heart and the property of having kidneys are coextensive, but we would never call them the same property. Classes are extensional, properties not.

We have no clear basis in general for saying what coextensive properties qualify as identical and what ones do not. In a word, properties lack a clear principle of *individuation*. Groping for such a basis, one settles for obscure talk of essence and necessity. Anything that can be described in terms of properties and not equally directly in terms of classes is unclear to my mind. I doubt that I have ever fully understood anything that I could not explain in extensional language.

Now that I have slipped back into the first person, I shall continue in that mode for a while; for the pertinent definitions are now explicitly before us. Afterward I shall resume the selfless business of making the world safe for extensionalism.

My first inarticulate hint of extensionalism may date from boy-

hood, when my liking for some Jewish schoolmates collided with someone's occasional derogatory remark about Jews. I reasoned in effect that a class is to be evaluated, if at all, by evaluating its members individually.

By my senior year in Oberlin College, 1929–1930, my extensionalism was full blown. I was majoring in mathematics with honors reading in mathematical logic. There was little mathematical logic in America, and none at Oberlin. But my professor, W. D. Cairns, had got me a reading list, culminating in Whitehead and Russell's *Principia Mathematica*.

My admiration for the three volumes, mostly in logical symbols, was almost unbounded. There was the spectacular analysis, the reduction of classical mathematics to a few basic notions of so-called logic, really logic and set theory. Further there was the rigor, explicitness, and clarity of the definitions, theorems, and proofs.

My admiration was not quite unbounded. It was bounded by the explanations in prose that were preposed and interposed as explanatory chapters and in briefer bits among the expanses of symbols.

Doubtful of the reality of classes, our authors undertook to accommodate them as fictions, eliminable by contextual definition in terms of purportedly more substantial things called propositional functions. These were functions which, when applied to objects, yielded propositions, which were the meanings of sentences. Thus the propositional functions of one variable were evidently identifiable with properties, and those of two or more variables were identifiable with relations "in intension."

Extensionality was seen by our authors as having to be worked for by devious contextual definition. The *intentional*, for all its failure of individuation, was the given. I suppose the reasoning was that, since the propositions and propositional functions are the *meanings* of sentences, adjective phrases, and verb phrases, surely they are clear to us insofar as we understand the expressions whose meanings they are.

If so, the authors' fallacy lay in tacitly taking in stride the giant step of reifying those meanings. Reification incurs the responsibility to individuate the reified entities, for there is no entity without identity. I suspect that our authors thus put undue weight on the adjective phrases that express the properties; for the phrases could differ conspicuously even if the properties did not.

Along with this regrettably intentional orientation, but independent of it, there was a detail that calls for notice only because of its disproportion-

ate consequences in subsequent literature. The truth-functional conditional “ $p \supset q$ ” or “not p or q ” was called material implication in *Principia* and read indifferently as “if p then q ” and “ p implies q .” On the face of it this was a grammatical aberration, independently of logical considerations. The grammar of “if p then q ” requires “ p ” and “ q ” to stand for sentences, whereas that of “ p implies q ” requires them to stand for nouns, in this case names of sentences.

C. I. Lewis lashed out against material implication,² but his objection was not to the grammar. He was protesting, and rightly, that the truth function “not-or” was a hollow mockery of implication, demanding as it does no semantic relevance of the one component sentence to the other. He supplanted the weak “ $\supset$ ” by a stronger connective “ $\rightarrow$ ” for what he called strict implication. He explained “ $p \rightarrow q$ ” as meaning “necessarily if p then q ,” and offered no further reduction. It was an emphatic departure from extensionalism, and it pioneered modal logic. Succeeding modal logicians have not all persisted in the grammatical confusion between “if-then” and “implies,” but they still sacrifice extensionality to their “necessarily if-then.”

I find the truth-functional conditional “ $p \supset q$ ” a satisfactory rendering of “if-then” in the indicative mood. Implication is quite another thing, in strength as well as in grammar. It is a relation between sentences, expressed by putting the verb “implies” between names of the sentences, and it is established by steps of deduction.

Whitehead and Russell’s regrettable use of “implies” virtually spoiled the word, prompting subsequent logicians to cast about for synonyms such as “entails” for the real thing. But I have been stubborn on that point.

Unlike the enduring intentionality of modal logic, the intentionality of propositional functions in *Principia* was mercifully just a flash in the pan. The propositional functions carry over only briefly into the formulas as values of quantified variables: only long enough to introduce classes as fictions by contextual definition. From there on the constructions proceed on greased wheels, greased by extensionality.

I was quite aware of these matters when I graduated from Oberlin in 1930, but my admiration for *Principia* was still almost unbounded. I proceeded to Harvard for graduate work in philosophy because Whitehead

2. C. I. Lewis, *A Survey of Symbolic Logic* (Berkeley: University of California Press, 1918), 222–239.

was in philosophy there. My doctoral dissertation was in mathematical logic still, under Whitehead's sponsorship, and was devoted to improving *Principia*. He seemed tickled by my little shortcuts and clarifications, except that I cannot have swayed him in my extensionalism. Anyway I imposed it in my dissertation. Individuals, classes, and sequences of them were all there was.

Mathematical logic, scarce in America, was sketchy even at Harvard. *Principia* was still the last word, and little was done even with it. I did not know that in Poland, Germany, and Austria the subject had been proceeding apace and that classes were the unquestioned staple from scratch. In 1931, while I wrote my dissertation, logicians in Europe were freely pursuing logic and set theory on the frugal conceptual basis of just truth functions, quantification, and membership.

This startling economy came of Kazimierz Kuratowski's discovery in 1921 that the ordered pair of x and y can be construed as the class whose two members are the class of x alone and the class of x and y . Kuratowski's definition had been anticipated by a slightly less elegant one in 1914 by a young American, Norbert Wiener, but his three-page paper in the *Proceedings of the Cambridge [England] Philosophical Society* escaped notice everywhere.³ Both contributions, Wiener's and Kuratowski's, escaped notice in America.

The economies over *Principia* that I achieved in my dissertation had thus been long surpassed, as I learned only after getting to Europe the following year, 1932, complete with doctorate. But in 1937 I published a more extreme reduction, assuming just class inclusion and class abstraction.⁴

The reduction of classical mathematics to one or another so meager a conceptual basis was amazing and illuminating, but calling it a reduction of mathematics to logic—*logicism*, in a word—gave the wrong message. Logic was proverbially slight and trivial. Mathematics proverbially ranged from the profound to the impenetrable, and reduction of mathematics to logic challenged belief, as indeed it well might. The reduction was to the unbridled theory of classes, or set theory, which, far from being slight and trivial, is so strong as to tangle itself in paradox until bridled in one way or

3. Norbert Wiener, "A Simplification of the Logic of Relations," *Proceedings of the Cambridge Philosophical Society* 17 (1914): 387–390.

4. "Logic Based on Inclusion and Abstraction," *Journal of Symbolic Logic* 2 (1937): 145–152. Reprinted in Quine, *Selected Logic Papers*, enlarged ed. (Cambridge, MA: Harvard University Press, 1995).

another. This is no fault of extensionalism, be it noted; properties are enmeshed in those paradoxes too. But what it shows is that the startling reduction of mathematics is to something far richer than traditional logic. I prefer to limit the term “logic” to the logic of truth functions, quantification, and identity, drawing the line at the reification of classes. Above that line we have set theory, the mathematics of classes.

I think of logic in this narrow sense as the grammar of strictly scientific theory. When a bit of science is thus regimented, the one place where extralogical vocabulary enters the picture is as interpretation of the schematic predicate letters. Within this grammar, extensionality prevails.

But extensionality had no evident charm for the Harvard philosophers during my two years of graduate study. Whitehead, Lewis, and Sheffer all swore by properties and propositions. It was with Carnap in Prague and Tarski, Lesniewski, and Łukasiewicz in Warsaw the following year that my extensionalism went without saying as a matter of course.

So much for reminiscence. But I have more to say of extensionalism, for properties and necessity are not its only hurdles. The domain of meanings of expressions is hopelessly intensional and in trouble over individuation. Propositions, seen as meanings of sentences, are conspicuous here. Properties themselves might be seen correspondingly as meanings of adjective phrases.

Properties, meanings, and necessity were violations of extensionality that I repudiated without regret. But the breach of extensionality that I cited early in this essay is of another sort: “Tom believes that Cicero denounced Catiline.” Those idioms of propositional attitude—belief, hope, regret, and the rest—are not to be lightly dismissed. It is not clear how to do without them. But there is a strategy by which, in the majority of cases, they can be rendered extensional.

I call it *semantic ascent*. It is the strategy of talking about expressions instead of using those expressions to talk about something more dubious. It already did us routine service in correcting Whitehead and Russell’s confused treatment of “implies” as a connective to be written between sentences. We lifted it to its rightful place as a transitive verb between names of sentences. Now the strategy can be used also on ascriptions of belief, artificially this time, by reconstruing belief as a relation between believers and sentences: thus

Tom believes “Cicero denounced Catiline”

or perhaps, for usage sticklers,

Tom believes true "Cicero denounced Catiline."

This reinterpretation of the propositional attitudes, as relating the person to the sentence, does not require him to know the language. The quoted sentence is the ascriber's expression of what he would be prompted to assert if he were in the state of mind in which he takes the subject to be. The effect of semantic ascent here is to seal the belief off from the context in which it is ascribed, so that Tom's disbelief of "Tully denounced Catiline" will not violate extensionality of the combined ascriptions of belief and disbelief.

Some ascriptions of propositional attitudes resist semantic ascent. For example,

There was an orator whom Tom believes to have denounced Catiline.

This example switches us from what are called propositional attitudes *de dicto* to attitudes *de re*. In ascribing a belief *de re* the ascriber ventures to assign a role within the ascribed belief to a denizen of the ascriber's real world. Such identifications can depend in varying degrees upon collateral information or conjecture about the subject's past behavior.⁵ Semantically these idioms *de re* of belief and other attitudes are comparable to the contrary-to-fact conditional, which depends so utterly for its truth upon tacit factual knowledge or assumptions that the interlocutors are assumed to share. The particles "you," "I," "here," "there," "now," and "then" are simpler examples of such dependence on circumstances of utterance. So the propositional attitudes *de re* belong with these extraneous idioms, ancillary to the self-contained language of scientific record.

Finally I turn to some further thoughts about our extensionalizing strategy of semantic ascent from use to mention. To the extensionalist eye, the ascent could seem paradoxically to be rather a descent from bad to worse, from frying pan to fire. We mention expressions by quoting them, and nothing could be less extensional on the face of it than quotation. Within a quotation you cannot supplant a word by even the strictest synonym without changing the designatum of the quotation, namely the quoted expression itself. Nothing could be farther from extensionality than quotation.

5. See Robert Sleigh, "On a Proposed System of Epistemic Logic," *Noûs* 2 (1968): 391–398.

This quandary is dispelled by recognizing the quotation as merely a graphic abbreviation, analyzable into spelling. We possess or coin a name for each of the simple signs of our language, and one for the space, and one for the operation of juxtaposition. Then we spell out the quoted expression. The spelling leaves no word of the quoted expression intact for replacement by a synonym. It thus blocks this latest little debacle before it begins.

Spelling is similar to polynomials and multi-digit numerals. It reduces similarly to truth functions, quantification, and predicates, with the help of contextual definition of singular description.

We have been seeing semantic ascent at work in achieving extensionality, but it has other uses. Someone's revolutionary scientific idea may prove difficult to promote because it undercuts one of the principles on which his colleagues' very thought and judgment depend. Holding that principle at bay for impartial assessment leaves the judge himself at a loss for a basis for judging. Semantic ascent, then, to the rescue. The innovator ascends from his subject matter to the formulas and laws themselves, dwells on their simplicity, and shows that they logically imply his strange new hypothesis together with essentials of the antecedent theory. The change of subject matter, from waves or quarks or fields to the formulas themselves, has bridged the gaps in his colleagues' intuition. Something like this perhaps went on at crucial points in the advance of science, though with no awareness of an ascent from use to mention.

Credits

Introduction: The Introduction is a modified version of “W. V. Quine Remembered,” *Harvard Review of Philosophy* 9 (2001): 106–111. Copyright © 2001 by *Harvard Review of Philosophy*.

Chapter 1: “Nominalism” was presented at the Harvard Philosophical Colloquium in Emerson B, Cambridge, MA, on March 11, 1946. The 87 handwritten 3×5 cards were transcribed by Paolo Mancosu and first published in Dean W. Zimmerman, ed., *Oxford Studies in Metaphysics*, vol. 4 (Oxford: Oxford University Press, 2008), 3–21.

Chapter 2: “On the Notion of an Analytic Statement” was presented at the University of Pennsylvania, Philadelphia, PA, on December 18, 1946 at 3:00 p.m. The 75 handwritten 3×5 cards were transcribed by Giancarlo Zanet. They are published for the first time in this volume.

Chapter 3: “Hume” was presented in 1946 as a summer term course, Philosophy 14c at Harvard University, Cambridge, MA. The 553 handwritten 3×5 cards were transcribed by James Buickerood. This chapter originally appeared as two articles: W. V. Quine, “1946 Lectures on David Hume’s Philosophy,” ed. James G. Buickerood, *Eighteenth-Century Thought* 1 (2003): 171–254, and J. G. Buickerood, “Addenda to Quine on Hume,” *Eighteenth-Century Thought* 2 (2004): 153–63. Copyright © 2003 and 2004, respectively, AMS Press, Inc. Used by permission. James Buickerood has merged the articles for publication in this volume.

Chapter 4: “The Importance of Logic for Philosophy” was presented at the Harvard Philosophy Club (public meeting) Phillips Brooks House, Cambridge, MA, on December 2, 1947, at 8:00 p.m. The 62 handwritten 3×5 cards were transcribed by Alexander Quine. They are published for the first time in this volume.

Chapter 5: “Where Logic Is Going” was presented at the Visiting Committee Dinner, Harvard University, Cambridge, MA, on December 13, 1947. The 22 handwritten 3×5 cards were transcribed by Alexander Quine. They are published for the first time in this volume.

Chapter 6: “Animadversions on the Notion of Meaning” was presented at the Philosophy Colloquium, University of Pennsylvania, PA, on December 6, 1949 (also as “Logical Aspect of the Problem of Meaning” on December 9, 1949, at the American Academy of Arts and Sciences, Boston, MA). The 18 handwritten 3×5 cards were transcribed by Giancarlo Zanet and Alexander Quine.

They are published for the first time in this volume.

Chapter 7: “The Entangled Philosophies of Mathematics” was presented at the Harvard Philosophy Club, Meeting Room, Harkness Commons, Cambridge, MA, on October 11, 1950, at 8:00 p.m. The 29 handwritten 3×5 cards (annotated with an expected duration of 52 minutes) were transcribed by Alexander Quine. They are published for the first time in this volume.

Chapter 8: “Meaning” was presented at the Center for Advanced Study in the Behavioral Sciences, Stanford, CA, in March 1959. The 12 handwritten 3×5 cards were transcribed by Alexander Quine. They are published for the first time in this volume.

Chapter 9: “The Way the World Is” was presented at Harvard’s 350th celebration, Cambridge, MA, in March 1986. It was cited in Burton Dreben, “Putnam, Quine—and the Facts,” *Philosophical Topics* 20, no. 1 (1992): 293–315. The eight-page 8.5 × 11 typescript is published for the first time in this volume.

Chapter 10: “Pressing Extensionality” was presented at the Føllesdal Festival, Oslo, Norway, on June 13, 1992. The 16 handwritten 4×6 cards were transcribed by Alexander Quine. They are published for the first time in this volume.

Chapter 11: “The Innate Foundational Endowments” is dated March and May 1996. The eleven-page cut-and-paste 8.5 × 11 manuscript was transcribed by Alexander Quine. It is published for the first time in this volume.

Chapter 12: “The Growth of Mind and Language” was given as the Karl Jaspers lecture at the University of Oldenburg, Germany, on June 5, 1997, under the title “Die Entwicklung von Bewußtsein und Sprache” and was published in German in *Logos* 5 (1998): 23–35. We are grateful to Professor Michael Sukale, University of Oldenburg, who was able to provide us with Quine’s original English manuscript, which is published for the first time in this volume.

Chapter 13: “Relations and Reason” was originally published in *Technology Review* 41 (1939): 299–301, 324–327. Reprinted with permission.

Chapter 14: “On the Reasons for Indeterminacy of Translation” was originally published in *The Journal of Philosophy* 67, no. 6 (March 26, 1970): 178–183. Copyright © 1970 by *The Journal of Philosophy*. Reprinted with the permission of The Journal of Philosophy.

Chapter 15: “Methodological Reflections on Current Linguistic Theory” was originally published in *Synthese* 21: 386–398. Copyright © 1970 by D. Reidel Publishing Company. It was reprinted in Donald Davidson and Gilbert Harman,

Semantics of Natural Language (Dordrecht, Holland: D. Reidel Publishing, 1972). Reprinted with the kind permission of Springer Science and Business Media.

Chapter 16: “On Empirically Equivalent Systems of the World” was originally published in *Erkenntnis* 9 (November 1975): 313–328. Copyright © 1975 by D. Reidel Publishing Company. Reprinted with the kind permission of Springer Science and Business Media.

Chapter 17: “Mind and Verbal Dispositions” was originally published in Samuel Guttenplan, ed., *Mind and Language* (Oxford: Oxford University Press, 1975), 83–95. Copyright © 1975 by Oxford University Press. Reprinted in W. V. Quine, *Quintessence*, ed. Roger Gibson (Cambridge, MA: Harvard University Press, 2004), 313–325. Reprinted with the permission of Oxford University Press.

Chapter 18: “The Nature of Natural Knowledge” was originally published in Samuel Guttenplan, ed., *Mind and Language* (Oxford: Oxford University Press, 1975), 67–82. Copyright © 1975 by Oxford University Press. Reprinted in W. V. Quine, *Quintessence*, ed. Roger Gibson (Cambridge, MA: Harvard University Press, 2004), 287–300. Reprinted with the permission of Oxford University Press.

Chapter 19: “Facts of the Matter” was originally presented at the sixth annual Oklahoma Conference on Philosophy held at the University of Oklahoma, March 31 to April 3, 1976. It was published in R. W. Shahan and K. R. Merrill, eds., *American Philosophy from Edwards to Quine* (Norman, OK: University of Oklahoma Press, 1977), 176–196. Copyright © 1977 by the University of Oklahoma Press, Norman. All rights reserved. It was reprinted in *Southwestern Journal of Philosophy* 9, no. 2 (1979): 155–169. Reprinted with the permission of the University of Oklahoma Press.

Chapter 20: “Cognitive Meaning” was originally published in *The Monist* 62, no. 2 (April 1979): 129–142. Copyright © 1979 *THE MONIST: An International Quarterly Journal of General Philosophical Inquiry*. Peru, Illinois, USA 61354. Reprinted by permission.

Chapter 21: “Grammar, Truth, and Logic” was originally published in S. Kanger and S. Ohman, eds., *Philosophy and Grammar: Papers on the Occasion of the Quincentennial of Uppsala University* (Dordrecht: D. Reidel Publishing, 1980), 17–28. Copyright © 1980 by D. Reidel Publishing Company. Reprinted with the kind permission of Springer Science and Business Media.

Chapter 22: “Ontology and Ideology Revisited” was originally published in *The Journal of Philosophy* 80, no. 9 (September 1983): 499–502. Copyright © 1983 by *The Journal of Philosophy*. Reprinted with the permission of The Journal of Philosophy.

Chapter 23: “Relativism and Absolutism” was originally published in *The Monist*

- 67, no. 3 (July 1984): 293–296. Copyright © 1984 *THE MONIST: An International Quarterly Journal of General Philosophical Inquiry*. Peru, Illinois, USA 61354. Reprinted by permission.
- Chapter 24:* “States of Mind” was originally published in *The Journal of Philosophy* 82, no. 1 (January 1985): 5–8. Copyright © 1985 by *The Journal of Philosophy*. Reprinted with the permission of The Journal of Philosophy.
- Chapter 25:* “The Sensory Support of Science” was originally published in *Discursos* (Granada: Universidad, 1986), 31–59 [even-numbered pages in Spanish; odd-numbered pages in English]. Reprinted with permission.
- Chapter 26:* “Panel on Reference” was originally published in *Philosophie et culture: Philosophy and Culture: Proceedings of the XVII World Congress of Philosophy* (Montreal: Editions Du Beffroi/Editions Montmorency, 1986), 225–227. Reprinted with permission.
- Chapter 27:* “Indeterminacy of Translation Again” was originally published in *The Journal of Philosophy* 84, no. 1 (January 1987): 5–10. Copyright © 1987 by *The Journal of Philosophy*. Reprinted with the permission of The Journal of Philosophy.
- Chapter 28:* “Mind, Brain, and Behavior” was originally published in Aaron J. Brownstein, ed., *Progress in Behavioral Studies*, vol. 1 (Mahwah: Lawrence Erlbaum Associates, 1989), 1–6. Reprinted with permission.
- Chapter 29:* “The Elusiveness of Reference” was originally published in Michael Sukale, ed., *Sprache, Theorie und Wirklichkeit: Studia Philosophica et Historica* (Frankfurt am Main: Peter Lang, 1990), 13–24. Reprinted with permission.
- Chapter 30:* “The Phoneme’s Long Shadow” was originally published in T. N. Headland, K. L. Pike, and M. Harris, eds., *Emics and Etics: The Insider/Outsider Debate*, *Frontiers of Anthropology* (Newbury Park, CA: Sage Publications, 1990), 164–167. Reprinted with permission.
- Chapter 31:* “Three Indeterminacies” was originally published in R. B. Barrett and R. F. Gibson, eds., *Perspectives on Quine* (Oxford: Blackwell, 1990), 1–16. Reprinted with permission.
- Chapter 32:* “Preface” was originally published in *The Logic of Sequences: A Generalization of Principia Mathematica*, Harvard Dissertations in Philosophy (New York: Garland Publishing, 1990). This was the first publication of the original text of Quine’s 1932 doctoral dissertation. The dissertation had previously been published in a very different form as *A System of Logistic* (Cambridge, MA: Harvard University Press, 1934). Reprinted with permission.
- Chapter 33:* “Two Dogmas in Retrospect” was originally published in the *Canadian Journal of Philosophy* 21 (1991). Copyright © 1991 by the *Canadian Journal of Philosophy*. Reprinted with the permission of the Canadian Journal of Philosophy in W. V. Quine, *Quintessence*, ed. Roger Gibson (Cambridge, MA: Harvard University Press, 2004), 54–63. Reprinted with permission of the Canadian Journal of Philosophy.

- Chapter 34:* “Structure and Nature” was originally published in *The Journal of Philosophy* 89, no. 1 (January 1992): 6–9. Copyright © 1992 by *The Journal of Philosophy*. Reprinted with the permission of *The Journal of Philosophy*.
- Chapter 35:* “Commensurability and the Alien Mind” was originally published in *Common Knowledge* 1, no. 3 (Winter 1992): 1–2. Copyright 1992, Duke University Press. Used by permission of the publisher.
- Chapter 36:* “In Praise of Observation Sentences” with “Appendix on Neural Intake” was originally published in *The Journal of Philosophy* 90, no. 3 (March 1993): 107–116. Copyright © 1993 by *The Journal of Philosophy*. Reprinted with the permission of *The Journal of Philosophy*.
- Chapter 37:* “Truth” was originally published in G. Fløistad, ed., *Philosophical Problems Today*, vol. 1 (Amsterdam: Kluwer Academic Publishers, 1994), 1–20, having been adapted from W. V. Quine, *Philosophy of Logic* (Englewood, NJ: Prentice Hall, 1970), 36–46. Copyright © 1994 Kluwer Academic Publishers. Reprinted with the kind permission of Springer Science and Business Media.
- Chapter 38:* “Promoting Extensionality” was originally published in *Synthese* 98, no. 1 (January 1994): 143–151. Copyright © 1994 by D. Reidel Publishing Company. Reprinted with the kind permission of Springer Science and Business Media.
- Chapter 39:* “Indeterminacy without Tears” was originally published in *Il Cannocchiale* 1/2 (1994): 57–58.
- Chapter 40:* “Assuming Objects” was originally published in *Theoria* 60, no. 3 (1994): 226–231. Reprinted with permission.
- Chapter 41:* “Naturalism; Or, Living within One’s Means” was originally published in *Dialectica* 49 (1995): 251–261. Copyright © 1995 by *Dialectica*. Reprinted with the permission of *Dialectica* and its editor in W. V. Quine, *Quintessence*, ed. Roger Gibson (Cambridge, MA: Harvard University Press, 2004), 275–286. Reprinted with permission.
- Chapter 42:* “Progress on Two Fronts” was originally published in *The Journal of Philosophy* 93, no. 4 (April 1996): 159–163. Copyright © 1996 by *The Journal of Philosophy*. Reprinted in W. V. Quine, *Quintessence*, ed. Roger Gibson (Cambridge, MA: Harvard University Press, 2004), 169–173. Reprinted with the permission of *The Journal of Philosophy*.
- Chapter 43:* “The Flowering of Thought in Language” was presented at the annual conference of the Royal Institute of Philosophy, University of Reading (UK), September 1996. It was originally published in John Preston, ed., *Thought and Language* (Cambridge: Cambridge University Press, 1997). Copyright © The Royal Institute of Philosophy and Contributors 1997. Reprinted with the permission of Cambridge University Press.
- Chapter 44:* “I, You, and It: An Epistemological Triangle” was originally published in Alex Orenstein and Petr Kotatko, eds., *Knowledge, Language and Logic: Questions for Quine*, Boston Studies in the Philosophy of Science (Dordrecht:

Kluwer Academic Publishers, 1999), 1–6. Copyright © 1999 by Kluwer Academic Publishers. Reprinted with the kind permission of Springer Science and Business Media.

Chapter 45: “Three Networks: Similarity, Implication, and Membership” (listed in the original program as “The Pre-Established Harmony of Subjective Perceptual Similarity”) was presented in Boston in August 1998. It was originally published in Akihiro Kanamori, ed., *Proceedings of the Twentieth World Congress of Philosophy*, vol. 6: *Analytic Philosophy and Logic* (Charlottesville: Philosophy Documentation Center, 2000), 287–291. Reprinted with permission.

Chapter 46: “Confessions of a Confirmed Extensionalist” was originally published in J. Floyd and S. Shieh, eds., *Future Pasts* (New York: Oxford University Press, 2001), 215–221. Copyright © 2001 by Oxford University Press. Reprinted in W. V. Quine, *Quintessence*, ed. Roger Gibson (Cambridge, MA: Harvard University Press, 2004), 329–337. Reprinted with the permission of Oxford University Press.

Index

- a priori*, 12, 18, 60, 112, 134, 135, 142–145, 161; *a posteriori*, 18, 143
- abstraction, 76, 77, 80, 84, 85, 87, 186, 199–201, 266, 309–311, 313, 357, 388, 503
- Ackermann’s construction, 470
- aesthetics, 120, 169, 364
- algebra, 16, 23, 87, 88, 89, 97, 105, 126, 144, 150, 196, 197, 200, 224
- American Academy of Arts and Sciences (venue), 508
- American Philosophical Association (venue), 323, 393
- analyticity, 2, 152–156, 223, 392–397, 453, 466
- ancestral, 16, 475
- anger, 122
- Anselm of Canterbury, 54, 92
- Apostel, Leo, 223
- Arabic numeral, 21, 49, 148
- Aristotelian essentialism. *See* essentialism
- Aristotle/Aristotelian/Aristotelean, 49, 53, 59, 60, 74, 107, 143, 153, 156, 179, 183, 186, 197, 429, 475
- assert(s, ed)/assertion, 13, 15, 18, 72, 82, 90, 91, 98, 126, 145, 154, 242, 343, 388, 410, 411, 413, 451, 505
- attributes, 11, 19, 20, 29, 54, 60, 65, 74, 75, 78, 89, 98, 102, 111, 116, 143, 163, 210, 253, 259, 285, 455
- Austria, 503
- awareness, 57, 60, 121, 131, 133, 219, 257, 258, 285, 315, 324, 327, 493, 506
- axiom, 18, 52, 54, 56, 151, 190, 201, 216, 233, 388, 391, 402, 433
- Babylonians (example), 354–355
- bachelor, 24, 25, 32, 33, 154, 155, 163, 292, 293, 395, 396
- Balaam’s ass, 62
- beauty, 114–116, 120, 129, 169
- Begriffsschrift* (Frege), 143
- behaviorism, 284, 341, 347, 348; armchair behaviorism, 164
- being: “to be is to be the value of a variable,” 1, 14, 146, 186, 451, 496
- Benacerraf, Paul, 401, 405
- benevolence, 122, 131
- Bentham, Jeremy, 460
- Bergström, Lars, 370, 372, 476, 489
- Berkeley, E. C. (Prudential Insurance), 207–208
- Berkeley, George (Bishop of Cloyne), 53, 61, 63, 66–67, 76–77, 85, 97–98, 104
- Bickerton, Derek, 186–187
- bilingual, 164, 165, 447
- Bishop, Erret, 433
- Bishop of Cloyne. *See* Berkeley, George
- bizarreness, 165, 218, 219, 429
- Blake, William, 417
- Boole, George/Boolean algebra, 16, 197–199
- Boolos, George Stephen, 402
- Boston (example), 116, 148, 152, 375, 420, 452
- Boston (MA), 62, 393
- Brentano, Franz, 325, 350
- Bridgman, Percy Williams, 143
- Brouwer, Luitzen E. J., 159, 431
- Buickerood, James, 8, 36, 507
- Burali-Forti, Cesare, 427
- Caesar, Julius (examples and problem), 90, 290–292, 470, 482
- Cairns, W. D., 501
- California, 374
- Cambridge, 154, 388, 503

- Campbell, Donald, 465
Canary Islands, 374
Cantor, Georg, 427
Carnap, Ina, 391
Carnap, Rudolf, 12, 33–34, 83, 103, 141–142, 152, 154, 219, 305, 335, 391–394, 397, 399, 410, 412, 418–419, 462, 504; *Der logische Aufbau der Welt*, 83, 141, 391, 399, 418; *Logische Syntax der Sprache*, 391
Carnap archives in Pittsburgh, 391
Carroll, Lewis, 67, 91, 138
Cartesian, 137, 258, 467
causality, 3, 41, 53, 67, 74, 89, 90, 91, 95, 97, 98, 104, 105, 107, 108, 109, 119, 135, 175, 349, 416
Center for Advanced Study in the Behavioral Sciences, 163
certainty, 50–53, 57–59, 85–87, 89, 90, 105, 106, 113, 126
Chinese classifiers, 375
Chisholm, Roderick, 350
Chomsky, Noam, 215, 217–222
Church, Ralph W., 39–40, 44–46, 72, 79, 149
Churchland, Patricia, 180, 349, 495
Churchland, Paul, 349, 400
Clarke, Samuel, 90
cogito ergo sum, 54, 55
commensurability, 407, 408, 413
concatenation, 425, 442
concept, 24, 26–29, 32, 35, 54, 100, 133, 156, 201, 202, 209, 248, 251, 365, 437, 472, 498
conceptual scheme, 14, 18, 158, 314, 406, 409, 493
concomitant variation, 94, 96, 272
concrete, 10–12, 29, 75, 77, 81, 188, 213, 271, 277, 281, 298, 316, 358, 376, 401–403, 422, 456, 459, 471, 496, 497
conscious(ly)/consciousness/unconscious, 101, 110, 183, 216, 217, 218, 320, 398, 399, 417, 419, 495; self-consciousness, 57, 495
constant conjunction, 91, 92, 94
construction: of contexts, 85; of ideas, 85, 89; of terms, 85
constructivism, 159, 162
contiguity, 37, 73, 74, 76–79, 81, 92, 93, 95, 96, 102, 104, 107, 114, 118, 125
contradiction, 16, 27, 75, 77, 91, 125, 159, 160, 185, 270, 336, 362, 396, 426–428
contrariety, 78, 79, 87, 96
Copenhagen interpretation, 467
cosmological argument, 110
Coué, Émile, 109
Cowan, Joseph, 231
Craig, William, 240–241
creole, 186–187
Crick, Francis, 349
Czechoslovakia, 391

Darwin, Charles/Darwinian, 60, 176, 179, 183–184, 260, 349, 474–475, 479, 493
Davidson, Donald, 249, 321, 325, 336–337, 350, 362, 370–372, 384–385, 415–416, 435, 437, 471, 474, 485
de re, 173–174, 295–298, 355–357, 442–443, 505
Dedekind, Richard, 158, 200
deduction, 29, 54, 55, 87, 138, 148, 382, 438, 439, 466, 502
definition, inductive, 373, 426, 427; definability, 425
Democritus, 63–65
demonstrative, 86, 88–90, 97, 126, 338, 352, 353, 421
Dennett, Daniel, 417
denote/denotation, 20, 21, 67, 83n, 152, 153, 188, 198, 199, 212, 275, 276, 287, 295, 297, 324, 353, 355, 356, 358, 375, 376, 391, 410, 412, 415, 416, 450, 451, 457–459, 461, 469, 496
Der logische Aufbau der Welt (Carnap), 83, 141, 391, 399, 418
Descartes, René, 53–59, 61–62, 65, 70, 86–87, 90–92, 98, 108, 176, 178–181, 244, 271, 347, 462
Dewey, John, 398
Dialogues Concerning Natural Religion (Hume), 38, 43, 50, 110–113
dictionary, 32, 33, 298, 299, 301
Dixon, Robert M. W., 216, 218
Donnellan, Keith, 353
Dreben, Burton, 209, 231, 302, 322, 323, 341, 370, 372, 386, 389, 437, 473–474, 492, 508
Dreyfus, Hubert, 374
Duhem, Pierre/Duhemian/Duhem-Quine thesis, 168, 228–229, 235, 337, 345, 380, 394, 398, 431
Dummett, Michael, 431–432
Dutch, 486

- Eaton, M., 57
economics, 139, 168, 190, 236, 307, 365,
388, 433, 444, 457, 463, 503
Eddington, Sir Arthur S., 10, 21, 414
Eddington's number, 21
Eddy, Mary Baker, 109–110
Einstein, Albert, 224, 408, 422
embodiment/disembodied, 81, 89, 104, 106,
162, 182, 311, 323, 331, 342, 439, 453,
483, 496
emotions, 68, 73, 82, 93, 114, 124, 125, 478
empathy, 120, 172, 343, 366, 371, 372, 440
empiricism, 2, 58, 66, 83, 89, 95, 97, 102,
107, 135, 141, 142, 144, 145, 146, 153,
154, 156, 390, 436, 469
Enquiry Concerning Human Understanding,
An (Hume), 38, 40, 43, 45, 50, 87–89, 97,
105–106, 108–110, 112–113, 124, 134,
136
entail/entailment, 73, 86, 154, 156, 503
Epimenides the Cretan, 423
epistemology, 1–3, 18, 24, 60, 103, 141, 145,
258, 317, 322, 327, 328, 361, 368, 369,
371, 391, 397, 403, 409–414, 418, 431,
462, 464, 467, 468, 473, 474, 485, 491
essence, 53, 54, 59–61, 67, 68, 87, 95, 117,
127, 153, 191, 337, 339, 355, 356, 394,
431, 466, 488, 500
essentialism, 173; Aristotelian essentialism,
153, 156
ethics, 50, 108, 126, 127, 134, 174
Euclidean, 434
Evening Star, 100, 152, 292, 294–297, 339,
353–356
evil, 109, 110n, 125, 127, 133, 461; problem
of evil, 109
existence. *See* being
explanation, 49, 56, 64, 71, 95, 102, 118,
122, 124, 125, 178, 184, 248, 253–256,
273, 286, 301, 351, 381, 400, 408, 493,
494, 495, 501
extensionality, 75, 172–175, 190–191, 438,
440–442, 444–445, 497, 498–502, 504–
506
faith, 9, 31, 89, 97, 98, 112–113, 220, 362,
365, 493
Feferman, Solomon, 190, 433
Feigl, Herbert, 337
Fermat, Pierre de, 149
Fermat's last theorem, 149
Feyerabend, Paul, 409, 490
fiction, 33, 83, 99, 132, 169, 174, 419, 501,
502
Findlay, John Niemeyer, 41
finitude/infinity, 82, 167, 216, 240
Floyd, Juliet, xi
Fodor, Jerry Alan, 222
Føllesdal, Dagfinn, 172–173, 175, 335–336,
339, 354–356, 370, 385, 437, 444, 474,
508
formalism, 157, 159–162
France, king of. *See* king of France
Frank, Philipp, 394
Frege, Gottlob, 143, 149, 152, 158, 186, 200,
339, 353, 387, 440, 460, 470; Frege's way,
172, 255; *Begriffsschrift*, 143;
Grundgesetze der Arithmetik, 143;
Grundlagen der Arithmetik, 143
Frege-Russell constructions, 470
French, 288, 357
fusion, 19, 20

Garland Publishing, 391
Gaulish, 364
gavagai, 209, 211–213, 344, 371–372, 374–
375, 474–475
Geach, Peter Thomas, 220, 225–226, 317,
360, 452
genetics, 103, 104, 178, 252, 264, 265, 349,
464
Geneva, 223
geometry, 52, 54, 56, 77, 83, 84, 86–89, 97,
105, 126, 149, 196, 373, 434, 485
George V, King. *See* King George V
Germany, 182, 503
gestalt, 327, 328, 399
Gettier, Edmund, 176
Gibson, Roger F., Jr., xi, 38, 49, 336, 385
God, 53–58, 90–92, 98, 108–111, 163, 180–
181, 429
Gödel, Kurt, 18–19, 30–31, 144–145, 149–
150, 161–162, 201, 213, 240–241, 388,
396, 433, 459; axiom of constructability,
433; completeness, 30, 145, 150; incom-
pleteness, 18, 19, 31, 144–145, 150, 161–
162, 201, 240–241, 396, 459
Goldbach's problem, 149
Goodman, Nelson, 1, 19–21, 23, 203, 393
Granada (venue), 510

- Green, Thomas Hill ("Hume's works"), 39, 44
- Greek(s), 296, 455
- Greig, J. Y. T. ("David Hume"), 44, 90
- Grelling's paradox, 426
- Grünbaum, Adolf, 381
- Grundgesetze der Arithmetik* (Frege), 143
- Grundlagen der Arithmetik* (Frege), 143
- Haas, William, 305
- habitual, 307, 451, 454, 488
- Hahn, Hans, 12
- hallucination, 119
- Hanson, N. R., 409, 490
- Harman, Gilbert, 213–214
- Harris, Marvin, 366
- Harris, Zellig, 224
- Harvard University, 36, 43, 148, 166, 387, 391–392, 502–504; Philosophical Colloquium, 9; Philosophy Club, 137, 157; Press, 388–389; Summer School, 8
- hatred, 120–123, 131, 136
- heap, 10, 11, 19
- Heidegger, Martin, 138
- Helmer, Dietrich, 142
- Hempel, Carl G. (Peter), 142, 392, 394
- Hendel, C. W., 38–39, 44–45
- Henderson, L. J., 65
- Heracleitus of Ephesus, 102
- Hesperus, 292, 294–297
- Heyting, Arend, 159
- Hilbert, David, 159–161
- Hindi, 365
- history, 22, 36, 43, 45, 46, 50, 53, 61, 131, 168, 208, 224, 283, 285, 331, 386, 414, 455, 462, 464, 467, 496
- Hobbes, Thomas, 65, 90, 131
- holism, 4, 168, 228–230, 335, 380–383, 393–395, 397, 398, 431, 432, 454, 467
- holophrastic, 212, 345, 374, 403, 405, 411–414, 461, 469, 476, 477, 491
- homeophone, 365
- Hubel, David, 329, 349, 417
- Hume, David, 8, 24, 36–136, 137, 140, 153, 271, 327; *Dialogues Concerning Natural Religion*, 38, 43, 50, 110–113; *An Enquiry Concerning Human Understanding*, 38, 40, 43, 45, 50, 87–89, 97, 105–106, 108–110, 112–113, 124, 134, 136; *Treatise of Human Nature*, 37–38, 40–43, 45–46, 48–50, 53, 62, 68–80, 82–85, 89–90, 105, 108–110, 113, 126–127
- Huntington, Edward Vermilye, 391
- Hutcheson, Francis, 127–128
- hypothetico-deductive, 261, 262, 276, 462, 466, 467, 469
- idea, simple, 41, 60, 61, 68–73, 77, 79–81, 106, 140
- idealism, 10, 60, 171, 179, 254, 280
- identity: "no entity without identity," 297, 455, 497, 501
- illusion, 52, 85, 163, 248, 256, 257, 258, 270, 491
- impressions. *See* perception
- indecision, 93, 109, 218
- indeterminacy of translation, 2, 209–213, 234n, 237, 251, 341, 344–346, 368, 373–374, 376, 407, 420
- indexical, 174
- individuation, 3, 102, 234–237, 275, 281, 417, 454, 475, 497, 500, 501, 504
- Indo-European (language roots), 227
- infinity, 20, 97, 190, 343, 442, 483
- injustice, 106, 111, 131
- inscrutability. *See* ontological relativity
- intension, 153, 162, 174, 190, 210, 325, 350–351, 361, 390–392, 501, 504
- intention, 157, 421, 429, 501–502
- intersubjective, 184, 263, 347, 348, 370, 371, 411, 416n, 456, 457, 474–476, 480, 486, 493–495; intersubjectivity, 103, 167, 373, 465
- intuition, 54–55, 87, 90n, 157–159, 163, 506; intuitionism, 157, 159–162
- intuitive, 18, 88, 90, 158, 159, 202, 234, 235, 250, 295, 307, 339, 405, 470
- irradiation pattern. *See* nerve
- Italian, 247
- Ivanhoe* (Scott), 287
- James, William, 73, 403, 462
- James-Lange theory, 73
- Janus-faced, 411–412
- Japanese (language), 213, 218, 345, 357, 375
- Jeffreys, Harold, 83, 85
- Jespersen, Otto, 215, 217
- Jewish, 501

- Johnson, William Ernest, 77
 Jungle (language), 341–344, 346, 369, 373–374, 413
- Kalaba (example), 366
 Kant, Immanuel/Kantian, 24, 32, 40, 48, 59, 88, 134–135, 143–144, 254, 379, 383, 472
 Katz, Jerrold, 222–223
 Keats, John, 169
 Kelvin, Lord, 17–18
 Kemp. *See* Smith, Norman Kemp
 kinaesthetic, 410
 kind, natural, 3, 179
 King George V (example), 227
 king of France (example), 139–140, 353
 Kirk, Robert, 373
 Korzybski, Alfred, 483
 Kripke, Saul, 339, 354, 356, 444
 Kronecker, Leopold, 159, 189, 242
 Kuhn, Thomas, 409, 477, 490
 Kuratowski, Kazimierz, 503
- Laird, John, 39, 40, 44, 46, 74, 77–78, 107
 Lambert, Karel, 499
 Latin, 215, 224, 296
 Lauener, Henri, 461, 466, 468, 471–472
 Leibniz, Gottfried, 56–58, 61, 88, 90, 106–109, 179, 475, 495
 Lepore, Ernest, 49
 Lesniewski, Stanisław, 504
 Lett, James, 367
 Lévy-Bruhl, Lucien, 88, 165
 Lewis, Clarence Irving, 143, 153, 391, 397, 444, 502, 504
 Lewis, David, 302, 401–402, 405
 Linnæus, Carl, 224
 Locke, John, 41, 48–49, 53, 57–68, 71–72, 74, 80–81, 85, 90, 95, 97, 105, 126, 131, 134, 137, 140, 176–177, 271
 logical: particles, 172, 305, 332, 342, 403, 425, 466; structure, 25, 26, 30, 185, 224–227, 382, 405, 415, 455, 460; truth, 25–26, 28–33, 88, 99, 106, 149, 154, 303, 305, 306, 308, 311–314, 335, 381, 396, 432; validity, 149; words, 26
 logicism, 157, 158, 160, 161, 503
Logische Syntax der Sprache (Carnap), 391
 London, 50, 174
 Lorenzen, Paul, 433
- Löwenheim, Leopold, 149, 282
 Łukasiewicz, Jan, 504
- MacColl, Hugh, 444
 Malebranche, Nicolas, 53, 91, 108
 malice, 122–123
 Mancosu, Paolo, xi, 7–8, 507
 Martin, Richard, 394
 Marx, Karl, 132
 Mary. *See* Queen Mary
 materialism, 254, 255, 282, 347, 467
 McGovern, Margaret Q. *See* Quine, Margaret
 meaningful, 14, 35, 89, 135, 140–142, 146, 169, 376, 385, 394, 395, 402, 432, 483; meaningless, 9, 11, 12, 14, 22, 91, 97, 104, 144, 146, 152, 156, 169, 405, 430, 435, 454, 468; Sinn, 152
 mechanical method, 148
 memory, 45, 68, 72–74, 76, 91–92, 97, 104, 107, 329, 390, 409, 418
 Mendeléeff, Dmitri, 64
Meno (Plato), 57
 mental entities, 11, 163, 244, 462; mentalism, 252, 349, 392, 473
 metaphor, 22, 154, 211, 274, 398, 399, 472, 500
 metaphysics, 1, 8, 9, 12, 14, 53, 56, 57, 60, 67, 108, 135, 153, 169, 173, 179, 325, 337, 355, 366, 395, 405, 406, 462, 475
 Michelson-Morley, 35
 Mill, John Stuart, 38, 96, 254, 272
 Miller, James Grier, 146, 182
 Missourian insistence, 319
 modality, 2, 3; modal logic, 173, 174, 191, 391, 392, 443, 444, 502; modal operators, 278, 358
 molecular, 11, 19, 171, 177, 234, 235, 237, 241, 258, 269, 270, 328, 329, 398, 403, 405, 410
 Molière (Jean-Baptiste Poquelin), 62
 Molyneux, William, 126
 Moncrieff, Bruse, 77
 monism, 325, 347
 moral, 43, 110, 126–131, 132, 133, 245, 250, 316, 429, 491
 Morning Star, 100, 152, 294–295, 297, 339, 353–356
 morphological, 364

- Næss, Arne, 222
Nagel, Ernest, 412
naming, 152, 198, 355
natural kinds, 3, 179
natural selection, 169, 176–181, 183, 184, 260–263, 274, 349, 415, 449, 453, 466, 475, 476, 479, 480, 486, 487, 493–495
naturalism/naturalistic, 2, 3, 112, 321, 361, 371, 395, 398, 403, 405–406, 419, 461–462, 464, 467, 468, 469, 471, 472,
necessarily, 29, 55, 57, 84, 100, 173, 191, 260, 278, 295, 355, 356, 381, 408, 444, 502
nerve, 184, 255, 320, 324, 328–329, 347, 349, 370–372, 405, 417, 454, 460, 463, 473, 474, 478, 485, 493, 494; irradiation pattern, 164; neural hookups, 165, 493; triggered receptors (extero receptors, nerve endings, neuroreceptors, sensory receptors), 177, 183–184, 370, 398–399, 410, 417–418, 463, 473–474, 478
Neurath, Otto, 410, 412
New England, 189
New Jersey (example), 154
Newton, Sir Isaac/Newtonian, 35, 74, 179, 228, 408, 414
Nile River, 405, 416
Nirvana, 337, 385, 435
Norway (venue), 508
notion. *See* concept
numerals, 21, 22, 38, 47, 49, 70, 72, 148, 293, 506; numerous, 21, 45, 47, 204, 300

Oberlin College, 390, 501–502
objective, 53, 65, 154, 166, 167, 177, 179, 180, 189, 216, 220, 231, 254, 267, 270, 275, 325, 334, 368, 377, 378, 381, 411, 415, 442, 451, 452, 453, 458, 460, 477, 478, 485, 496; objectivity, 220, 225, 330, 367, 465
objects of belief, 172, 441
Occam/William of Ockham, 15, 395, 400
Oklahoma, University of, 523
 ω -inconsistent, 238
ontological relativity, 213, 334, 375, 376, 458; “ontology recapitulates philology” 146, 182; inscrutability of reference, 361, 368, 374, 375, 376, 496
opacity, 253, 326
opaque, 253; transparent, 98, 242, 423

Oppenheim, Paul, 142
Orcutt, Bernard J., 173, 441–443
Osgood, Charles E., 165
Oslo (Norway) (venue), 172
ostension, 213, 231, 265, 325, 338, 353, 402, 411, 413, 452, 456, 457, 479

paradigm(s), 21, 27, 28, 163, 265, 276, 281, 316, 334, 365, 375, 452, 462, 500
paradox(es), 18, 28, 102, 152, 158, 160, 225, 242, 248, 281, 384, 402, 423, 424, 426, 427, 428, 440, 472, 503, 504, 505
Paris (France)/Parisian, 22, 50, 70
particulars, 9, 10, 12, 13, 15–19, 21, 22, 59, 79, 232, 259, 318, 421, 458, 482
passions, 72, 82, 113, 114, 116n, 118, 119, 122, 124–126, 130, 131
pattern of irradiation. *See* nerve: irradiation pattern
Peano, Giuseppe, 158, 200, 310, 388
Pegasus, 139, 140, 152, 188, 353, 354, 357, 358, 482, 499
Peirce, Charles Sanders, 142, 158, 472
Pennsylvania, University of, 24, 152
perceive, 2, 3, 69, 70, 72, 73, 96, 97, 104, 107, 325, 326, 327, 404, 440, 475
perception: and empathy, 172; shade of blue, 70, 80n; simple impression, 69, 71, 72, 79, 80, 82, 84, 100, 105, 107, 140
perspective(s), 36, 37, 103, 176, 177, 189, 275, 337, 384, 406, 436, 475, 476, 477, 478, 485
Phaedo (Plato), 57
phenomenalism, 409, 413, 418, 419, 464, 477, 491
phlogiston, 435
phoneme, 218, 223, 224, 287, 288, 340, 364, 365, 374, 441, 442
pidgin/Pidgin, 186–187
Pike, Kenneth L., 364, 366
Pilate, Pontius, 428
Pittsburgh, 391
planet, 21, 295, 327, 328, 339, 353–356, 361, 456
Plato/Platonist, 1, 11, 19–20, 49, 52–53, 56–58, 86, 145–147, 150, 158, 220, 390, 402n
Pluto, 430, 456
Poincaré, Henri, 159, 237, 241–242, 270, 337, 383–384, 434, 436

- Poland, 503
 Polyanna, 109
 Popper, Sir Karl, 332, 466
 Portuguese, 16, 125
 positivism, 9, 11, 12, 24, 88, 97, 113, 242, 405, 467
 Postal, Paul M., 216, 218, 227
 potential, 57, 85n, 162, 208, 438, 445
 pragmatism, 26, 27, 83, 135, 142, 158–162, 397, 430, 471
 Prague, 391, 504
 praise and blame, 108–110, 128, 133
 predicate logic. *See* quantification theory
 preestablished harmony, 108, 179, 464, 473–475, 485–487
 Premack, David, 246–247, 374
 Price, H. H., 39, 44–45
Principia Mathematica (Whitehead and Russell), 16, 387–388, 390–391, 501–503
 probability, 51, 52, 56, 95, 97, 142, 175, 217, 218, 445, 469
 Procrustean, 316
 promise, 132
 pronouns, 14, 15, 188, 212, 269, 316, 359, 373, 374, 421, 465; demonstrative, 352; essential, 452, 455, 483, 488; of laziness, 452, 455; relative, 267, 315, 359
 propositional attitudes, 172–174, 442–444, 505
 Protagoras/Protagorean, 179, 181
 proxy, 19, 189, 415, 458; functions, 189, 316–318, 321, 322, 339, 340, 345, 361–363, 375, 376, 404, 433, 434, 458, 496
 Prudential Insurance Company, 207
 pseudoproblem, 418n
 psychology, 2, 62, 95, 99, 103, 104, 109, 113, 118, 267, 286, 318, 328, 343, 348, 349, 360, 390, 395, 417, 462, 464, 474
 Putnam, Hilary, 278, 358

 qualities, primary/secondary, 63, 64, 65, 66, 97
 quanta, 10, 21, 22, 64
 quantification theory, 29, 30, 145, 149, 159, 498; predicate logic, 172, 358, 438, 439, 442; quantifying into modal contexts, 2; quantifying into opaque contexts, 253
 quantum, 175, 277, 282, 394, 445, 467
 quasi-Darwinian, 349
 quasi-existence, 393
 quasi-instances, 92
 quasi-repeated, 92
 Queen Mary (example), 227
 Quine, Alexander, xi, 507–508
 Quine, Douglas, 3, 7–8, 37–38, 49, 247
 Quine, Margaret, 247
 Quine, Willard Van Orman web site, 3, 7–8

 Rabbit (Volkswagen), 374
 Ramsey, F. P./Ramsey sentences, 235, 401–402, 405, 459, 470
 rational/rationality, 89n, 124, 127, 134, 153, 246, 413, 419
 realism, 1, 10, 60, 170, 179, 321, 346, 405, 491
 recursion, 219, 343, 427
 reduction, epistemological, 141
 referent, 200
 referential opacity. *See* opacity
 Reichenbach, Hans, 412
 Reid, Thomas, 107
 reinforce, 102, 133, 177, 180, 187, 244–246, 259, 271, 329, 341, 349, 351, 375, 376, 399, 419, 463, 478, 486, 495
 religion, 50, 108, 112, 113, 169, 337
 reminiscence, 57, 58, 504
 Riemannian, 434
 roundness, 11–14
 rules, general (Hume), 96, 129, 133, 309
 Russell, Sir Bertrand, 13, 15, 18, 20, 61, 83, 85, 103, 139, 153, 157–158, 200, 305, 383, 387–388, 390, 392, 424, 427–428, 430, 439, 470, 497, 499–502, 504; *Principia Mathematica*, 16, 387–388, 390–391, 501–503
 Ryle, Gilbert, 252, 255

 Saarnio, Uuno, 20
 Sargasso Sea of mentalism, 252
 schematic, 291, 312, 314, 394, 404, 491, 498, 504
 schematism, 382, 404, 454, 456, 487
 Scott, Sir Walter, 350, 361
 sea-fight tomorrow (Aristotle), 429
 self-consciousness. *See* conscious(ly)/consciousness/unconscious
 Sellars, Wilfrid, 252, 255
 semantical rules, 33, 154

- sense data, 102, 140, 155, 257, 263, 327–329, 369, 409, 413, 462, 464, 487, 491
- sensory receptors, 167, 170, 171, 189, 328, 329, 370, 403, 410, 416, 418, 463, 469, 493
- Shannon, Claude E., 204, 208
- Sheffer, Henry Maurice, 141, 391, 504
- Sherwin-Williams (paint), 372
- significance, 150, 152–155, 211, 241, 273, 369, 393, 443
- Sinn. *See* meaningful
- skeptic/sceptic, 2, 53, 58–59, 65, 67–68, 97–98, 108, 113, 135–136, 257–258
- skill, 111, 198
- Skinner, Burrhus Frederic, 61, 95, 259, 348–349
- Skolem-Löwenheim theorem, 282
- Sleigh, Robert, 174, 354, 443, 505
- Smart, J. J. C., 260
- Smith, Norman Kemp, 39–40, 46, 68, 73, 127–128
- social, 131–133, 155, 163, 245, 264, 271, 272, 274, 289, 301, 330, 366, 462, 467, 479
- Society of Fellows, 65, 95
- Sophistry, 53, 91
- Spanish, 125, 374
- spatial, 10, 85, 86, 224, 274, 454
- spatio-temporal, 10, 19, 232, 283
- St. Louis conference, 370
- Stanford, 163, 370–371, 385, 393, 474
- state(s) of affairs, 175, 383, 487
- Stetson, Raymond, 390
- Stevens, Smith, 222
- Stevenson, Charles Leslie, 143
- Strawson, Sir Peter F., 226, 339
- strict implication, 502
- structuralism, 401, 402, 404–406, 459; “save the structure and you save all,” 170, 405
- struthionism, 11, 12
- subjectivity, 11, 60, 65, 97, 102, 259, 260, 273
- substitutivity of identity, 253, 354, 356, 357, 440–444, 499
- Sukale, Michael, xi
- Sumerian(s), 295–296, 339
- surface 84, 85, 228, 252, 258, 328, 345, 357, 371, 372, 389, 392, 398, 406, 416, 418, 434, 463, 474; “save the surface and you save all,” 372
- sympathy, 117–120, 122, 123, 129–131, 133
- syncategorematic, 14, 15, 20, 21, 188, 496
- synonym, 12, 28, 154, 284, 292–296, 299, 502, 505, 506
- Tännsjö, Torbjörn, 174
- Tarski, Alfred/Tarskian, 8, 27, 152, 249, 314, 318, 321, 360, 375, 388, 392–393, 415, 422, 424–427, 459, 471–472, 504
- telepathy, 166, 177, 284, 328
- Tennant, Neil, 391
- theology, 112
- theory-laden, 407, 409, 412, 418, 476
- time. *See* spatio-temporal
- Toletus, Francisco, 62
- Tooke, John Horne, 61–63, 67, 82–83, 88, 106, 271–272
- Tooke’s doctrine of abbreviations, 61
- Toronto, 390, 393
- transcendent(al), 406, 461, 472
- translate, 12, 135, 164, 210, 237, 315, 337, 358, 359, 391, 427
- transparent. *See* opaque
- Treatise of Human Nature* (Hume), 37–38, 40–43, 45–46, 48–50, 53, 62, 68–80, 82–85, 89–90, 105, 108–110, 113, 126–127
- triggered receptors (extero receptors, nerve endings, neuroreceptors, sensory receptors). *See* nerve
- Tully (example), 172, 499–500, 505
- Tymoczko, Thomas, 335, 379
- Ullian, Joe, 398, 491
- under-determination, 228, 230, 232, 237–242, 269
- unicorn, 55, 91, 92, 188, 278
- universals, linguistic, 219, 220, 366
- Uppsala (venue), 509
- Urban, Wilbur Marshall, 12
- Urmensch*, 481
- utility, 125, 126, 132, 135, 149, 158, 160, 258, 281, 334, 365, 372, 379, 391, 402, 424, 439, 455, 460, 496
- vague, 73, 93, 153, 154, 179, 202, 242, 251, 276, 289, 300, 307, 321, 390, 396, 421, 428, 467, 469, 489
- vagueness, 75, 178, 212, 231, 242, 322, 388, 421, 490
- verification, 24, 153, 380, 383
- Vienna, 409

-
- Vienna, thinking about (example), 219, 305, 350
 Vienna Circle, 12, 32
 virtuous, 128–131
 vivacity, 68, 73, 80, 92, 118 (Hume)
 Volkswagen (Rabbit), 374
 Voltaire (François-Marie Arouet), 163
 Von Neumann, John, 213, 255, 470

 Wang, Hao, 190, 433
 Warsaw, 504
 Watson, John B., 244–245, 255, 390
Waverley (book example), 139–140, 287, 350, 361, 452
 Wedberg, Anders, 224
 Weyl, Hermann, 159, 190, 433
 Whatmough, Joshua, 364
 White, Morton (Morty), 393
 Whitehead, Alfred North, 25, 83n, 200, 387–388, 390–391, 439, 497, 501–504;
 Principia Mathematica, 16, 387–388, 390–391, 501–503
 Whorf, Benjamin Lee, 165
 Wiener, Norbert, 158, 349, 388–389, 503
 Wiesel, Torsten N., 329, 349, 417
 Williams, Donald Carey, 42–44, 155
 Wilson, Neil, 237
 Wittgenstein, Ludwig Josef Johann, 91, 99–100, 139, 142, 164, 169, 271, 273
 wvquine.org, 3, 7–8

 yellowness, 13

 Zanet, Giancarlo, xi, 24, 507, 508
 Zermelo, Ernst, 213, 255
 Zimmerman, Dean W., 8
Zuordnungsdefinitionen, 412