



The Role of Notation in Mathematics

by Edwin Coleman MA (ANU) MSc(Adelaide)

A thesis submitted in accordance
with the requirements of the
degree of Doctor of Philosophy

Department of Philosophy
University of Adelaide
South Australia
May 1988

Contents

Title Page

Contents

Synopsis

Summary

Statement

Acknowledgement

Part I

Notation and philosophy of mathematics

| | | |
|-----------|---|----|
| Chapter 1 | The nature, effects and importance of notation | 3 |
| Chapter 2 | Problems and positions in philosophy of mathematics | 20 |

Part II

Mathematical text, context and praxis

| | | |
|------------|------------|-----|
| Chapter 3 | Expression | 68 |
| Chapter 4 | Text | 91 |
| Chapter 5 | Writing | 126 |
| Chapter 6 | Diagrams | 169 |
| Chapter 7 | Notation | 200 |
| Chapter 8 | Paragraphy | 246 |
| Chapter 9 | Context | 270 |
| Chapter 10 | Praxis | 298 |

Part III

A conception of mathematics

| | | |
|------------|------------------------|-----|
| Chapter 11 | Making mathematics | 332 |
| Chapter 12 | Mathematical knowledge | 378 |

Bibliography

Exhibits

(In pocket at back)

Synopsis

1

§1 Introduction

§2 Philosophical problems about mathematics

§3 The nature, effects and importance of notation in mathematics

§4 Notation allows generality

§5 Notation allows us to discuss what is too abstract, remote from experience

§6 Notation permits a useful polysemy

§7 Notation gives ease, makes things practical, overcomes our limitations

§8 Notation makes calculation possible

§9 Notation directly expresses the things

§10 Notation permits precision and certainty of inference

§11 Notation is suggestive, carries you along, thinks for you

2

§12 Notation and philosophy of mathematics

§13 Seven problems of philosophy of mathematics

§14 The mathematisation of logic

§15 The foundations problem

§16 Logicism, Intuitionism, Formalism

§17 The impact of Gödel's theorems

§18 The logico-formalist hegemony

§19 Will nothing work?

§20 Conceptions of the role of notation in mathematics

§21 Intentions and methods of the investigation to follow

§22 Recapitulation of the problem and summary of the argument to follow.

3

§23 Bleak Intuitionism as a view about notation

§24 Four Intuitionist theses about constructions, language and logic

§25 Signs for intuitions are required in mathematics

§26 Signs for defined terms are necessary in mathematics

§27 Signs for theorems are necessary in mathematics

§28 Signs for proofs are needed in mathematics

§29 Some difficulties raised by Dummett

§30 A doubtful argument

4

§31 Examination of texts

§32 Any mathematical text has a context of use

§33 The signsystems of mathematics

§34 The Words tell the tale

§35 The Diagram shows the idea being exemplified

§36 Notation makes the inferences evident

§37 Paragraphy ties the text together

§38 What this text does is illustrate a technique.

§39 Local and global eliminability of signs

§40 What this example indicates about mathematical text

§41 What the logico-formalist hegemony says about the sample text

§42 Goodman's concepts of notation, character and score

§43 Goodman on diagrams and icons, and language

§44 Goodman's forgery puzzle and his solution of it

§45 Difficulties with this solution to the puzzle

§46 The error underlying Goodman's conception of symbolsystem

§47 Text in mathematical didactics

5

§48 Role of speech in mathematics

§49 Antecedents and development of writing

§50 Characterisation of writing

§51 The cognitive significance of writing

§52 Speech Acts

§53 Hard cases for speech act theory

§54 Limitations of speech act theory

§55 Text acts

- §56 Text act, text and discourse
- §57 The points of text acts
- §58 Felicity conditions for text acts
- §59 Text effects
- §60 Definition as a speech act and as a text act
- 6
- §61 Kinds of diagram
- §62 The formal elements of diagrams
- §63 Diagrams are formally articulated by juxtaposition, alignment, spacing incidence and measure.
- §64 Diagrams as text acts
- §65 A diagram is a grounded iconic graph
- §66 The semantics of diagrams
- §67 Tables, lists and forms are diagrams
- §68 The logicist attempt to extirpate diagrams from mathematics
- §69 Why mathematics needs diagrams
- §70 Tu quoque
- 7
- §71 What notation comprises
- §72 An example described in logico-formalist manner
- §73 Some omissions from this account
- §74 Notation abbreviates
- §75 Notation is formal
- §76 Notation is iconic
- §77 Notation must be grounded in Words
- §78 Begriffsschrift as perfect logical language
- §79 Begriffsschrift as notation
- §80 Analysis, replacement and avoiding ontological commitments
- §81 Throwing away the ladder ?
- §82 Does regimentation reveal ontological commitment?
- §83 Surveyability
- 8
- §84 Nature of the Paragraphy signsystem
- §85 What Paragraphy comprises
- §86 Formal elements of Paragraphy
- §87 Functional elements of Paragraphy
- §88 Elementary paragraphic text acts
- §89Text articulation
- §90Text grounding, discourse and Archive
- §91Textual objectification
- 9
- §92 Mathematical text and context
- §93 The trace of praxis
- §94 The Archive is presupposed by all mathematical texts
- §95 Contents of the mathematical Archive
- §96 The material environment of mathematics is largely mathematised
- §97 Contents of the mathematised world
- §98 Diagrams are constitutive of the mathematised world
- §99 The third facet of the context of mathematical text is a readership
- §100 Readership is constituted through didactic text and social stratification
- §101The mathematical context is product and precondition of mathematics.
- §102 How context affects the understanding of mathematical text.
- §103The myth of the common content (again)
- 10
- §104 The concept of mathematical praxis
- §105 Lakatos on proofs and refutations
- § 106 Lakatos on notation
- §107 Some limitations of Lakatos' position
- §108 Wilder's cultural evolutionism
- §109 Kitcher on mathematical practice.
- §110 Kitcher's conventional semiotics
- §111 Kitcher's platonist mathematical ontology
- §112 Retreat from pragmatics

- §113 Consolidated criticism of these writers
- §114 Mathematical activity is work
- §115 Mathematical work is textual
- §116 Mathematical work is phenomenologically misleading
- §117 What mathematics comprises
- 11
- §118 Creation and application in praxis
- §119 Wittgenstein's later philosophy of mathematics
- §120 Combination of the two
- §121 Mathematical invention
- §122 How Conway found a new simple group
- §123 Novelty, invention and application in mathematics
- §124 Two accounts of application
- §125 The sources of vagueness and precision
- §126 The logic of vagueness a red herring
- §127 Kyburg on measurement
- §128 Measurement's pragmatic presuppositions
- §129 Wittgenstein's conventionalism: following rules
- §130 Mathematical versus non-mathematical sentences
- §131 The programmed formula challenge
- §132 Alternative mathematics: is it comprehensible?
- §133 Craig's challenge
- §134 The spectrum of application
- §135 Making mathematics and mathematical praxis
- 12
- §136 Problems remaining
- §137 Objects
- §138 The kind of thing known only through signs
- §139 Infinity as a text effect
- §140 Mathematical knowledge
- §141 Mathematical knowledge as text effects of proof
- §142 A proof achieves bivalent textual truths by deploying notation.
- §143 Proofs have their power through being surveyable.
- §144 Surveyability comes from written expression
- §145 A proof has a place in an archive, presupposing P-signs
- §146 A proof is a construction using the iconicity of the diagram system
- §147 Hilbert's Epistemology of Mathematics
- §148 Observing pawns and surveying signs
- §149 The logicist muddle about definition
- §150 Definition in mathematics
- §151 What do definitions and notations achieve?
- §152 Semiotic ascent
- §153 Concluding remarks

Summary

I propose a conception of mathematics derived from an investigation of the role of notation in mathematics.

In Part I, I introduce the question of the role of notation in mathematics. I identify seven problems standardly discussed as the philosophy of mathematics, and show how what I call the *logico-formalist hegemony*, a widely accepted conception of signification in mathematics, underlies the general unsatisfactoriness of most extant philosophies of mathematics. I identify two main conceptions of the role of notation in mathematics which support it, *Dispensability* and *Replacement* conceptions.

In Part II, I argue against these conceptions and in favour of what I call an *Extension* conception instead. I do this by analysing the signsystem of mathematics, which I argue to be specific in two ways. First, it is a *written* system, and second, it consists of four relatively independent subsystems, those of *Words*, *Diagrams*, *Notation* and *Paragraphy*. I show that these four subsystems play mutually irreducible roles in mathematics, and that weaknesses in extant conceptions of mathematics derive from conflating them. I argue against prevailing conceptions of writing as mere transcription: the writtleness of mathematics permits the constitution of mathematics as an accumulative tradition relying on a context of *archives*, *a mathematised world* and *a readership*. Thus I develop an account of mathematical *texts* and *context*, and in the course of this development I argue that a number of widespread conceptions are flawed by mistaking the role of notation in mathematics - the Intuitionist conception of mathematical activity, the Formalist conception of evidence, the Logicist conception of definition, the Logicist thesis that diagrams are not needed in mathematics, the Positivist idea that a notation could be a complete language for science, and the Holist argument for platonism in mathematics. I criticise the conceptions of mathematical practice of Lakatos, Kitcher, and Wilder. I contend that although these give some recognition to context by considering mathematical processes, it is not enough. I suggest a notion of *mathematical praxis* based on *textual work*.

In Part III, using the preceding analysis of mathematical texts, context and praxis, I expound a coherent conception of *mathematics as a system of text effects*. I defend this conception by interpreting mathematical proof, truth, objects, knowledge, application and creation, as aspects of textual praxis.

Statement

This thesis contains no material which has been accepted for the award of any other degree or diploma and to the best of my knowledge and belief contains no material previously published or written by another person except where due reference is made.

I give permission for the thesis, if accepted, to be loaned or photocopied.

Edwin Ronald Coleman

Acknowledgement

The research for this thesis was carried out under the supervision of Dr Chris Mortensen and I wish to thank him for the constructive criticism and enthusiastic discussion which he lavished on its execution.

I thank Michael Rowan particularly, and other philosophical colleagues for many helpful comments.

I thank Helen Measday for many helpful comments and much valued support.

I thank my children for distracting me often enough.

Part I

Notation and philosophy of mathematics

Chapter 1

The nature, effects and importance of notation



§1 Introduction

The aim of this thesis is to present a conception of mathematics formed through an investigation of the role played by notation in mathematics, an investigation which considers all the aspects of signification which are actually to be found in real mathematics.

In the present Part I, I will present the issues which the thesis addresses, relating questions about the role of notation in mathematics to problems in the philosophy of mathematics. In the following Part II, I will describe the various elements of mathematical signification. I emphasise the role of writing, of the several mutually irreducible signsystems employed in mathematics of which notation is one, and the nature of the context of mathematical work. In the concluding Part III, I will give an integrated account of the consequent conception of mathematics. I argue that a number of the "problems" of philosophy of mathematics are pseudo-problems, but I recognise mathematics as a distinctive form of knowledge, and I account for its distinctiveness in terms of the analysis of texts and contexts developed in Part II. My conception is hyperformalist and conventionalist, but also materialist: what this means will become clear later.

In the following discussion I shall frequently refer to one or another of the examples of mathematics which are collected together in the pocket at the back of this thesis as *the exhibits*. These are all samples of actual mathematical texts. I will describe them at various places in the sequel. There is a complete list appended to them, which includes references to the numbers of the sections where they are first discussed.

§2 Philosophical problems about mathematics

There are many philosophical problems discussed in the name of philosophy of mathematics, but we shall make contact with most of them if we examine one of the exhibits. Consider exhibit #1. This table forms part of the financial records of a small organisation, and it records the wages paid for the last

working fortnight of 1987. I have chosen to begin the discussion with this slightly unusual example for reasons which will be made clear later. More traditional examples can be found in the exhibits too, such as the passages from Euclid, Hardy and other pure mathematicians (exhibits ##5,6 etc). Most, though not quite all, of the following discussion applies to any of them. While I do not propose to discuss this example in great detail here¹, consider for the present just the first line where we find in successive boxes the numbers 14, 8.5 and 119. These record the hours worked, hourly wage rate (in \$) and total wage (in \$) for the first listed employee, EB, in the period in question. The third figure is obtained by multiplying² together the first two: $14 * 8.5 = 119$. I have here made explicit a mathematical fact which was used in the construction of the exhibit, but is not explicit there. We can ask a number of puzzling questions about this fact. We seem to know for certain that $14 * 8.5$ could be nothing but 119. This kind of knowledge seems special. But why is ' $14 * 8.5 = 119$ ' certain, and what is it about, how can it be useful? These general and vague questions need to be specified into more numerous and precise ones. Unfortunately the result is a great proliferation of "problems", and of suggested answers to them, none of which has been terribly satisfactory.

Three aspects of mathematical knowledge which can all seem mysterious are the kind of knowledge that we have in mathematics, the things we know of in such knowledge, and what can be done with it. Consider the peculiarities of knowing that $14 * 8.5 = 119$. It is different from the great bulk of our knowledge in its exactness, its certainty and its absoluteness. ' $14 * 8.5 = 119$ ' holds exactly, but 'France is hexagonal' is only true up to a point. ' $14 * 8.5 = 119$ ' could not be false, but France might easily have been square, and indeed may become so. ' $14 * 8.5 = 119$ ' holds here and now and in China 5000 years ago, for me for you and for Confucius, but 'France is hexagonal' depends for its truth on when and where it is known. The differences

¹ But see also §12 and Chapter 11.

² Here and in later chapters I use '*' for multiplication, not the traditional '.' or 'x'. Both of these have ambiguities; '*' is a notational innovation consequent on computerisation, and a good one.

seem to be related to the use of proofs in mathematics, since it can be conclusively demonstrated that 14×8.5 is 119, but it is not clear why this is so or why it makes such a difference. Accounting for these differences and melding the two kinds of knowledge into a coherent account may be summed up as a question about mathematical knowing: why is it peculiarly secure?

The second aspect of the peculiarity of mathematical knowledge can be put like this: if we consider mathematics as presented in texts like the exhibits, we find that it is concerned with numbers, circles and other *objects* which do not seem to be in the world of flesh and blood. Yet in the examples, knowledge of these things is brought to bear on that world. How can we, as flesh and blood creatures, get to know about these unwordly entities, and how can that knowledge be at all relevant to the real world? How can knowledge about numbers be of any relevance to the world of numbness and numbats? This is to ask questions about the mathematical knowns: what sort of thing are mathematical objects? How does their manner of being influence our knowledge of them, how is it possible for us to obtain acquaintance with them, and how are they related to everyday items, or to those which science seems to tell us compose sticks and stones? How is a mathematical "fact" like any ordinary one (however that is)?

A feature of exhibit #1, which some of the others lack, is that it involves the application of mathematics. This can seem very mysterious, and the mystery is deepened if we consider some of the more amazing events achieved using mathematics, for example landing on the moon or the construction of cameras. In doing so we face a third group of philosophical problems: how is it that ("pure") mathematics can be applied to the world? How can all these "chicken scratches" (as some mathematicians self-mockingly call their signs) play any useful role in facilitating such courses of action? How can it be that in many cases, as I shall argue in detail later, the events and items in question could not be created or happen without them?

§3 The nature, effects and importance of notation in mathematics

If mathematics has cognitive peculiarities, as the puzzles just sketched indicate, this must have some source. This thesis is an investigation of the idea that notation is the source. It begins from some simple questions. ***What is the role of notation in mathematics? Does notation have a necessary role in mathematics? Is there mathematics worth the name which does not rely on notation?*** And if, as I shall argue, the answers to these questions are yes, yes and no respectively, then there are further questions. ***What is the specific role which notation plays in mathematics which cannot be played by words, or diagrams or some other element? What is the peculiarity of mathematics as a human activity that it requires this tool? Are there others, for example music, which have a similar reliance, and if so what is it they have in common? How is that mathematics can have originated in cultures without notation, if it is really necessary?***

Before giving some reasons for taking these questions seriously, let me indicate in a preliminary way what I understand by notation. A glance at the exhibits will confirm that mathematical text employs signs of the following kinds.

| | |
|--|-----------------------|
| letter combinations which are not words, like | $x f Z SL_2$ |
| specially introduced "mathematical" signs, like | $= + - \Sigma$ |
| other signs, like brackets, underlines and affixes | $) \underline{y} x^y$ |

The boundary between special signs and ordinary written words is fluid and vague, and the functions which are divided in mathematical text between word and notation are therefore similarly mobile. In Chapter 7, I will consider what notation is in more detail.

There are several prima facie reasons for thinking notation to be essential for mathematics. First, folk wisdom would have it that mathematics is all about arcane formulas, that it essentially depends somehow on the use of "symbols", in some sense of that rather overused word. Common antipathy for mathematics identifies it through its reliance on notation - meaningless gobbledygook to most

people. Second, mathematical sophistication sometimes has a similar idea, though less plainly and less officially. Many writers have commented on the crucial role of well-chosen notation in mathematics. Cajori³ gives scores of remarks by mathematicians commenting on the relative advantage of one notation against another, and the profits which flowed from their adoption. It is not difficult to find eminent mathematicians prepared to rate very highly the value of a good symbolism. Leibniz and Peano are well-known examples. It has been seriously suggested⁴ that British mathematics contributed nothing to analysis in the century and a half after Newton because it insisted on clinging to Newton's inferior notation for the calculus, whereas on the Continent they went ahead in leaps and bounds using that of Leibniz. Doubtless this is a gross oversimplification of a complex historical fact, but its popularity is some evidence for the idea that notation has a crucial importance all the same. Third, it is in any case obvious that mathematics as normally understood *does* make use of notation to a greater or lesser extent, whether or not it *must*. Mathematics as found in practice is heavily reliant on notation. The exhibits make this obvious. So we can certainly ask whether it must make such use, and if so why? Or if not, why is it used in fact and what does this show us about mathematical and other knowledge? These considerations seem sufficient to make the questions worth investigation.

Another reason for considering the nature and role of notation in mathematics is mathematical logic. As I shall argue in Chapter 2, a certain set of assumptions about language and mathematics have become general this century in association with the logico-formalist characterisation of mathematics, which not only consists in the development of a notation generalised from that of arithmetic, but also identifies mathematical discourse with formal theories in "Fregean" languages. This is an idea to which I shall revert constantly below.

³ Cf Cajori [1929]. For example, Laplace on p70, Cardano p142, Stiefel p144, Oughtred p199, Wallis p218ff, and many others *passim*.

⁴ Cf the debate about "dotage" (Newton's notation) referred to in Cajori [1929: 211ff].

The importance of notation in mathematics is often acknowledged, but it is rarely explained. I shall argue in this thesis that the nature of notation and its role in mathematics can only be understood when it is seen as an element in the mathematical sign system as a whole, that this system is a writing system and so therefore is notation, and that an appreciation of why and how these are so will illuminate the traditional questions in philosophy of mathematics which I will outline in Chapter 2. Parts II and III will be devoted to these tasks. In this chapter, I will classify some opinions as to the importance and role of notation which can be found in the literature of mathematics. At least eight **effects** of notation are remarked upon by philosophical and mathematical writers. I only identify these here; they will be evaluated in Chapter 7. In the following eight sections I gather together some illustrative quotations about the importance of notation, extracting the main ideas which seem to have been put forward to account for that importance. These quotations all occur in the manner of asides, for there is no single work which I can find devoted specifically to the questions which I have raised⁵. But there are a number of common themes to be found in these statements, and many more which no doubt could be assembled, for they seem to cover most of the common ideas about the value of notation. It will be evident from these representative samples that little effort has been devoted to careful analysis of the different aspects and effects of notation.

§4 Notation allows generality

Jourdain⁶ expresses an aspect of the *economic* aspect of notation:

"...what is a great advance in economy of thought and labour, a part of calculation could be done with symbols instead of numbers, so that the one result stated, in a manner analogous to Greek geometry, a proposition valid for a whole infinite class of different numbers."

⁵ Cajori's fascinating two-volume work Cajori [1929] is basically a history like Menninger's [1969], and although it does attempt to draw some morals, they seem to be mainly concerned with Cajori's desire for uniformity of usage, rather than any real attempt to divine the nature, effects and role of notation in mathematics.

⁶ P Jourdain, *The Nature of Mathematics*, in Newman [1956: I, 61]

By itself, this is not very helpful: generality is made possible in many ways in language, but we are interested in what is specific to notation. On the other hand it is certainly the case that many advances in mathematics have come from an increased generality, and that notational innovations underlie such advances. The rise of abstract algebra is a case in point. But it is not generality alone which is significant since extreme generality is easily available in words: 'everything is what it is and not another thing' - true, even perhaps profound, and completely general. The effect at which Jourdain is perhaps hinting is a particular kind of generality, a general applicability of the calculus within which mathematical statements have their place, perhaps like the point made by Frege and Putnam among others, that numbers can be used to count *anything*.⁷

However, we must not dismiss the idea of economy altogether; it is certainly true that notation makes possible economies of thought and action. For example, the simple multiplication of exhibit#1, $14 \times 8.5 = 119$, which takes a few seconds in decimal notation, would take minutes in successor arithmetic⁸; moreover if it were done in cents and minutes, rather than dollars and half-hours, it would take thirty times as long again. This might be dismissed as "in principle" irrelevant, but in practice it is not. Calculations which cannot be completed in time to be useful are not made and cannot become part of a social institution like employment. If all calculation were as onerous as this, our culture could never have become a mathematical one and we could therefore have no mathematical knowledge to speak of.

⁷ But is it true? Can we count fundamental particles, or the sea, or feelings of unease? There are difficulties with many such projects; counting seems to me no more universally applicable than color vocabulary. (I use American spellings for words like 'color'.)

⁸That is, in the primitive notation of stroke symbols which the Peano axioms concern before definitions are introduced.

§5 Notation allows us to discuss what is too abstract, remote from experience

Russell and Whitehead devote several paragraphs in the preface to *Principia Mathematica* to justifying their use of notation. Many different ideas are touched on in this passage⁹:

"The symbolic form of the work has been forced on us by necessity: without its help we should have been unable to perform the requisite reasoning... [It use] was forced upon us by the consistent pursuit of the above three purposes [of a minimal basis, perfectly precise expression and avoiding paradoxes]...

Any use of words would require unnatural limitations on their ordinary meanings...

the very abstract simplicity of the ideas in this work defeats language...

The adaptation of the rules of the symbolism to the processes of deduction aids the intuition in regions too abstract for the imagination readily to present to the mind the true relation between the ideas employed...

The terseness of the symbolism enables a whole proposition to be presented to the eyesight as one whole..."

Several ideas are mingled here, as in most of my specimen quotations, which I am trying to disentangle. The appeal to notation to cope with "remote" experience seems to me to be a very obscure idea, as though words are somehow limited to the expression of familiar experience, or as though working on higher mathematics is not a kind of experience. But there is at least the following sense in which remoteness enters formalised mathematics: conclusions can be deductively very remote from the premisses ultimately relied on. Inferences of such length as are routinely required in mathematics seem to be most unreliable in the vernacular, and the thought is natural that notation contributes to the contrasting reliability in mathematics. But it cannot be as simple as that, for the paradigm of such inferential structures, Euclid's *Elements*, while not quite perfect, does succeed very well with little notation, and Hilbert's *Foundations of Geometry*, correcting Euclid, is also highly verbal. A related idea which seems to be present in the quotation is this. In much modern mathematics one is dealing with objects which cannot be visualised (or not in any simple fashion). "imagination [cannot] readily present to the mind the true relations between the ideas employed.." Higher dimensional knot theory is an example here (who can visualise six-dimensional

⁹ Whitehead and Russell [1962: vii and 1-3]

knots embedded in eight-dimensional space?¹⁰), as are functional analysis and large cardinal theory. If one is reasoning about things which cannot be visualised then extra care may be needed and notation plays a role in securing it. Logicism seems to make a virtue of necessity here, and goes on to deny any essential role for any other faculties, under the boo term "intuition". I shall argue in Chapters 6 and 7 that not only is this a nonsequitur, its conclusion is false, since notation *and* words are needed in proofs.

§6 Notation permits a useful polysemy

Peano says about the importance of symbols in mathematics¹¹

"The first advantage that one sees in the digits is their brevity....[but] ...these symbols are not just shorthand, i.e., abbreviations of ordinary language, but correspond to a new classification of ideas...The use of digits not only makes our expressions shorter, but makes arithmetical calculation essentially easier, and hence makes certain tasks possible, and certain results obtainable, which could not otherwise be the case in practice....algebraic symbols represent ideas and not words... The symbol '+' allows us to represent that which in ordinary language is expressed by 'plus, sum' and even 'addition, term, polynomial'...Algebraic symbols are much less numerous than the words they allow us to represent....[but 'geometrical' symbols] are merely shorthand symbols and do not lend themselves to any calculus.... The first advantage of the symbols of logic is the brevity they produce.. [they] represent ideas...[they] make reasoning easier..."

These remarks of Peano's include an expression of the economy obtained through the use of notation, as well as several other ideas to which I shall return below. The point about using a symbol to stand for various words ("polysemy") is perhaps a way of adverting to a crucial property of notation which is not really remarked in the other quotations given so far, which is that it can be *empty* or uninterpreted. This formalist idea is not always recognised as a valuable property of notation, though it is the basis of the metamathematical attempt to give mathematics foundations. It may also be the basis of the kind of generality on which I was commenting in §4. It is ironic that there is so little comment on this prevalent ambiguity in the notations chosen in mathematics - how many uses has

¹⁰ Non-mathematicians often think that mathematicians are people who do this kind of thing, but it is not true.

¹¹ Peano [1915: 227-31]

the sign '+' in current mathematics? - considering the extensive fulminations of the founding fathers of logistic about the ambiguities of ordinary language! There is a difference, I admit, which is that notations are explicitly introduced so that any ambiguity *can* be resolved by context. But to see why this resolution is possible is rather a complex task. It requires us to observe that explicit introduction and resolution by context are possible in virtue of the archiving of written discourse. To show this requires in turn that we take seriously what is a crucial but almost entirely unremarked fact about notation: it is written. I shall discuss it at length below from Chapter 4 onwards.

§7 Notation gives ease, makes things practical, overcomes our limitations

This is the most remarked feature of notation, the preceding quotations all advert to it. Turnbull¹² says

"The study of notation is interesting, and covers a wider sphere than at first sight might be supposed...A good notation is a valuable tool; it brings its own fitness and suggestiveness, it is easy to recognise and is comfortable to use. Given this tool and the material to work on, advance may be expected. In their own language and in their geometrical notation the Greeks were well favoured; and a due succession of triumphs followed. But their arithmetic and algebra only advanced in spite of an unfortunate notation. For the Greeks were hampered by their use of letters a, b, c for the numbers 1, 2, 3 and this concealed from them the flexibility of ordinary arithmetic calculations. On the other hand, the very excellence of our decimal notation made these operations well-nigh trivial....."

(So letters don't automatically confer generality!)

Similarly Bell¹³ says

"The importance of an easily manipulated symbolism, as implied by De Morgan, is that it enables those who are not great mathematicians in their generation to do without effort mathematics which would have baffled the greatest of their predecessors....Unless elementary algebra had become "a purely symbolical science" by the end of the sixteenth century, it seems unlikely that analytic geometry, the differential and integral calculus, the theory of probability, the theory of numbers and dynamics could have taken root and flourished as they did in the seventeenth century...it may not be too much to claim that the perfection of algebraic symbolism was a major contributor to the unprecedented speed with which mathematics developed after the publication of Descartes' geometry in 1637.....almost anyone will agree that lack of an appropriate symbolism constrained the Greek arithmeticians and algebraists to consider special cases of what might have been

¹² Turnbull, in *The Great Mathematicians*, reprinted in Newman [1956: I, 114]

¹³ Bell [1945:128-130]

their problems, and prevented the Indians and the Moslems from producing an elementary algebra within the capacities of ordinary adolescents."

Indeed, everyone comments on the utility of notation in general terms such as these. The tenor of many such comments is that notation makes things easier which we could do anyway with more trouble. In these last four sections I have identified some commonly noticed effects of notation which focus on its abbreviatory nature and its consequent economy of thought and labour, its making old things easier to do. I have given them a section each because the literature, as these examples show, does not pay much attention to the distinctions among these ideas.

The more interesting effects of notation which I now go on to identify are those which suggest that notation makes *new* things possible. Recall that we are interested in what the specific role of notation may be in securing the peculiarity of mathematical knowledge. It must have irreducible properties if there really is such a specific role and peculiarity.

§8 Notation makes calculation possible

Peano¹⁴ notices a point which I shall argue in Chapter 7 to be a crucial feature of notation:

"The use of digits not only makes our expressions shorter, but makes arithmetical calculation essentially easier, and hence makes certain tasks possible, and certain results obtainable, which could not otherwise be the case in practice....algebraic symbols represent ideas and not words... [but 'geometrical' symbols] are merely shorthand symbols and do not lend themselves to any calculus."

Some signs are apt for a calculus, others are not. Perhaps many of the remarks about the ease which notation affords are actually motivated by this effect. For example, when Bell specified in the earlier quotation that symbolism be easily manipulable, perhaps he means that the increased powers conferred by notation

¹⁴ Peano [1915: 228, 230]

are as much a result of the manipulability of the notation as of this manipulation being easy.

The point is best made by example perhaps. If I say that "six dozen less six brace is three score", no doubt the reader will verify the correctness of this claim after a few seconds thought. What kind of thought? Well, almost certainly, re-expression of the number terms into more usual ones like 'seventy-two' and then of them to digital form like '72'. It is obvious to the eye, and therefore to the practised inner eye, that $72 - 12 = 60$. This obviousness is simply not there in the spoken form from which I began. Putting the point more abstractly, for numbers expressed in digital notation we have perspicuous algorithms for addition and other arithmetical tasks. For numbers expressed in the vernacular this is simply not the case. Decimal notation is apt for a calculus, number words are not. The basis of this distinction is twofold, the visible form of decimal notation and the place value (recursive) formation system. I will discuss the *perspicuity of inference* to which this leads later, in Chapters 7 and 12.

§9 Notation directly expresses the things

Russell and Whitehead say that notation "directly expresses" things, Peano that it "reclassifies ideas" (cf the quotations in §§5 and 6). These are rather obscure suggestions; a clearer formulation was given by Leibniz¹⁵:

"In signs one observes an advantage in discovery which is greatest when they express the exact nature of a thing briefly and, as it were, picture it; then indeed the labor of thought is wonderfully diminished".

I shall take this up in chapter 7 as the idea that notation is **iconic**. By this I mean that the logico-formalist semantics is a precise explication of the "picture theory of meaning" in which signs for objects and their relations stand in isomorphic projection for the objects and relations represented. Iconic expressions of facts are transparent in this way: if a statement is about three individuals, there will be three individual terms in its expression; and so on. Commonly this is not

¹⁵ Quoted by Cajori [1929: II,184]

true for statements in ordinary words, and it is here that notation has an advantage over them. Insofar as it is true for systems of notation (and formal systems are designed and interpreted so that it will be) the advantages accrue which Leibniz mentions. The utility of this procedure for science needs some explanation of course, which I shall offer in Chapter 11. There is[^] rather large metaphysical claim about how things really are which cannot be allowed to remain implicit in such remarks about notation.

§10 Notation permits precision and certainty of inference

An idea motivating Frege (though much older) is that notation can improve on language in the expression of inferences. For example Oughtred says¹⁶

"...this specious and symbolically manner, neither racketh the memory with multiplicity of words, nor chargeth the phantasie with comparing and laying things together; but plainly presenteth to the eye the whole course and processe of every operation and argumentation." {sic}

Repeating the quotation from Russell and Whitehead¹⁷

"The symbolic form of the work has been forced on us by necessity: without its help we should have been unable to perform the requisite reasoning... [It use] was forced upon us by the consistent pursuit of the above three purposes [of a minimal basis, perfectly precise expression and to avoid paradoxes]... Any use of words would require unnatural limitations on their ordinary meanings... the very abstract simplicity of the ideas in this work defeats language... The adaptation of the rules of the symbolism to the processes of deduction aids the intuition in regions too abstract for the imagination readily to present to the mind the true relation between the ideas employed... The terseness of the symbolism enables a whole proposition to be presented to the eyesight as one whole..."

The idea here is that notation can escape the uncertainty of expression and inference in ordinary language. Notice that Whitehead and Russell allow the intuition (and the imagination) a role in following reasoning, though in more forthright moments logicians deny that mathematics is in any way synthetic; and that Oughtred mentions the presentation to the *eye* of the argument in notation. I readily grant that there are inferences in notation which are superior in point of

¹⁶ Quoted by Cajori [1929: I, 199]

¹⁷ Whitehead and Russell [1962: vii and 1-3]

precision and certainty to most of those in speech. But I shall argue below that this does not mean that the former class can wholly replace the latter, even for the purposes of mathematics. To see why, it is necessary to see *why* notational inferences can be especially secure. Once again, this brings us back to the written nature of notation and the visual perspicuity of mathematical proofs.

§11 Notation is suggestive, carries you along, thinks for you

A number of writers have made remarks about the way in which notation can be assimilated into the mind:

Ernst Mach¹⁸

"The student of mathematics often finds it hard to throw off the uncomfortable feeling that his science, in the person of his pencil, surpasses him in intelligence"

Gottlob Frege¹⁹

"Symbols have the same importance for thought that discovering how to use the wind to sail against the wind has for navigation. Thus, let no one despise symbols! A great deal depends upon choosing them properly....without symbols we would scarcely lift ourselves to conceptual thinking...."

Whitehead²⁰

"By relieving the brain of all unnecessary work, a good notation sets it free to concentrate on more advanced problems, and in effect increases the mental power of the race... By the aid of symbolism, we can make transitions in reasoning almost mechanically by the eye, which otherwise would call into play the higher faculties of the brain...Civilisation advances by extending the number of important operations which we can perform without thinking about them."

Jacob Klein²¹

"The creation of a formal mathematical language was of decisive significance for the constitution of modern mathematical physics. If the mathematical presentation is regarded as a mere device, preferred only because the insights of natural science can be expressed by "symbols" in the simplest and most exact manner possible, the meaning of the symbolism as well as the special methods of

¹⁸ Ernst Mach quoted by Nagel [1979:171]

¹⁹ Frege, in Bynum [1972: 84]

²⁰ Whitehead [1911: 59]

²¹ From Klein [1968: 1-2]

the physical disciplines in general will be misunderstood. True, in the seventeenth and eighteenth century it was still possible to express and communicate discoveries concerning the "natural" relations of objects in non-mathematical terms, yet even then - or, rather, particularly then - it was precisely the mathematical form, the mos geometricus, which secured their dependability and trustworthiness. After three centuries of intensive development, it has finally become impossible to separate the content of mathematical physics from its form. The fact that elementary presentations of physical science which are to a certain degree nonmathematical and appear quite free of presuppositions in their fundamental concepts (having recourse, throughout, to immediate "intuition") are still in vogue should not deceive us about the fact that it is impossible, and has always been impossible, to grasp the meaning of what we nowadays call physics independently of its mathematical form. Thence arise the insurmountable difficulties in which discussions of modern physical theories become entangled as soon as physicist or nonphysicists attempt to disregard the mathematical apparatus and to present the results of scientific research in popular form. The intimate connection of the formal mathematical language with the content of mathematical physics stems from the special kind of conceptualization which is a concomitant of modern science and which was of fundamental importance in its formation...

.....The creation of the formal language of mathematics is identical with the foundation of modern algebra."

This effect of notation is, I believe, the most profound to be remarked on in the literature. The previous three effects - making calculation possible, directly expressing things, making precise inferences possible - are three important ways in which notation underpins cognitive advance. Such statements as these quotations from Mach, Frege, Whitehead and Klein reflect, albeit somewhat obscurely, an appreciation of the fact that our minds and our concepts are not given but developed over time, and that the construction of notation is an important element in *improving* our mental equipment. Such a point of view is quite contrary to the re-expressionism of the logico-formalist hegemony which I shall describe in the next chapter, in which we already have all the ideas but must struggle to find improved expression to replace the vernacular.

These last eight sections have identified some effects of notation which will all be discussed further in what follows. For the present it will suffice to group the ideas involved as three aspects of notation: its abbreviatory, economic, facilitating role, including its power of generality and abstraction; its iconic, picturing, direct expression of things making possible calculation; its constituting novel conceptual elements in our cognitive achievements which I shall argue to underlie its

improvement of our inferences. Views of the role of notation in mathematics, such as they are, tend to concentrate on the first or second of these three facets²². In this work I shall be emphasising the third, which is less generally accepted, and trying to fit the other two together with it. In the next chapter I will examine the literature of the philosophy of mathematics with a view to finding some account of the role of notation in mathematics. In fact, as I shall show, such an account is largely lacking except insofar as the development of mathematical logic provides one implicitly.

²² Historically, early individual notations tend to serve mostly the first facets, later systems of notation, the second. Some writers do not seem to realise that introducing a formal system is just as much a matter of defining notation as are explicit definitions of individual signs.

Chapter 2

Problems and positions in philosophy of mathematics

§12 Notation and philosophy of mathematics

The effects of notation which I have identified in Chapter 1 are all candidates for an account of the source and nature of the cognitive peculiarity of mathematics, which is at the bottom of philosophical puzzles about it. The three aspects of mathematical knowledge which I discussed in Chapter 1 do not provide a very useful classification of philosophical puzzles about mathematics, because they assume a certain picture of knowledge which I believe to be misleading, and which I shall be opposing. In the introductions to their standard book of readings, Benacerraf and Putnam identify a number of "problem areas" for philosophy of mathematics, without claiming any systematic classification: infinity, mathematical truth, set theory, "access", mathematical objects and intuitions about them¹. I agree with Benacerraf and Putnam that it is not easy to give a neat list of the concerns of philosophy of mathematics. The following questions are commonly regarded as important in philosophy of mathematics. Are mathematical truths necessarily true? What role do proofs play in their being so? What objects is (pure) mathematics about? How do we come to know mathematical truths? Is mathematics really concerned with the infinite, and if so how can this be reconciled with our apparently finite minds and capacities? How is mathematics in the pure sense related to mathematics in the applied sense, and how can mathematics be applied so variously and so effectively to the world? Less commonly discussed, but in my opinion equally important, is the question: how is mathematics made? These questions are all interrelated, and though the order in which I have asked them here is a "logical" one, the discussion of any one tends to involve the others too. In the next section, I will identify seven **problems** which I

¹ See Benacerraf and Putnam [1964] and [1983]. Two of these "areas" are discussed in several sections of their introduction, certain writers are treated as topics in themselves and Wittgenstein's philosophy of mathematics is dismissed from the second edition - on rather flimsy grounds, I suggest below. A more organised scheme may be found in Kielkopf's book [1970: ch1]. Kielkopf's scheme, he says himself, derives from Körner's simpler set of test questions for the three classical philosophies which he examines [Körner 1960: 9-11]. Hookway sums up the problems as they appeared to Peirce as [1985: 184] "how there can be a discipline which provides certain, exact, necessary, surprising and, above all, useful, results."

think any philosophy of mathematics must address. I will discuss them in an order which allows me to introduce a number of philosophies of mathematics in historical sequence. If a point of view can deal satisfactorily with these problems, other related problems will also be tractable. None of these traditional problems directly engages the questions about notation which I have asked, but I shall argue that they may all be better understood, and answers attempted, by examining the question of the role of notation in mathematics. They will be addressed explicitly in Part III.

§13 Seven problems of philosophy of mathematics

The object problem

Platonism is the original philosophy of mathematics, whose main point is that numbers and shapes are beings of a reality greater, because timeless and changeless, than that of ordinary things. Mathematics is about nonphysical, platonic *objects*. Mathematics is not only objectively true, it is *objective*. Such a view is the natural response to the invention of a form of discourse in which numbers and shapes are the things under discussion. Apparent reference is real reference. '14*8.5=119' states a relation holding among eternal, nonphysical objects to which '14', '8.5' and '119' refer just as 'the cat sat on the mat' reports the holding of a relation among physical objects to which 'cat' and 'mat' refer. This opinion is natural to working mathematicians. Virtually all of them operate with such a viewpoint, though these days the mathematical objects would generally be claimed to be no more real than, but merely just as real as, physical ones. Most will admit to it too²; only a few can be shamed by its evident philosophical difficulties into a rather hypocritical sunday formalism. (In Part III I will explain this tendency without subscribing to it.) Plato's invention of the world of Forms, or any similarly realist attitude to the subject-matter of mathematics, carries with it

² The classical admission in our time is that of Hardy [1929].

natural answers to other questions about mathematical knowledge. Mathematicians describe a particularly stable part of reality; the peculiarity of mathematical knowledge derives from its being knowledge about definite, unchanging and eternal Forms. The relation of applied mathematics to pure, as we would put it now, is for the platonist the relation between ordinary apples and the Form of the apple, one of "participation" - a kind of identification made unclear by embodiment. One does not create mathematics on this view, but only attains to a kind of observation, freed of the illusions of "appearance".

Platonism as a position in philosophy of mathematics is a source of endless difficulties, since to treat the number 3 as an object as real as, or more real than, a cat is to invite questions whose answer is obvious for the latter object but not the former. We don't feel numbers, or feed them, or see them move from mats. We don't know about them by perception at all. How then do we know things about them? For platonism most of the problems about mathematics concern access and arise from taking a simple and direct approach to **the object problem**. The object problem is the problem of giving some "scientifically" credible account of the subject-matter of mathematics. What is a number, a topos, a graph, a random variable? An account is needed which seems credible given what we generally assume, granted the scientific picture of the world, about ourselves and the kinds of things we deal with. The "surfaces" which Griffiths is really analysing in exhibit #10 are not like the crude paper models which he apparently discusses, three dimensional and rough, subject to wear and tear, with only nearly straight edges. Close reading of the text shows they are they are purely two-dimensional, not made of any particular material, ideally straightedged.

What we find from this and many other examples is that mathematical discourse apparently refers to a whole realm of objects which are quite unlike the objects with which we are as physical beings surrounded and which provide us with our knowledge of the world via perception. We have a rough idea of the way perception and manipulation of the environment leads to a knowledge of it in

animals such as ourselves. How can we possibly have come to know so much about things which simply are not in the world for us to perceive and manipulate? Is mathematics really about things with these ideal, that is unreal, properties of infinite thinness or complete precision or eternal unchangeableness³? A platonist viewpoint on mathematics leads to difficulties in accounting for our access to the objects it postulates, and is therefore to be avoided if at all possible for philosophers who take seriously the "scientific image" of the world. The difficulty then is to find an alternative account for the apparent objects which mathematics concerns. This is the object problem.

The infinity problem

A variant on Plato's opinion about mathematical objects is contained in Aristotle's philosophy. For Aristotle, Forms are not supposed to have separate existence, but only in their instances. Aristotle had a much more concrete, observational, this-worldly outlook than did Plato. Exactly what Aristotle's views on mathematics were is not too clear; but structuralism, a view I shall mention below⁴, has clear affinities with his general modification of platonism. There is interesting recent work⁵ on Aristotle's views about geometry and arithmetic, though the two seem not very well integrated. I shall discuss in Part III his idea that mathematical objects arise by our taking physical triangles "as" purely planar, precisely bounded etc.

One does not go far in mathematics without the use of terms which appear to refer to infinite objects, processes or properties. For example, the set of real numbers with which Hardy's first chapter is concerned (Hardy [1963: ch1]; cf exhibit #12Bi). How is it possible to refer successfully to a set whose element-by-element identification not merely exhausts all possible recognitionary action by finite creatures in space and time, but in principle resists such a

³ Lehman calls them "queer entities". Cf his [1979], passim.

⁴ See §19.

⁵ See Mueller [1970], Annas [1976] and Lear [1982]

procedure at all - this set is not only infinite, it is uncountable? What does 'reference' mean if we can refer to this set? In Euclid it is specifically laid down⁶ that lines have no breadth and so on - yet no lines are really so, or really intersect in unique points, or extend indefinitely. Indeed it is doubtful if there are any line segments, let alone lines in the world. How can we with our coarse faculties grasp properties of objects which seem to presuppose an indefinitely extended power of discrimination and perception?

Aristotle's is the first significant discussion of the role of infinity in mathematics. Are there really infinitely long lines, as geometry assumes, infinitely many prime numbers, as number theory shows? On the platonic view, yes, certainly (in for penny, in for a pound). On any view which wishes to avoid allowing that mathematical objects exist in a realm independent of the physical world, there is a new problem. Aristotle's theory of universals, that they exist but only instantiated, seems to be driven to postulating physical infinities to instantiate such particular mathematical objects. In order to avoid such a postulation Aristotle distinguished between actual and potential infinity, and declared that mathematics has no need for the former. So any straight line is potentially infinite, in that it may be produced further at will, but is not actually infinite, it has some particular finite extent. Whether this can really suffice for mathematics is still a live issue. Various constructivist positions are attempts to show that it can. Classical, nonconstructive mathematics at least appears to require completed actual infinities, most notoriously in set theory where infinite sets are actually operated on. **The infinity problem** is the problem of understanding how finite animals can grasp, or deal with or understand infinite structures like the integers, \mathbb{Z} , infinite processes like differentiation or infinitely thin lines. Or if this is impossible, then we must account for the apparent involvement of these things in mathematics, to give a positive account of the role of infinite objects, processes and properties in mathematics. The infinity problem seems to me to be mainly a case,

⁶ Heath [1956:153]

though a special one, of the general problem about objects. I shall not have much to say about it separately.

The truth problem

In the philosophy of Plato two kinds of knowledge, very differently valued, are possible for us, through the reason and the senses. In the beginning of the scientific age, the role of the reason in providing knowledge once again came to be counterposed to that of the senses, after a long period in which revelation was a more important source of knowledge than either. The rationalist philosophers, in particular Descartes and Leibniz, created a version of the older distinction which transferred its emphasis from the mode of cognition to the things known. These mathematical philosophers provided important ideas about the peculiar nature of the knowledge we have in mathematics. Descartes' radical distinction between mind and body, echoed Plato's opposition of the reason and the senses. Together with Leibniz' distinction between truths of reason and of fact, it underlies the continuation of a great divide in modern thought, mediated by the analytic and the synthetic, and still at work in modal logics where some theses are necessary and others not⁷. From perceptual knowledge and rational knowledge, emphasising their sources, we are turned toward empirical knowledge and logical knowledge, emphasising the nature of the things known: empirical knowledge is knowledge of how things are, logical of how they must be. Leibniz proposed the idea of truth in all possible worlds to account for the nature of logical facts. Both philosophers emphasised the role of analysis in the formation of knowledge. Descartes' criterion for truth lay in its exhibition as analysed into clear and distinct ideas, while Leibniz called logical truths analytic to emphasise their reducibility to the law of non-contradiction. In these as in many other writers of the time the ideal of knowledge is given by the organisation of mathematics in Euclid's Elements, and

⁷ There is of course an enormous literature distinguishing among all these distinctions, and I am not trying to give any kind of analysis of them here.

their ideas are attempts to construe all knowledge in ways which make it out as similarly founded and structured.

These various distinctions underlie the problem of accounting for the felt peculiarity of mathematical knowledge. In what consists the truth of mathematical truths; are they true in a special way? Are they true in all possible worlds - what is a possible world? Could a mathematical truth be false? For example, Pythagoras' theorem, Euclid 1.47, seems to compel assent in a different way to that achieved by a true sentence like 'Edwin Coleman has two arms', no matter what evidence is marshalled for it. That the time and place of a body moving in a parabola are related as Newton shows they are, seems to be true in a less accidental way than that a particular body, say a cat, is now at a certain place, on the mat. The one seems to be necessarily true, the other not. That employee EB, who worked 14 hours at \$8.50 per hour *therefore* earned \$119, has a kind of irrefutability which the working 14 hours does not - the record might be in error. In all these cases and many more which can be found in the exhibits, the truth of mathematical assertions seems to have a quality of necessity which other, "non-mathematical", assertions in the same contexts do not. Consequently it seems that mathematical knowledge, the knowing, is certain as non-mathematical knowings are not; and that mathematical knowledge, the known, is true in some more final sense than are non-mathematical facts.

But if there is a distinct kind of truth to which mathematics attains then one will need some account of how such truths are related to the more homely truths of empirical knowledge, how the apparently empty tautologies of logic and mathematics can play any informative role in science, and of how if not through perceptual evidence, these truths are established, founded and known. The **truth problem** is the problem of explaining why mathematical knowledge has a *specific and distinct* kind and level of security. Some explanations consist in denying such a distinction, Mill's and Quine's for example. (One can distinguish the problem of

necessity from that of the peculiar nature of mathematical knowledge, but I resist this because it tends to presuppose platonism.)

The application problem

Kant's difficult philosophy has a negative part: the antinomies arise from extending beyond all possible experience, concepts which are useful for dealing with experience. (Intuitionism develops this idea.) It also has a positive account of the nature of thought to which it is difficult to do justice briefly. Since his system is essentially an explanation of the "necessary truth" of the Newtonian physical world-view, the important thesis is that mathematical truths are *synthetic a priori*. That is, they are not gleaned from experience, they are a priori not a posteriori. Nevertheless they are not "mere" logical tautologies, as Leibniz claims and Hume following him, because they involve *constructions in intuition*. They are synthetic not analytic. Previously, for example by Hume and Leibniz, the a priori and analytic had been identified as the logical, as had a posteriori and synthetic as the empirical. Kant was provoked to recognising the novel possibility that these two distinctions do not coincide, by the problem in accounting for the necessary truth of Newtonian science which seems to arise from any radical division of logical truths from factual ones.

The **application problem** is this problem of understanding how it is possible to make such effective use of mathematics, with its apparently ideal subject-matter, in dealing with the empirical world. The necessary fact of mathematics that $14 \times 8.5 = 119$ is applied in exhibit #1 to the wages owing to EB. How can the abstract relationship be "applied" - surely it either holds or does not in any case, what could one do to make it hold by "application"? The figures themselves presuppose application - why do we take it that exactly 14 hours were worked? This kind of application cannot be understood without some account of measurement, which is a particular kind of application of mathematics⁸.

⁸ See Chapter 11 for a discussion of measurement.

Kant's theory of experience provides an answer to the application problem which manages to retain the distinction between the logical and the empirical, essentially by interpreting the logical as the framework of concepts within which we must necessarily have any perceptual experience. Space and time are the forms of our intuition, mathematics the explication of their properties. Now there are numerous difficulties which can be raised about Kant's suggestion, but the important thing for the present is this: if Kant's account of the relation of the theoretical and the empirical is rejected, then we must make clear in some other way how the theoretical reasoning of mathematics is brought to bear on the empirical world. That is the application problem.

The access problem

Mill's views are hardly as influential as those of Kant, so far as recent writing on the philosophy of mathematics is concerned. But they form an interesting contrast, combining a return to the simple cleavage between the logical and the empirical, which is the leitmotif of English philosophy before and after Kant, with some advance in the understanding of logic and therefore of logical truths. Mill tried⁹ to make mathematics certain by making out its inferences as all deductive, yet empirical because its axioms are drawn from observation by abstraction. In Mill's view the assertions of mathematics are simply very general empirical assertions. Frege lampooned¹⁰ Mill's rather simple idea of abstraction, but Mill's view in any case can hardly account for the mathematics of nonEuclidean geometries and the plethora of alternative algebras which was growing around him in the nineteenth century¹¹. In any case the defect of any such theory as Mill's is that it can account only for the mathematics which is actually applied in science and not for any mathematics which is not or cannot be applied.

⁹ Mill [1843]

¹⁰ Frege [1884: §§7, 23]

¹¹ Kitcher [1983: passim, ch 4 and 6 particularly] and Resnik [1980: 157ff] have recently tried to rehabilitate Mill's views somewhat.

The problem is that by denying mathematics any subject matter of its own, besides the physical world as it is, there is no longer any room for alternative subject matter which the various geometries or set theories treat of. The full development of the axiomatic method has made this a rather embarrassing problem for such an empiricist view. The further development of logic at the end of the nineteenth century was the anvil on which the presently prevalent views on mathematics were hammered out.

On the other hand, the **access problem** arises from virtually any answer to the object problem which does give mathematics a subject matter of its own: how can material creatures know ideal objects? (Under this heading Benacerraf and Putnam, discuss both this and the preceding set of questions I have discussed as the object problem. Of course they are all interrelated.) In Euclid and in Hilbert and in other works of geometry, and in many other mathematical texts, axioms are laid down at the beginning. Except in certain views, which will be considered later, these are taken to be truths about the objects of the text. How have we achieved knowledge of these truths, which are not empirically testable? In most mathematics, for example in exhibit #5, inferences are made from certain mathematical facts to others. The validity of these inferences is generally explained in terms of the impossibility of things being otherwise. How can we possibly know such brute-fact-transcendent generalities? We know very little about what is the case about our own bodies - what causes aging? is schizophrenia somatic? - yet we seem to be able to make *certain* statements about all possible right-angled triangles - no one of which can be interrogated in any way which anyone has ever satisfactorily explained. In some bold-faced writers, an ability to directly apprehend such mathematical objects is posited - mathematical intuition¹². This solves the immediate problem,

¹² Gödel's declaration [1947:483f] is the most respected, though it contains almost no explanations at all; Hardy's [1929] is the most persuasive, since it powerfully records the experience of the practising mathematician. Recent attempts to make sense of the idea are Parsons [1979] and Maddy [1980]; see also Steiner [1975: ch4]. Benacerraf [1973] and Kitcher [1983: ch 3] point out some of the difficulties of this move. I discuss the Intuitionist philosophy (which is slightly different) in Chapter 3.

but only to create worse ones: how is this intuition to be fitted into the rest of our picture of the human organism? How are different intuitions to be reconciled? How does it work? The problem of finding some way to answer these questions, to explain how despite their disconnection from perceptual information-gathering we can achieve knowledge about triangles, toposes and transcendental numbers, is the **access problem**.

The proof problem

We return then to the idea that mathematics has its own proper objects since the ordinary objects of the world won't do. They won't do because they are quite unlike the objects which mathematics treats of, which are exact, eternal, unchanging, causally inert and so on. (Subsequently, particularly in Chapter 12, I shall consider the idea that mathematics does not treat of objects at all.) I have mentioned Frege's harsh criticism of Mill's attempt at an empiricist ontology for mathematics; Frege embraced the obvious alternative to Mill's, that mathematics does have objects of its own. Then of course he must provide some kind of answer to the access problem. But the kind of answer which Frege gave avoids the attempt to interpret mathematical facts on the model of empirical facts, as translations or constructions from some mysterious counterpart of the perceptions which are usually thought to back empirical facts. Rather, Frege puts emphasis on the role of proof. He wishes to show that mathematical knowledge is justified through being obtained by rigorous proofs from premisses which are justified by their being logically necessary. The "basic laws of arithmetic" are logically necessary, and so are all the deductions from them which give us the not so basic laws. He tried in his theory of sense to defend this Leibnizian answer, that mathematical knowledge is analytic, from the charge that it makes mathematics trivial. Before this problem became urgent however, the first task was to clarify just what proof in mathematics is. Frege found those proofs offered in the text books of his day

shockingly lacking in rigour. He set to work to improve them, and his labours initiated developments which resulted in the present concept of a formal proof.

Mathematics is certainly full of proofs, though I would not accept the common identification of mathematics with its proofs (see Part III below). Many mathematical claims are justified by proofs, and it seems obvious that the peculiarity of mathematical certainty arises from this source. But are things as they seem: is mathematics certain, and if so are proofs the reason? In what do proofs consist, how do they differ from arguments, what makes a text a proof? Is there is a useful contrast between formal and informal proofs? Do formal proofs really explain the role of proofs in mathematics? What is the nature of the compulsion which proofs are said to convey? To answer questions like these is the **proof problem**.

The origin problem

A problem which is often overlooked in philosophy of mathematics concerns the origin of mathematics. Much philosophy of mathematics assumes some kind of platonism or is otherwise fixated on the crystal world of mathematical product, the theorems known rather than the knowings of them. Consequently the process of mathematical creation, or more generally of mathematical thought, is usually pushed aside as of merely psychological interest. This is largely attributable to the influence of Frege, whose polemic against his particular idea of psychologism (logic as the science of thought) has been accepted by most philosophers this century as obviously correct. The mind has been excluded from philosophy of mathematics ever since. By the same token, so has mathematical knowing.

There are however deeply puzzling questions which can be raised about mathematical creation. Plato's *Meno* recognises some of them in its attempt to think through the way a piece of mathematics might be learnt, in such a way as to mimic the possible origin of the theorem Euclid I.47. What is the nature of the

process whereby we actually come to recognise the truth of a piece of mathematics? How do we arrive at conjectures, and how do we evaluate them? Why is mathematics difficult to learn and why is it only a recent (in evolutionary terms) presence in human culture? How are mathematical truths discovered, proofs found? How is mathematics preserved? How do we know that it is appropriate or legitimate to apply old mathematics in particular new contexts? Questions like these constitute the **origin problem**.

This completes my sketch of the seven problems which I regard as central to the philosophical interpretation of mathematics. I shall not attempt to give my solutions to these seven problems explicitly until Part III, though the problems will be mentioned repeatedly in the following discussion of various later philosophies of mathematics which have been proposed to deal with them, as well as in various places in Part II. In the next six sections I will give a sketch of developments in philosophy of mathematics since Frege, because current writing on these issues is dominated by a particular point of view which has been consolidated over the last century. There occurred a change of focus from the classical problems which I have just described, to a new problem: the foundations problem. I shall argue in what follows that the reason for the unsatisfactory current attempts at philosophies of mathematics can be traced to the analysis of mathematical language which we know as the mathematisation of logic; for this analysis led to the reinterpretation of the traditional problems as *waiting on a solution to the foundations problem*. That problem has proved to be insoluble; hence the present disarray. But this disarray presupposes various assumptions of the *logico-formalist analysis of mathematical discourse*, as I shall call it. I believe that a new kind of analysis of mathematical discourse can cast new light on these old problems. In the next section I will outline the mathematisation of logic, and in the one after that the foundations problem to which it gave rise. In subsequent sections I will review the positions on offer in the philosophy of mathematics in the light of the effect of

these developments. In Part II I provide a new analysis of mathematical discourse, and I bring it to bear on the seven problems in Part III.

§14 The mathematisation of logic

Before Frege, there was no such separate discourse as the philosophy of mathematics, although it is perfectly reasonable and possible to identify the positions of earlier writers on the characteristic problems of the philosophy of mathematics, in delineating "Kant's philosophy of mathematics", and so on. The views of Plato, Aristotle, Descartes, Leibniz, Kant, and Mill amongst others are important, but all the same the idea of treating mathematics by itself as a subject of philosophical attention is really a twentieth century development. The differences in views to be found prior to Frege, for example the contrast between Plato and Aristotle on the separate existence of the objects of mathematics, or between Kant and Leibniz on the analyticity of mathematical truths, are relatively timeless ones. They presuppose the existence of mathematics, but not of any *particular* mathematics. Euclid's Elements is all the example any of them ever need (and the one they generally have in mind!). This is no longer true of the philosophies of mathematics created by and after Frege.

During what might be called in retrospect the Fregean period, he along with Russell and others created the modern formal methodology called "symbolic logic", building on the efforts of many other nineteenth century logicians - Venn, DeMorgan, Boole, Schroeder, Peirce and many others - to find a mathematical logic. To ask whether what was sought was a logic adequate for mathematics, or a mathematical formulation of logic is perhaps not a fruitful dichotomy, for what happened was a considerable rapprochement of logic and mathematics culminating in the logicist attempt, beginning with Frege, to exhibit mathematics as actually part of logic. Naturally, any logic for which this could be shown would have a rather mathematical formulation. The present methodology of formal languages and systems was developed over the period roughly from Frege to Gödel. It is the

realization in *Principia Mathematica* and similar treatises, to the satisfaction of many writers, of the ideal of a fully explicit notation for mathematics which is simultaneously a perfect logical language. This made it possible to think of mathematics as a distinguishable discourse which could be accounted for separately from everything else. The original impulse behind this effort would seem to have been the need to clarify analysis, which became difficult to understand in the nineteenth century as more and more "non-intuitive" functions were required by Fourier Analysis and related developments in analysis. Cantor's researches in the theory of sets of points arose in this way. The obscurities towards whose clarification this work was directed should not be confused with those which it was responsible for creating. In particular, the problem of the set-theoretical paradoxes, which became a great spur to logical work after their crystallization around the turn of the century, was not in the least part of Frege's motivation, nor even of Russell's to begin with. But the new logical technique made the paradoxes so rigorous, that naturally (but erroneously) their solution was sought in the same technique.

The paradoxes¹³ created a new problem for the philosophy of mathematics, that of its security, namely, *is mathematics secure?* Can we always rely on it? Does it have the epistemological unimpeachability it was traditionally accorded? If not, how can we make it so? This had never been seriously in question before. The philosophy of mathematics was transformed into the problem of foundations.

§15 The foundations problem

The paradoxes are as old as philosophy, and some would say, as pointless; in particular, the paradox of the liar is of great antiquity. They are puzzling, even infuriating, but not really *important*. What really turns on their resolution?¹⁴ However, around the end of the nineteenth century several different versions of

¹³ Cf Fraenkel [1973] for an account of the paradoxes as a spur to mathematical logic.

¹⁴ I agree with Wittgenstein on this point but I shall not discuss the issue here.

such paradoxes converged to create a crisis in confidence in mathematics from which it has never recovered. Philosophical confidence, that is - bridge-builders never paused to worry about them, a significant fact to which we shall return in Chapter 10, nor did the great majority of mathematicians¹⁵.

Several paradoxes were formulated in the nascent theory of sets of Cantor, some of them only apparent paradoxes in that no real contradiction was involved, but only a challenge to accepted ideas, as in Cantor's theorems showing the rationals countable, and so on. But the paradoxes of Burali-Forti, and several "semantic" paradoxes (Berry, Richard) worried many mathematicians. When Russell discovered an inconsistency in Frege's system, which they had thought was going to give mathematics its true foundations, and formulated it in the simple form known as Russell's paradox, the roof had fallen in. Those numerous mathematicians who had distrusted Cantor's theory from its inception, in the tradition of Kronecker distrusting all nonconstructive arguments, felt vindicated. Those who liked it, or thought they needed it, were challenged to action.

The problem now seen as urgent was to meet the threat to the cognitive status of mathematics: not only were the sublime truths of mathematics no longer inhumanly certain, they were based, or so it seemed, on outright contradictions and therefore quite valueless. Worse, from the worried mathematician's point of view, one no longer knew what results to trust. It was essential in order that mathematical work should go on (no one considered that perhaps it should stop) to find a formulation of mathematics which could be convincingly shown to be free of contradictions and yet adequate for science and technology. Finding this formulation and the demonstrations of its consistency and adequacy became serious mathematical tasks - and the philosophy of mathematics became little more than an ideological battleground for contending programs aimed at achieving this task.

¹⁵ Even Hilbert was motivated not by insecurity but by the need to repudiate the attempted intuitionistic decimation of the mathematics science needs.

§16 Logicism, Intuitionism, Formalism

A research program was initiated of great scope, that of the **founding** of mathematics, the complete systemisation of mathematics on a rock-solid basis. There was some considerable difference of opinion however as to what this basis should be. For Russell, Frege and their heirs it was logic, as perfected in their own researches; for Brouwer, the intuition; for Hilbert and his school, formal arithmetic. The competition between these views led to the customary¹⁶ trinary division of philosophy of mathematics into logicism, intuitionism and formalism which has been rather enshrined in much of the literature. But it was really always a distinction between schools of mathematics. Beeson suggests¹⁷ that nonconstructive mathematics was invented during the 19th century, that until then all mathematics was (essentially) constructive. Functions were given by rules, proofs were basically computations or constructions. There are some famous protests about the incursion of nonconstructivity into mathematics: Kronecker's dictum¹⁸, Gordan's protest about a proof of Hilbert's¹⁹, even Frege's controversy with Hilbert²⁰ has elements of it. But the first sustained effort to swim against the nonconstructive and Platonist tide in mathematics around the end of the 19th century is Brouwer's. This tide had several sources - the development of analysis principally, Fourier analysis being at the root of Cantor's set theory, but pathological examples of Weierstrass's and others arose in the course of solving various problems of analysis. The structuralist implications of Galois' theory of equations were also inexorable once they gained recognition²¹. Apart from the

¹⁶ Cf Körner [1960], Benacerraf and Putnam [1965].

¹⁷ Beeson [1985: 418ff]

¹⁸ 'God made the integers, the rest is the work of man.'

¹⁹ 'This is not mathematics but theology.'

²⁰ See Frege [1984b]

²¹ Emmy Noether symbolises the transformation of algebra to its present abstract form; her father was a mathematician of the Gordan school, but she was a pupil of Hilbert. Students of hers such as van der Waerden completed the formulation of Galois theory in its present abstract form.

protests of the old guard, what gave the rush to idealism pause was only the paradoxes. And even these were for most mathematicians to be circumvented, not retreated from. As Hilbert said: "No-one shall drive us from the paradise which Cantor has created for us."

Logicism

Logicism, from Frege through Russell and beyond²², focussed with great insistence on the fact that mathematical argument is *deductive*. In fact it proposes to show that there is nothing to mathematics but deduction. Contrary to the view of Kant, on this view, there is no need for any faculty of intuition in mathematics. All the basic terms and propositions in mathematics can be derived from logic alone, all mathematical inferences are purely deductive, so mathematics is logic. This program failed, essentially because no-one could convince anyone else that the necessary axioms, in particular those of infinity and choice, were logical - or even should be taken as axioms. Russell made them explicit hypotheses in the theorems which need them, thereby casting mathematics into a hypothetical light which has sometimes been adopted as a philosophy as "if-thenism" or "postulationism"²³. The logicist program showed at the very most that mathematics is set theory. I shall argue below that this is still far from the truth. But its ideology lives on, particularly its Euclideanism and its "rejection" of intuition. Logicism will be discussed at several points in the sequel, particularly in Chapters 7 and 12.

Intuitionism

Intuitionism takes from Kant as the salient fact about mathematics that it consists in *actually effected mental constructions*. Consequently real paradoxes

²² Frege [1879], [1884], [1893]; Russell [1905], Whitehead and Russell [1910-13], [1962]; Church [1956]; Bostock [1974] and [1979], Steiner [1975], Mayberry [1980].

²³ Cf Lehman [1979 : ch 1,2], Russell [1937] and Putnam [1967b]. The Russell form of logicism also required the axiom of reducibility, which even he acknowledged to be simply a trick to get around some side-effects of his type theory. ZF does not need types, so this is less of a difficulty for more modern logicians (or logico-formalists as I shall call them: see below). Cf Ramsey [1925].

are impossible and only paper tigers - they arise from illegitimate attempts to represent in language results of mathematical activity which could not in fact occur. Could not occur because they presuppose the completion of infinite processes. For example, some statements about the complete decimal expansion of π have no place in genuine mathematics, for Intuitionists. This research program, though confined mostly to Holland, achieved a certain philosophical importance. Its denial of the law of excluded middle (see Chapter 3), and its proposal to abandon much traditional analysis as unrecoverably infected with infinitistic thinking, galled many mathematicians immeasurably, who looked to the satisfactory completion of one of its rivals to *justify* their labours. I will discuss some of the philosophical aspects of Intuitionism in Chapter 3. References are given there.

Formalism

Formalism puts in the centre of the mathematical stage the *calculations* at the heart of arithmetic and geometry, and attempts to show that the advanced theories of mathematics on which analysts work are merely instrumental: mathematics is to be exhibited as a conservative extension of a finitary core which is self-evident²⁴. It is not true, what many of formalism's opponents said of it²⁵, that it considers mathematics to be meaningless game of marks on paper. On the contrary, the idea was to treat mathematics **as if** it were only a game of marks on paper in order to prove, without any possibility of being misled by their interpretation, that its procedures are reliable, that $1=0$ can never be proved. This "as if" implies that mathematics is not really such a game. It is true that formalists have mostly been interested in the search for foundations, and much less in

²⁴ Cf Hilbert [1926], [1928]; Curry [1952], [1963]. See also Detlefsen [1986] on what Hilbert's program really is, and Resnik [1980: 76-104] and Kitcher [1976] for some difficulties in its philosophy. I discuss some of these difficulties in Chapter 12.

²⁵ For example, Russell [1959: 110]

providing a comprehensive philosophy of mathematics. But this confusion of the problem of foundations and the problems of philosophy of mathematics is one which is endemic in modern writings. Following the later Wittgenstein²⁶, I will argue en passant below that the problem of foundations is not a *philosophical* problem at all. But it should be stated now that the conception of mathematics presented in Part III of this thesis might well be described as a kind of hyperformalism, not in the sense of an attempt at foundations, but in its deliberate concentration on the actual processes of signification in mathematics as providing both the material in need of understanding, and the means of gaining it.

All three of these twentieth century philosophies of mathematics have roots in earlier doctrines. Logicism is derived from the rationalists Descartes and Leibniz while Intuitionism is self-consciously Kantian. Formalism is more novel as a philosophy of mathematics (and less developed) but it too has clear forebears in the nominalist strand of the debate about universals²⁷. But equally important, each school of thought was a program of mathematical activity. Each sought to create a formulation of mathematics which could escape the foundations problem. The debate in the first three decades of this century was essentially between the Intuitionists and the Logico-formalists²⁸. The former proposed to solve the foundations problem by eliminating most of the analysis actually needed for science and technology, with the rationalisation that it was based on the invalid use of LEM for infinite totalities. The latter were not prepared to reduce mathematics to a safe but useless Dutch artform, and sought to create formal systems which could be seen to be adequate for analysis and proved to be sound. The Logistic method was the basis for Hilbert's plan, the Formalist turn consisted in applying it to

²⁶ Wittgenstein's views will be discussed in Part III.

²⁷ We shall see that connection in Field's program (Field [1980]), discussed in Chapters 7 and 12.

²⁸ Logicians and Formalists were not really opposed, the Formalist program succeeding the Logician one. Nineteenth century Formalists like Thomae had been silenced by Frege, but Hilbert was actually provoked towards a satisfactory metamathematics by his controversy with Frege. See also Resnik [1980].

mathematics itself - the creation of metamathematics. So the Intuitionists on one side created their new mathematics of real number generators, of spreads and species and of theorems false in classical mathematics but lacking many basic theorems of analysis; and the Logico-formalists on the other created a myriad of formal languages and systems and theorems about them, providing as a by-product a new level of consciousness about the notation used in mathematics and a new standard of rigour in its deployment. The basic tool of meta-mathematics, the formal system, was perfected during the first three decades of this century, going beyond the early logicist account in its disconnection from interpretation, in its self-consciousness about derivation, and in the arbitrariness of the axiomatic bases considered. Work proceeded apace on the Formalist reconstruction program until the impact of Gödel's discoveries²⁹.

§17 The impact of Gödel's theorems

The theorems proved by Gödel in 1931, specifically his second theorem, destroyed Hilbert's program³⁰. Consequently the philosophy of mathematics virtually disappeared for thirty years, having become identified with the foundations problem which now seemed insoluble. But more, not less, meta-mathematical work followed Gödel's; and a new philosophical use was found for the logico-formalist methodology by positivist philosophers of science. These claims will be justified in this section.

Gödel's second theorem showed that no consistent formal system containing arithmetic (Gödel's title refers to systems like that of *Principia Mathematica*) could also contain a proof of its own consistency. This destroys Hilbert's program as then conceived, since it means that no formulation of mathematics could at once be adequate for analysis, that is, fall into the class studied by Gödel, without thereby eluding what Hilbert sought, namely a proof of its consistency which would be

²⁹ Gödel [1962]

³⁰ Some have denied this recently, see Detlefsen [1986]; but I think no-one denies that it was taken that way in the thirties.

acceptable to sceptics like the Intuitionists. For if such a proof existed it would need to be frugal enough in its requirements to be formulable in just such a system as was being proposed. Gödel's theorem showed that this could not be. The opinions of the Intuitionists about the uselessness of formalisation were confirmed by the impossibility of the logico-formalist plan implied by Gödel's theorems, and other limitative theorems such as Tarski's and Löwenheim's. But their own plan gained few adherents and only really became clear, and clearly inadequate, as formal methods were applied to it by Heyting, Kreisel and others. But almost no-one had wanted it to work anyway, because what was generally being sought was a *justification* of the mathematics used in actual practice, not a mutilation of it. Since Logicism had not convinced anyone that mathematics was logic (rather than set theory with some fairly hefty assumptions), the effect of Gödel's 1931 paper was to leave the foundations problem without any possible solution. Everyone despaired.

Gödel's results halted the confident building of "foundations" which had been going on, and the philosophy of mathematics remained static for about thirty years. The philosophical interpretation of all the theorems being proved became less programmatic and less vociferous in this period, though many profound facts were discovered. But no new strands were added to the philosophy of mathematics. The writings of Wittgenstein only began to be discussed late in this period, since his unpublished writings on these topics dribbled out only slowly after his death. Ironically, the forum in *Erkenntnis*³¹, in which the Logician, Intuitionist and Formalist "foundations" are presented together, only appeared when it had become clear that none of them could work. Körner's book of 1960, thirty years later, might equally well have been written then, for all those thirty years seem to have contributed. Mehlberg's survey of the situation in philosophy of mathematics at the end of the period³² is equally revealing. It consists of a recital of the deficiencies

³¹ Reprinted in part in Benacerraf and Putnam [1983].

³² Mehlberg [1962]

which have been found in these same three programs, and proposes as a "solution" to the foundations problem what he calls "pluralistic logicism". Clearly descending from Carnap's Principle of Tolerance³³, this is nothing more than the suggestion that providing foundations consists in doing meta-mathematical work, while resiling from any claim to be able to achieve a unified final result. Not only have problems of the philosophy of mathematics been entirely identified with those of the foundations problem, the solutions have become nothing more than work on that problem! Again, as late as 1969, Hintikka defiantly prefaced his collection of readings in philosophy of mathematics³⁴ with a dismissal of (then) current writing on philosophy of mathematics as worthless and a declaration that articles of the kind he had chosen were "by far the most useful for a student of the philosophy of mathematics". In the articles in question, as he says, "no specifically philosophical problems seem to be discussed". They are samples of meta-mathematical work. He then goes on to summarise the recent developments in mathematical logic which his selections represent and ignores their possible philosophical implications. There is a deep confusion at work here of which we can find a symptom in Hintikka's comparison between trying to do philosophy of mathematics without a familiarity with such material, and discussion of legal philosophy "not based on any knowledge of the actual operation of the law". Now I agree with the thrust of this remark; but to what is Hintikka comparing knowing the operation of the law? One would expect the parallel to be a knowledge of the operation of mathematics, but he actually means metamathematics. A knowledge of the latter is certainly a prerequisite for the philosophy of meta-mathematics - but is that the same as philosophy of mathematics? Only on a certain assumption about what actual mathematics comprises - namely that it is metamathematics (Curry's view) or at least that meta-mathematics is a faithful representation of it. I

³³ See next section.

³⁴ Hintikka [1969]

shall be trying to undermine this prevalent assumption in this thesis, and I will discuss it further in the next section.

The logico-formalist research program in mathematics which had been set up had too many successes and too much momentum to be stopped by Gödel's results and meta-mathematical activity continued undisturbed, indeed it intensified. The modern divisions³⁶ of mathematical logic - model theory, recursion theory, proof theory, set theory - were all developed in this period. A very great deal of profound and interesting knowledge about formal systems accumulated. Unfortunately it could not be interpreted as solving the foundations problem with which philosophy of mathematics had been replaced. Nature abhors a vacuum, and (perhaps by the processes dubbed by Hegel the cunning of reason) a different ideological function for all this metamathematical knowledge was found: it was used to consolidate the positivist paradigm in the philosophy of science by the attempt to extend to the language and discourse of science the "successful" analysis of those of mathematics. The leaders in this movement were Carnap and Hempel; Quine's role was also important, if more ambiguous.

What occurred in that period of roughly thirty years was however in some ways just as significant for philosophy of mathematics, despite the lull in activity explicitly addressed to its problems. What happened was a massive export drive, which still goes on, to convince the wider philosophical world that the logico-formalist method can solve general philosophical problems. Quine wrote a series of papers³⁷ which provided a new method for dealing with ontological questions, which can be traced back to Russell's "On Denoting"³⁸, but is much more firmly wedded to the logico-formalist paradigm of expression. The "criterion of ontological commitment" which he put forward in those papers has nowadays achieved an almost unquestioned status among philosophers writing in English on

³⁶ Cf Barwise [1977]

³⁷ Cf Quine [1961]

³⁸ Russell [1905]

mathematics and science³⁹. I will argue in Chapter 7 that it is delusory. Carnap argued⁴⁰ for a combination of logicism and formalism as the solution to the problem of foundations - formalism to provide a syntax for mathematics, together with further logical apparatus which will permit the representation of the application of mathematics, modelled on the logicist definition of '2'. This suggestion was merely part of a more general program to replace the philosophy of science with the "logical syntax of language" (Carnap). Thus the philosophy of science, of language, and various areas of epistemology became colonised by the logico-formalist methodology, so much so that various of its presuppositions have become a kind of dominant paradigm, that is now reapplied to philosophy of mathematics itself in a kind of vicious circle of self-confirmation. I believe that this circle of ideas is what Wittgenstein was trying somewhat obscurely to confront and attack in his later work. I shall do so more explicitly here. In the next section I will try to summarise what I take this dominant paradigm to be.

§18 The logico-formalist hegemony

The logico-formalist *characterisation* of mathematics was accompanied by the evolution of *conceptions* of language and mathematics which were not always made explicit, and not actually consequences of it anyway. These implicit conceptions of language and mathematics have become both pervasive and largely unquestioned, and form a dominant paradigm or hegemony, within which most recent writing on the philosophy of mathematics has flourished. I call this dominant paradigm, which I shall outline in this section, the **logico-formalist hegemony**. It is the offspring of the venerable but oversimplified conception of language, which Wittgenstein attacked as the "Augustinian picture" in §§1-59 of the *Investigations*⁴¹, and of the new technique of representation created in pursuit of

³⁹ For example, Field's program (Field [1980]), and many discussions of it, simply assume the correctness of this criterion.

⁴⁰ Carnap [1937: 325ff]

⁴¹ Wittgenstein [1953].

the programs of Logicism and Formalism. Underneath a surface clash of philosophies of mathematics in the literature there are largely shared assumptions about just what counts as the mathematics which is the subject of the philosophising, and about the nature of its expression and how that relates to "ordinary" language.

The logico-formalist formulation of mathematics was achieved by a careful scrutiny of the language and the special signs used in mathematics, and the extension of notational devices to the parts of mathematics usually expressed in words. Frege's work provided the necessary notational innovation for this development⁴², the school of his old opponent Hilbert perfected the notion of formal system, and Tarskian semantics connected the notions of proof and truth. The characterisation of mathematics which was created to solve the foundations problem, its actual formulation as formal systems with Tarskian semantics, did not in fact do so; but as an adequate expression of mathematics, it has been implicitly accepted by many writers. There is no explicit conception of mathematics, no set of answers to the classical problems of the philosophy of mathematics in such a characterisation; but certain kinds of answer are implicit, and philosophy of mathematics since then is largely attempts to rationalise that characterisation.

The basic idea in the Augustinian picture⁴³ is that words name objects, namely their "meanings", which are taught by ostension. A word's meaning an object consists in the intention with which it is used, and sentences are simply combinations of words. The relation of a name to its bearer, as of 'Edwin' to the present writer, is taken as clear and unproblematic, and as the prototype and fundamental relation underlying the role of every word. The second assumption is that every word is learned in situations of the kind often called "baptismal" in

⁴²Cf the full title of the *Begriffsschrift*: "Conceptual notation a formula language of pure thought modelled upon the formula language of arithmetic." For further discussion, see Chapter 7.

⁴³ For further exegesis, cf Baker and Hacker [1980, 21-124]

recent discussions of naming and necessity - for example, the naming of Devitt's cat⁴⁴. Thirdly, by identifying meaning and intending to mean, the Augustinian picture implicitly excludes writing from language. Finally the articulation of words is taken to be simple concatenation, corresponding to a simple compresence of features of the environment. The Augustinian picture is a conception of language as reflection, as consisting of strings of tokens produced in direct response to jointly present saliences.

Baker and Hacker point out that the Augustinian picture is not in itself a theory of meaning. Rather, it is the ground from which many spring, by the elaboration of ways to evade objections to it. The logico-formalist hegemony has arisen through the development of an Augustinian theory of meaning for mathematical discourse which has two distinctive features, contributed by Frege and by Russell. The Fregean idea is the *functional* conception of predication and the symbolisation of inference⁴⁵ which is based on it; Russell's contribution⁴⁶ is the idea of the "analysis" of a sentence to "reveal" its "logical form". Difficult issues in Fregean philosophy concern the attempt to treat predicates, linguistic items expressing relations among objects, as themselves names for objects (or, as the hegemony has it, as being referential). The tough-minded approach is to adopt extensionalism and cash them as sets of ordered pairs. Thus the possibility of the logico-formalist conception of language rests on the development of mathematical logic which has three main features. These are the extension of notation to the words in mathematics, first satisfactorily achieved by Frege; the perfection of the idea of a formal system, which may be taken to be completed by Hilbert's school in the 1920s; and model-theoretical (or referential) semantics which was created on this basis by Tarski and provides the final requirement, an account of truth.

⁴⁴ Devitt [1981: 26ff].

⁴⁵ Frege [1879]

⁴⁶ Russell [1905]

The characterisation of mathematics to which the logico-formalist characterisation leads is this⁴⁷. Mathematics consists of theories and models for them. Theories are certain classes of signs, which are conceived as abstract objects; models are other abstract objects, usually sets, bearing an isomorphic relation to the signs, a relation I shall be calling *iconic*. Mathematical discourse consists of wffs, which are (finite) strings of signs, taken from a fixed finite alphabet of primitive signs, satisfying certain recursive rules of formation which enable any string to be finitely classified as wff or not. The set of wffs determined on a given alphabet by a set of such rules is a formal language; a logic for such a language is a further set of recursive rules specifying a class of accepted consequences, which are ordered n -tuples of wffs; the rules specify permissible final members for all possible initial $(n-1)$ -tuples. A theory is a class of wffs determined by a language, a logic and a set of axioms (simply a set of wffs) as the class of final members of the closure of the axioms under the consequences. Proofs are certain sequences of such final members. In metamathematics the relations of theories and models are formalised. An interpretation of a theory is a map from the algebra of terms to a set algebra which preserves relevant structural features. Interesting properties of theories such as consistency and completeness, what kind of models it has, and so on, can be defined in various ways and treated rigorously in this framework. The class of proofs can also be studied. The possibility of these studies underlies the present near-consensus that the characterisation of mathematics on which it is based is acceptable.

This characterisation of mathematics is usually associated with a "picture" like this. Mathematics is a vast structure of certain and exact propositions arranged in a myriad of inter-related formal systems, describing abstract objects. Actual mathematics is approximated by the expressions we use in our concrete

⁴⁷ The following picture is presented in many books on the "foundations" of mathematics, for example, Wilder [1952: part 1], Church [1956], Kleene [1952]; more recent literature is less definitive about this, for example Hatcher [1982], but it is by now the accepted wisdom imbuing most undergraduate texts.

statements, "mathematics", which are genuinely mathematics insofar as their logical form is that of genuine propositions, with which they would be replaced by less limited creatures than ourselves, who would use entirely formalised mathematics and not the rather mongrel mixture of formal signs and vernacular commonly found in "mathematics". Many features of "mathematics" are in principle or logically irrelevant and do not occur in mathematics, such as definitions, diagrams and motivational remarks. The actual presentation, and a fortiori its context are quite irrelevant to understanding what it is. The truth of a mathematical thesis consists in its iconic copying of a fact about all the possible models of the theory to which it belongs. Its truth and its necessity coincide. Mathematics is applied by taking a real system as a "model" (it cannot *really* be one of course - models are abstract structures), whereupon the theorems may, if we have chosen well, be approximately true of it. How we choose, why it works and so on are dismissed as "pragmatic" questions irrelevant to understanding the logical facts about mathematics (but there is a common presumption that it works because the world can be "reduced" to objects of which mathematics is literally true.)

There is a tendency, which is common among philosophers impressed by mathematics and science, and some others, to extend to so-called natural language the account of mathematical discourse given by the logico-formalist characterisation. This seems reasonable, since it is generally thought that the logico-formalist characterisation of mathematics is a model of mathematical discourse, which of course usually involves considerable wordage as well as the usual kinds of notation, and has developed historically from predominantly verbal discourse. It is suggested, then, that there is a formal system (which we must try to discover) in which there is a counterpart for every sentence of natural language (revealing its logical form) whose meaning is given by its truth-conditions, which is a recursive specification for its interpretation in terms of its constituents. On such a view every sentence is to be construed as an attempt at or a disguised version of a statement, and its truth consists in its "picturing" the facts in the iconic mode

made precise by Tarski. This conception was first and influentially stated in a bald form by Wittgenstein in the *Tractatus* in 1922. Nowadays quite a few people are busy trying to create a formal system which can be plausibly claimed to "represent English" in this way - Montague was perhaps the most forthright, but there are plenty of others. Chomskyan linguistics provided a great boost for such endeavours in recent decades.

The rise of model theoretic semantics, a largely Polish achievement of the 1930s with Tarski as leader, provided a "model" for such a philosophy of language which has grown steadily more predominant despite the continued failure of its advocates to meet a large number of serious objections. Baker and Hacker have recently assembled many of these⁴⁸ and it is interesting to observe how Lycan's book⁴⁹, published in the same year, 1984, addresses none of the philosophical issues raised by Baker and Hacker, essentially trying only to show how the state of the logico-formalist art copes with various apparent limits on the scope of such a theory. Yet this "coping" consists in so many concessions, restrictions on the scope and aims of the enterprise, and endless complications to the necessary apparatus that it is tempting to agree with Baker and Hacker (and Wittgenstein of course) that such stuff is produced by people "held captive by a picture". But this more general debate is not my present concern, since it is only the inadequacy of the standard semiotic for philosophy of mathematics which is relevant, and which will be argued in Part II.

One reason for the resilience of this point of view is that the intended scope is unclear: some writers make no explicit claim that the logico-formalist characterisation can be extended to all language, some expect that some natural language is outside its scope but probably unimportant, some claim that it can indeed be made universal. Russell and the early Wittgenstein certainly fell into the third category. Later writers like Quine are impressed by the possibility of

⁴⁸ In Baker and Hacker [1984a]

⁴⁹ Lycan [1984]

perspicuous re-expression in the "canonical notation" of *Principia Mathematica*, taking Russell's analysis of descriptions as successful and paradigmatic, and dismiss recalcitrant aspects of natural language as scientifically irrelevant. But most writers on philosophy of mathematics simply assume either that mathematics can be treated in isolation, or that what works for it, *must* be universalisable. Benacerraf's "Mathematical truth" is a rather inchoate reflection on the difficulties of this conception.

The leading ideas about signification which all this has made part of the conventional wisdom in philosophy of mathematics are these. Reference and truth are the basic semantic notions; language is basically fact-stating. Other semantic features and pragmatic uses are to be explained in terms of these basic ones. Natural languages approximate to the strict logico-formalist account, which really only applies to artificial "languages"; but these are assumed to be real languages. Other forms of signification are approximations to language. Every language has a recursively specifiable syntax which can be revealed by correct "analysis" of the actual surface form of its signs. All languages, and so all sign-systems, are intertranslatable. There is no essential difference between words and notation, so proofs don't really need both words and notation. Re-expression in a "canonical notation" will preserve the essential logical form of a sign and eliminate misleading surface features.

§19 Will nothing work?

In 1979, Putnam despairingly declared⁵⁰ that in the philosophy of mathematics "nothing works". But his survey of the available positions omits the one which is to be found in the Nachlass of Wittgenstein, an omission consistent with significant changes between the first and second editions of his and Benacerraf's collection of readings. The positions which he does consider are unified by their implicit subscription to the logico-formalist hegemony. But the

⁵⁰ Putnam [1979: 386]

writings of Wittgenstein, and some others which have emerged in the most recent period, to some extent escape the dominant assumptions, and so they do provide more satisfactory answers to the problems of philosophy of mathematics, as I shall show later.

Putnam's article surveys the positions which he regards as on offer, and remarks that each of them has lethal difficulties. To those of logicism, formalism and intuitionism which I have already mentioned, he adds *logical positivism*, which he interprets as the view that mathematical statements are true because they are conventional rules of language. He gives few references, but it is not unreasonable to suppose Carnap is a major exponent of this view⁵¹. It is commonly supposed, as by Putnam here, that Quine's [1936] refutes this view, but Mortensen [1988] and the discussion of Wittgenstein's conventionalism in which I join in Part III argue not. But conventionalism about mathematical statements can hardly be called a philosophy of mathematics without considerable expansion. Another view which Putnam sadly admits to fail is Quine's *holism*⁵², which interprets mathematics as simply part of the scientific enterprise, postulating, tested and justified corporately. Putnam argues that this cannot be accepted since the aim of that enterprise, chez Quine, is to explain our sensory stimuli: mathematics just doesn't. One way out here would be to abandon Quine's behaviourism-influenced notion of science. But Putnam rather considers his own amendment⁵³ to Quinean holism, what he calls *quasi-empirical realism*, in which mathematical facts ("combinatorial" facts, such as $1+2=2+1$, that is) are added to the data of science (which for Quine is only sensations) and mathematical intuitions are added to the constraints of simplicity and conservatism in theory change. This won't do, he admits, because it provides no sort of account beyond arbitrary convention of why intuitively acceptable theories should be true. (One might also look for more detail as to just what sort of thing these "facts" are.) The only other view which Putnam

⁵¹ Cf Carnap [1937]

⁵² Cf Quine [1951] and [1962].

⁵³ Putnam [1975b]

can find to consider is what he calls *modalism*, a view attributable to Parsons, though Putnam himself has also espoused a version of it⁵⁴. On this view mathematics is the study of logically possible structures; but such a view has just as bad, really just the same, epistemological problems as does platonism. (The ontological status of possible worlds would need to be clarified before this kind of view could be seen as superior to platonism.)

There appears to be a certain parochialism in Putnam's list of possible philosophies of mathematics. As late as 1979 older philosophers like Putnam were still largely under the sway of the logico-formalist hegemony. The strange idea became current as part of the logico-formalist hegemony that a perspicuous representation of mathematics would provide ipso facto the correct philosophy of mathematics. So when all the reconstruction programs were shown to be hopeless by Gödel, logico-formalist writers had nowhere to turn for a philosophy. That is Putnam's problem. To solve it requires the repudiation of the logico-formalist hegemony. Putnam's unwillingness to do so is shown by his repeating Kreisel's inaccurate and deprecatory dismissal of the later Wittgenstein as a "strict finitist" revision of logical positivism⁵⁵. A similar refusal can be found in Benacerraf's "Mathematical truth" where⁵⁶ he remarks that to reject a referential characterisation of "mathematese", as he calls it, which would ease the task of epistemology for mathematics, requires a conscious rejection of the "standard" interpretation of mathematical language (that of the logico-formalist hegemony which I have sketched above) and provision of an alternative. He says that no-one has done this, but there I think that he is mistaken: Wittgenstein's later philosophy is the beginning of such an alternative semantics. But to one in the sway of the logico-formalist hegemony, it cannot be any use because it is *not formal*. Putnam and Benacerraf confirm this interpretation of their attitude by their

⁵⁴ Parsons [1977] ; Putnam [1967a]. Field has also bolstered his new nominalism [1980] with a view of this kind, [1984a].

⁵⁵ Kreisel [1958 *passim*] ; Putnam [1979: 387 and 396].

⁵⁶ Benacerraf [1973: especially 410-1]

omission of the Wittgenstein material from the second edition of their book of readings⁵⁷. But Wittgenstein's later ideas⁵⁸ are important because they are in large measure a critique of the logico-formalist hegemony, being a multiform and complex attack on the system of ideas about language and mind which he had earlier expressed in the *Tractatus*⁵⁹, which is a very influential statement of the logico-formalist account of signification.

Some other writers have escaped the logico-formalist hegemony in various ways in the last quarter of a century and it is to these, insofar as they have repudiated it, that I shall be looking for support in seeking satisfactory answers to the problems of philosophy of mathematics. Wilder, Kitcher and Lakatos are such writers whom I shall consider in some detail⁶⁰ in Chapter 10. Some writing on philosophy of mathematics which was available at the time of Putnam's paper, but which avoids one or other of the presumptions of the logico-formalist hegemony has been collected by Tymoczko⁶¹. Apart from Wilder and Kitcher, this mainly includes pieces which challenge developed theses rather than main presuppositions of the logico-formalist hegemony. The theme of the collection is the temporal development of mathematics, and this is certainly an aspect ignored by logico-formalism. The idea of mathematics as a social process will be explored sympathetically in Chapter 10 below.

Since Putnam's "report" there has been more and more published discussion of philosophy of mathematics, which becomes harder and harder to classify. Perhaps the two main developments still firmly in the logico-formalism tradition

⁵⁷ Their given reasons (lack of space, availability elsewhere) seem inadequate for the complete extirpation of Wittgenstein's ideas unless they regard them as unimportant. Does all the material on "the iterative concept of set" justifiably supplant the Wittgenstein material?

⁵⁸ Wittgenstein [1935], [1956], [1974], [1976], [1978], [1979]. See Part III for discussion.

⁵⁹ Wittgenstein [1922]

⁶⁰ Wilder [1973], [1981]; Kitcher [1983]; Lakatos [1974], [1978a]

⁶¹ Tymoczko [1986]

are the *structuralist* position⁶², which is a kind of Aristotelean response to Benacerraf's [1967] difficulty for platonism about numbers as sets, and Field's novel *nominalism*⁶³. Some more recent writing on mathematics dips into what the logico-formalist hegemony would dismiss as psychologism, for example Miller's and Tragesser's elucidations of Husserl's views⁶⁴, to which Tieszen⁶⁵ on intuition is also relevant. There is something mathematically very significant going on in constructive mathematics in the last two decades⁶⁶, but it has not produced much philosophical work as yet. Very recently, Shanker has written a book about Wittgenstein's philosophy of mathematics which is *not* mainly interested in the philosophy of language, so the discussion of Wittgenstein's philosophy of mathematics may be about to mature. I shall discuss it in Part III.

Having outlined the main problems of the philosophy of mathematics, and argued that they have been subordinated to the foundations problem by the logico-formalist hegemony which has dominated most writing about mathematics this century, let us consider what conceptions of notation are presupposed, since they are not set forth, in philosophy of mathematics. The hegemony derives from the logico-formalist analysis of mathematical language and its particular conceptions of language and notation and their relationship, and I will be arguing that the logico-formalist conceptions are in fact misconceptions.

§20 Conceptions of the role of notation in mathematics

The various ideas about the importance of good notation which I assembled in the previous chapter are rarely put forward as essential parts of a philosophy of mathematics. Rather, these ideas about notation lie behind the particular forms

⁶² Jubien [1977]; Resnik [1981], [1982]; Shapiro [1983]. Putnam [1967], Benacerraf [1973], and Steiner [1975] foreshadow the explicit statement of this view. Kitcher's [1983] contains a related view - see Chapter 10 below.

⁶³ Field [1980]. See also Field [1982] and [1984]. Field's view, and its dependence on some logico-formalist assumptions, is discussed in Chapter 7 below.

⁶⁴ Miller [1982], Tragesser [1984]. The latter continues the argument of his [1977].

⁶⁵ Tieszen [1984].

⁶⁶ See Beeson [1985] and the literature he cites.

which philosophy of mathematics takes. Most of the writers whom I quoted show an appreciation of most of these eight effects, but since views on the role of notation in mathematics are rarely very articulated, their relative significance is not grasped. In fact, there is a general and fundamental equivocation involved to which I have already adverted. Is notation actually *necessary* for mathematics? If it is, it must be because notation allows possibilities which are impossible without it. New things must be possible, not just a new facility in doing old things. We can solve problems or make things which we couldn't at all before - not just more efficiently. Now opinions differ on whether we are able to do new things in mathematics. If a reductionist attitude is taken, according to which we are not "really" doing new things (it can all be reduced to strokesymbols), then the role of notation cannot be anything essential beyond what language already affords us. On the other hand, if there is genuinely novel possibility in mathematics and notation is its vehicle, then notation is not optional. The question then is, do we really need ordinary language? The predominant opinion in this century seems to have been that notation is necessary, but that "in principle" we can dispense with words, indeed that notation is all we need. A related debate, as to the relative needs for words and "symbols" in geometry, is reported by Cajori⁶⁷ as current among geometers in the seventeenth century, though there it was more clearly a matter of taste and emphasis; writers allowed some kind of mixture to be needed but they disagreed about the best proportions. More extreme positions have been common recently, because the extension of notation to logic seems to make every word eliminable from text which can be put into notation. Of course it does not follow that because each word is eliminable individually, that all are collectively; but it is the conclusion usually drawn.

So, there are basically three general ideas about the place of notation in mathematics. There is the **dispensability view** that it is tremendously useful, but a mere auxiliary. We do not do anything with notation that we couldn't do without it.

⁶⁷ Cajori [1929: §§385ff]

Such views focus mainly on the abbreviatory aspects of notation. Secondly there is the **replacement view** that because notation does allow us to do new things, which ordinary language does not, that notation must replace ordinary language for our scientific purposes at least. This latter view is the dominant view, its dominance deriving from the logico-formalist hegemony, which centres on the iconic nature of notation and its consequences. Thirdly there is the **extension view** that notation supplements, in an essentially symbiotic way, ordinary language. Most writers in the logico-formalist tradition vacillate between the dispensability and replacement views. I shall be proposing an extension account. In the rest of this section I will describe these views further.

Dispensability views

The dispensability idea is that notation is not necessary for mathematics, that it could all be eliminated (perhaps "in principle"). The basis for this view is that the essence of notation is shown in the introduction of new signs in definitions, that notation abbreviates what could be expressed in longer signs, so that if definitions were all eliminated in turn we would have mathematics expressed in words alone, perhaps somewhat cumbrously. I shall consider here two versions of this viewpoint, the Intuitionist conception of mathematics, and the Logician point of view on definition.

The most radical dispensability view is that *notation is dispensable in favour of thought*. According to this view language, indeed any kind of expression, is not really necessary for mathematics, because it is a languageless activity of pure thought. This is the viewpoint of classical Intuitionism as stated by Brouwer and Heyting, though more honoured in the breach than the observance by third generation "intuitionists" like Troelstra. The role of notation is merely part of that auxiliary function of language, to communicate mathematics among mathematicians. I will argue in Chapter 3 that this idea is quite mistaken, because mathematics **needs** expression.

The viewpoint of logicism as launched by Frege and Russell is more ambivalent. On the one hand, notation is needed to guarantee the inferences; on the other, notation is merely convenient abbreviation and could always be dispensed with. Indeed Frege devoted a whole book to an exposition of his main argument without any use of his Begriffsschrift (yet he also held that it needed expression in that form too for precision and certainty). What this view would be if, made quite explicit, is that *notation may be dispensed with in favour of written words*, but all its adherents also hold the very widespread view that *written words are dispensable in favour of speech anyway*. This very basic presumption about language is one which I shall dispute precisely because it underpins mistaken ideas about notation.

Replacement views

The replacement idea about notation is that notation is necessary for mathematics, but ordinary language is not; a suitable notation can be found for the purposes of mathematics, perhaps for science, using which all ordinary language can be eliminated. What underlies this point of view is the idea that notation can allow us to improve on the inferences of ordinary language through its iconism. Notation is thought of as a system, as a language, indeed as the ideal language desired by Leibniz. This is the most common conception of notation. Many logicians held this view (along with the dispensability view)⁶⁸, and many formalist mathematicians too (but not, I think, Hilbert, and definitely not Curry). Many philosophers of language have adopted a view of this kind in the hunt for a "theory of meaning".

In the presentations of mathematics of Frege and of Whitehead and Russell, there is a deliberate effort to separate the introductory explanations from the "mathematics proper", and to make it clear that the former is not a part of the latter. I shall discuss the significance of this device in Chapter 7. For the present I

⁶⁸ Russell explicitly says in his paper on vagueness [1923:84] that his *Principia Mathematica* notation is a *language* in which vagueness is avoided.

wish only to point it out, and the great success it has had. For the common wisdom is that the "real" mathematics is the "symbolic" part, the rest is mere rhetorical help for dull readers which "in principle" is unnecessary. Unlike Russell, Frege did not, at least in the major part of his career, claim that his notation was a real language; but there have been plenty to do it for him, and I will discuss some of them in Chapter 7. The logicist successors of Frege and Russell abandoned any pretence that notation was mere re-expression and openly adopted a replacement approach⁶⁹.

Many versions of Formalism adopt a replacement viewpoint, based on the assumption that the endeavours of the Logicist and metamathematical schools have provided us with a complete and correct account of what notation is, in the concept of a formal system. But actually it is a very widespread assumption that signs such as words and diagrams are "in principle" dispensible from scientific discourse, and ideally should be. Curry explicitly declares that mathematics is the science of formal systems, and many have agreed taking this to mean the science done "in principle" in formal systems; though Curry did not mean that, as his opening chapter make clear in Curry [1963]. The replacement idea of notation also underlies programs which rigorously separate object language (notation) and metalanguage (words etc), and attempt to complicate the object language sufficiently for it to cope with everything science needs.

In Chapter 6 I shall attack the idea that diagrams are unnecessary for mathematics, and a similar argument will be given in Chapter 8 for other elements deemed inessential by logico-formalists. The idea that notation can be used without ordinary language will be attacked in Chapter 7.

Extension views

The extension idea about notation is that notation is necessary for mathematics, but so is ordinary language. Notation is needed in addition to ordinary language because it makes it possible for us to say and know new kinds of

⁶⁹ As in Quine [1960], Field [1980]

things. Our concepts are "stretched" by work with notation and we come to be able to see new things.

Views of this kind are less common than dispensability and replacement views. The distinction between object-language and meta-language could be the basis for such a view, though often it is not. In the case of Curry, it certainly is⁷⁰, and Hilbert's position seems to me to be similarly extensionist, though less clearly so since his philosophy is not very explicit. Hilbert's formalism is a more complex view than many take formalism to be, and needs to be distinguished from the parodies of it offered by critics. For Hilbert, mathematical activity may be real or ideal. Notation is what permits the evident truth of real mathematics to be conserved in ideal mathematics. It does so because it is diagrammatic and expresses our intuitions of space. Curry's viewpoint has been well presented recently by Meyer⁷¹. Wittgenstein's is another such view. Wittgenstein construes mathematics as a congeries of ways to give rules for the use of mathematical terms in practice. He interprets proofs as transformations of designs which show us techniques to govern our use of terms in non-mathematical language. I shall discuss these views in Chapter 12. In Chapter 10 I shall discuss several process-oriented philosophies of mathematics, which permit the dynamic interplay between words and notation to be observed to the advantage of our understanding of both.

§21 Intentions and methods of the investigation to follow

The view which I shall develop in this thesis is an extension view of notation, treating notational innovations as additions to our semiotic resources. The fundamental advance was the invention of writing which in itself changed the kind of knowledge we could have; mathematics has grown to its present extent and variety on the basis of a refined technique of signifying expression which notation has built up in many small steps over several millenia. The perspective on

⁷⁰ Cf his discussions of the U language [1963: 28ff].

⁷¹ Meyer [1987]

mathematics I shall develop here is an inclusive, and so eclectic one. What Hilbert and Wittgenstein have in common is an appreciation of the *diagrammatic* element in notation. The strength of Wittgenstein's philosophy of mathematics is that he sees that mathematics is embedded in the social activity of language, the strength of Hilbert's, that he holds fast to the phenomenological facts of the use of notation. Where both are lacking is in overlooking first, the written nature of mathematics, and second, the scientific and power contexts of the use of mathematical texts. Lakatos and Kitcher focus on mathematics itself as a process, not idealised reconstructions (or plans for them). However I do not think that any of these more congenial and realistic philosophies of mathematics is acceptable and I shall suggest that this is to some extent⁷² due to an insufficient appreciation, even for these writers, of the nature and role of notation. In Part II, I shall try to refute views of the eliminative and replacement kinds, and I shall try to find ways to improve views of the third kind to give a more adequate conceptualisation of mathematics, which I shall present in Part III through a discussion of the seven problems identified in Chapter 1.

I do not intend to discuss philosophical method at length, but I wish to point out a few working assumptions which I am making. First and foremost, I accept Wittgenstein's demonstration that our concepts are not to be analysed into necessary and sufficient conditions; rather, they have the character of family resemblance. I shall not be arguing for this thesis, I take it that it is at least a respectable minority position in the philosophy of language, though it has few adherents in the philosophy of mathematics. Accepting the force of his arguments implies that all concepts, in particular those of knowledge, proof, definition, even mathematics must be considered as more or less like 'game'. There are *no* terms of ordinary language which have necessary and sufficient conditions; one of the bad

⁷² The other reason is a certain *political* innocence which is endemic in philosophy of mathematics. I shall only be able to slightly sully it here.

effects of mathematics as a paradigm of knowledge, I believe, has been the illusion that such definitions, because they occur within mathematics, must be aped outside it, in particular in discussing it. Conceptual *analysis* is a will-o'-the-wisp, but conceptual *clarification* may help us to a more perspicuous overview of a human activity such as mathematics.

A second methodological point is that I am wary of idealisation. Many poor philosophies have been produced with the appellation "scientific", and this will not be claimed here. But certain kinds of simplification common in philosophy are often justified on the ground that idealisation is necessary in science. An example which I shall discuss in Chapter 4 is Searle's theory of speech acts, in which many complications are brushed aside explicitly on this ground. I do not accept this move as a legitimate one in philosophy. Another such simplification which I shall eschew is one which philosophers of mathematics seem to have picked up from linguistics: the use of invented single sentences as data for explanation. For example, Benacerraf's paper on mathematical truth discusses only one example, a sentence he has clearly made up on the spot. In this thesis I shall look at sample texts from **actual mathematics** - hence the exhibits. This will also help to combat the tendency which the logico-formalist hegemony encourages of sweeping applied mathematics under the carpet as not part of the real data of philosophy of mathematics, but a sort of volume 2 problem to be solved separately and later. A related point which will recur frequently is that I mostly reject appeals to "in principle" as hiding difficulties under the carpet of "scientific idealisation".

I have already indicated by choice of writers for discussion, that I shall be following the current trend to focus on the *process* of mathematics as well as the *product*, a fixation on the latter being one of the hallmarks of the logico-formalist hegemony. It is not easy to focus on mathematical process; studies of the process of science are relatively new and there are special difficulties in attempting any

empirical study of mathematics⁷³ which perhaps explain its neglect by philosophers of science and mathematics.

These last two points underlie a common refrain in what follows: I believe the philosophy of mathematics must address the actual mathematics which is so important in modern life, an activity of real human beings with distinctive capacities and limitations. I am very wary of any argument which makes a claim about mathematical knowledge which abstracts "in principle" from the nature of those mathematical knowers with which we are acquainted, namely ourselves, and invokes hypothetical cognisers not limited and driven in the ways that we are.

§22 Recapitulation of the problem and summary of the argument to follow.

In this Part I, I have pointed out the kind of puzzle which the cognitive peculiarities of mathematics give rise to, and suggested reasons for thinking that some of the properties of notation may be responsible. I have distinguished a number of these effects of notation, and shown through a discussion of the literature that the literature of philosophy of mathematics has become dominated in this century by a certain conception of mathematical signification related to the use of notation in mathematical logic. The aim of the rest of the thesis is to cast some light on the problems of philosophy of mathematics which I have outlined by a detailed investigation of the actual role notation in real mathematics (not reconstructions like *Principia Mathematica*).

In Part II, I will give an account of notation as one element in the system of mathematical signification. In Part III, I will describe the conception of mathematics to which this account of notation leads.

⁷³ I am thinking of parallels for mathematics, of the kind of work on science done recently for example by Latour and Woolgar [1986] or Mulkay [1985] (cf references there). The difficulties include the extreme individual privacy of much mathematical work. But see also Chapter 11 on creation of mathematics.

In Part II, I begin in Chapter 3 by attacking the Intuitionist conception of mathematics as a languageless activity of pure thought. I argue that, on the contrary, for there to be mathematics there must be expression. Specifically I show that signs for intuitions, constructions, theorems and proofs are all necessary. I also comment on some difficulties for Intuitionism raised by Dummett and Tait. The conclusion of this chapter is that mathematics **requires** expression.

In Chapter 4, I examine a sample of mathematical expression, a text chosen for its brevity. I find in it four signsystems at work, **Written Word, Diagram, Notation** and **Paragraphy**. I point out that the different kinds signs in the text have different functions. Seeking an adequate conceptualisation of this variety of signsystems, I consider but reject the account of sign-systems given by Goodman in his book *Languages of Art*. My main objection to Goodman's account is that it conflates all sign-functions with naming.

In the next four chapters I give my own accounts of the four signsystems of mathematics, interleaved with criticisms of some logico-formalist views. I begin in Chapter 5 by showing that **Writing** is a different signsystem to speech by presenting the concept of **text act**, which I derive from a critique of the theory of speech acts. I also introduce the idea of **text effect**. I argue that some needed mathematical activities such as definition and proof are text acts, whose significance depends on text effects. I illustrate these claims by discussing some aspects of definition.

In Chapter 6, I go on to an analysis of the **Diagram** system. I review the great variety of diagrams in mathematics and some ways they have been classified. Then I propose a functional classification of diagrams. I identify five important functions which diagrams have in mathematics which can be understood in terms of their functional articulation. I argue that diagrams are **grounded icons**, and that the semantics of some diagram elements have no parallel in speech. On the basis of my analysis I argue that the logicist rejection of diagrams from mathematics is mistaken. Diagrams are needed to pose problems and motivate methods; moreover

the logicist idea that intuition is not needed in mathematics is not demonstrated by the removal of diagrams.

In Chapter 7 I turn to the consideration of the *Notation* system itself and, while accepting the logico-formalist account in some sense, find it inadequate in its narrowness, that is, in its assumption that in notation we merely re-express verbal mathematics. I argue that on the contrary, mathematics could never be prosecuted in a notation alone. A notation, even a formal system, is not a language and cannot be one, because languages need grounding - which is incompatible with pure formality. To illustrate this argument, I consider the problem of communication with extraterrestrials and the question whether and how an invented "language" like LINCOS could work for this purpose. I go on to consider the ideal language view which was certainly adopted by the early Wittgenstein and middle Russell, but has been erroneously foisted onto Frege. I rehearse Frege's reasons for taking an extension, rather than a replacement, view about notation. I argue finally that Quinean re-expressionism of the kind relied on by Field's program assumes an erroneous replacement conception of notation.

In Chapter 8, I give an analysis of *Paragraphy* and its functions. I point out that many signs to be found in mathematical text are not part of the matter under discussion but have the role of articulation of the text - pagenumbers, titles and so on. After making some distinctions among these signs, I turn to justifying my calling them collectively a sign-system. Then I go on to argue that there are two main aspects of text articulation, the stratification of the text and its integration into its context. These functions are actually essential for text as a distinct means of communication. The most crucial function of this neglected element in mathematical signification is its role in the constitution of the *archive*. This is discussed further in the next chapter.

In Chapter 9, I describe the kinds of *context* in which mathematical processes go on. There are three dimensions of this context. First, the textual archive which has been built up by past mathematical processes. The role of

paragraphy is to make the use of such an archive, and therefore its existence, possible. Second, the mathematically formed environment within which mathematical practitioners live. Third, the mathematically educated community of which the individual mathematical actor is a member. I illustrate the contributions of these dimensions of context to the understanding of a typical mathematical text.

In Chapter 10, I consider how well the practice-oriented philosophies of mathematics of Wilder, Lakatos, and Kitcher take account of this context. I argue that Wilder tries unsuccessfully to ignore philosophical issues, while Lakatos and Kitcher are too respectful of classical logic, not recognising its textual basis, and too insouciant about the power context of mathematics. All three assume the logico-formalist account of language. I argue that the element missing from an account of practice which can be obtained from their several accounts is that of *mathematical work*. I suggest *mathematical praxis* as the term for my own conception.

In Part III, I return to the problems identified in Part I as the main concerns of philosophy of mathematics and attempt to say something illuminating about each of them from the perspective which I have been developing in Part II.

In Chapter 11, I argue that a praxiology of mathematics must recognise not only its textual basis but its textualised context. Both application and creation of mathematics proceed through text acts of various kinds in which the specific features of the mathematical sign-system play crucial roles. Most application involves measurement, and I show that this has pragmatic presuppositions which are illuminated by my analysis of the context of mathematical praxis. I defend Wittgenstein's interpretation of the simplest kind of application of arithmetic with arguments drawing on my analysis of context and praxis which fill out his ideas.

In Chapter 12, I argue that the logical and epistemological problems can only be understood if mathematics is conceived as a panoply of text effects. Mathematics is constituted textually. Since the kind of knowledge we have in mathematics is text effects, as I argue, its objects must be accepted as of a kind that

can only be known through signs - textual signs at that. The object problem is an artefact of a lingering essentialism applied to notation. The "problem" of infinity is equally illusory. But an account of the specific character of mathematical knowledge, its exactness abstractness and certain truth, can be given by exhibiting definition and proof as text acts, which I do.

Part II

Mathematical text, context and praxis

Chapter 3

Expression

§23 Bleak Intuitionism as a view about notation

Brouwer's Intuitionism, which I distinguish from the views of later intuitionists and other constructivists, I call bleak because it must make fairly heroic dismissals of what is ordinarily understood by mathematics. Not merely those parts of the *content* which Intuitionism is usually criticised for rejecting, but also most aspects of the *reality* of mathematics which make it a social force, a tradition of knowledge, something other than a solipsist amusement. Intuitionism does not attempt to provide an account of applied mathematics and this is already a damning indictment. I shall argue in this chapter that its account of "pure" mathematics is inadequate too. The bleak Intuitionist view is that notation is unnecessary for mathematics. It is unnecessary because language, indeed all expression, is unnecessary for mathematics, and the role of notation is *merely* to improve the reliability of language in communication between mathematicians. Such communication of results is not necessary for mathematics, because mathematics is *essentially* a languageless activity of constructions in pure thought. For bleak Intuitionism, notation is dispensable in favour of thought. This is an extreme kind of Dispensability view about notation. The aim of this chapter is to refute this Intuitionist thesis; I don't consider here other aspects of Intuitionism.

I shall argue in this chapter that mathematics cannot be essentially pure thought, mental constructions in intuition, as Brouwer has it, but must have public expression. I give four arguments against the view that written or spoken proofs merely report mental constructions, and are not essential for mathematical activity¹. First, signs for "basic intuitions" are not merely convenient but necessary in mathematics. Second, signs for defined items are necessary - the Intuitionist view makes the possibility of definition difficult to understand. Third, signs for relations of constructions are necessary for theorems to be possible. Fourth, signs for proofs must be used which can be surveyed and checked, as mental

¹ For the sake of argument, in this chapter I restrict my argument to "pure" mathematical activity. Not that I accept such a limitation as making much sense, as later chapters will make clear.

constructions can not. I comment on a group of arguments put by Dummett against the Intuitionist conception of proof, and I consider an argument of Tait's against the languagelessness thesis. I suggest that he makes a false assumption which I shall reject in the following chapters. What this chapter shows is that mathematics must be expressed. *How* it is expressed is the subject of the rest of Part II.

§24 Four Intuitionist theses about constructions, language and logic

In this section I shall formulate bleak Intuitionism, as it concerns me here, in four theses. The basic question which identifies a constructive position is this: under what conditions is a mathematical item given to us? The constructivist answer is: when we have constructed an instance. But this could be taken in various ways, for example as 'How is a mathematical item introduced?' or 'When do we have access to an item?' or even 'When may we introduce a name for an item?' Dummett² prefers the formulation 'When is an expression involving a mathematical term meaningful?' But these last two questions beg the question against Intuitionism, since it asserts that mathematics is an essentially languageless activity. So let us remain with the original question for the present: under what conditions is a mathematical item given to us?

As the name suggests, Intuitionism begins from the concept of *intuition*³. This term is hardly explained, but refers to something like the Kantian "form of inner sense". So the starting point for the Intuitionist philosophy is the fact that mathematics is an activity of thought, a process. In particular, Intuitionism claims that there is a primordial intuition of the "bare two-oneness" on which all of mathematics is *constructed* by processes of repetition and convolution. Thus mathematics consists of *constructions in intuition*, that is, private, purely internal acts of the mind, independent of perception and other empirical input. In his

² Dummett [1975a]

³ The concept of intuition used by Gödel and discussed by Maddy, Parsons and others is not quite the same; but see Tieszen [1984] for some connections. I don't have space to go into this interesting question here.

inaugural address⁴, Brouwer presents intuitionism as a refurbished form of the Kantian philosophy of mathematics. He interprets the developments in geometry in the nineteenth century as gradually making Kant's position untenable; he proposes that since geometry can be reduced to arithmetic, a Kantian position can be regained by interpreting mathematics as concerned only with the intuition of inner sense, that is, time. This seems untrue; but I shall later⁵ give arguments for something like the related Kantian thesis that mathematics is synthetic a priori.

For Brouwer in [1912] the elements of mathematics were four⁶. They comprise [1] "the basal intuition of the bare two-oneness"; the mental actions of [2] "create a finite ordinal number", and [3] "create omega"; and [4] an intuition of the continuum. Brouwer restated this to some extent at different times, though he never gave much detail about the phenomenology of Intuitionist mathematics. He assumes, but nowhere explains, that it is possible to put these elements together in various ways in constructions. It will be useful to quote in full a late formulation of his⁷.

"FIRST ACT OF INTUITIONISM

Completely separating mathematics from mathematical language and hence from the phenomena of language described by theoretical logic, recognising that intuitionistic mathematics is an essentially languageless activity of the mind having its origin in the perception of a move of time. This perception of a move of time may be described as the falling apart of a life moment into two distinct things, one of which gives way to the other but is retained by memory. If the twofold thus born is divested of all quality, it passes into the empty form of the common substratum of all twofolds. And it is this substratum, this empty form, which is the basic intuition of mathematics. Inner experience reveals how, by unlimited unfolding of the basic intuition, much of 'separable' mathematics can be rebuilt in a suitably modified form."

Thus we have Intuitionist thesis number 1:

[IT1] A mathematical item is given to me when I have constructed it in my mind.

Mathematical entities are mental constructions from basic intuitions.

⁴ Brouwer [1912]

⁵ In Chapter 6 (synthetic) and Chapter 12 (a priori).

⁶ Somewhat later he added a further element in the theory of the "creative subject." I shall not discuss the latter (even Intuitionistically controversial) extension at all.

⁷ Brouwer [1982:4-5]

Brouwer asserts that mathematics is an essentially languageless activity, and the second and third generation Intuitionists, principally Heyting and Troelstra respectively⁸, have echoed this claim, though without much further explanation, and with some significant qualifications ('Language is not essential to the *idealised* mathematician" - Troelstra, my italics.) Language plays a purely auxiliary role in mathematics, for the purposes of aiding the memory of the individual mathematician, and for the mutual reconciliation of the results of a community of mathematicians. For Intuitionism, any expression of a piece of mathematics in language is subject to error and imprecision by the nature of the case. Heyting says⁹ that language does not *represent* mathematics. I shall discuss this thesis at length in the later part of this chapter. I shall argue that it is false but in an interesting way. So we have

[IT2] Mathematics does not require language.

Nevertheless, Intuitionists do use language. When they do, mathematical assertions must be properly understood. For the Intuitionist the truth of a mathematical proposition requires its being correctly formulable¹⁰. Thus we have

[IT3] The form of a proper mathematical assertion is "I have effected a certain construction in my mind" .

As such it is clearly tensed, indexical, first-personal and about a mental content.

Intuitionism holds that mathematics founds logic, not vice versa. In particular, mathematical experience shows that the classical law of the excluded middle (LEM) cannot be applied unrestrictedly in mathematics - though it has a role. Brouwer asserts that the use of the formalist's axioms, in particular LEM, on infinite sets *inevitably*¹¹ leads to contradictions. So he declares that, although the intuitionist is happy to agree with the formalist that certain reasonings are valid, some kinds of formulation of the *results* of such reasonings are "without

⁸ Heyting [1931: 52f], [1980: 4-5 etc], Troelstra [1969: 4]

⁹ Heyting [1931: 53]

¹⁰ Heyting [1980: 18-19, 103]

¹¹ Cf Brouwer [1912: 83]

meaning". Examples given by Brouwer include the essence of Cantor's argument that no 1:1 mapping can be made from the integers to the reals between 0 and 1. What Brouwer rejects is a positive formulation of such an impossibility result, such as that "the power of the set of reals exceeds that of the integers". Results with such contentious formulations all consist in, or at least presuppose, a claim that a certain mathematical entity such as "the power set of the set of reals" exists without an explicit indication of how it is to be constructed. A proof leading to such an assertion will be a proof of a disjunction which establishes neither of the disjuncts, or else a proof of an existential thesis which does not construct an actual case. Since the results of such reasonings are not available in the "meaningless" form which employs such a locution, any later arguments which take such a proposition as a premise are entirely without force for the intuitionist.

An example that is given by Dummett¹² is clearer than any of Brouwer's. This is the proof that there are solutions a, b, c to the equation $a^b = c$ with c rational and both a and b irrational. It proceeds by arguing from the alternatives that *either* $\sqrt{2}^{\sqrt{2}}$ is rational and *or else* $\sqrt{2}^{\sqrt{2}}$ is irrational; at present no one knows which of these holds. But if the first is true, then we can take $a = b = \sqrt{2}$, $c = \sqrt{2}^{\sqrt{2}}$ as a solution, since $\sqrt{2}$ is known to be irrational; while if the second, then we can take $a = \sqrt{2}^{\sqrt{2}}$ and $b = \sqrt{2}$ and $c = 2$ (because $(\sqrt{2}^{\sqrt{2}})^{\sqrt{2}} = (\sqrt{2})^2 = 2$). In this argument a formulation "A or not A" (A being " $\sqrt{2}^{\sqrt{2}}$ is rational"), with neither disjunct known, is used as a premise, and this is not acceptable to Brouwer. So the conclusion that the given equation "has solutions" is meaningless and not usable in further reasoning.

The logic of all this was made clearer later by Heyting, Kreisel and others¹³, and can be stated relatively simply in formal terms. But it is important not to beg the question against the Intuitionist by taking such a logic as basic or primary, as the nonIntuitionist is constantly tempted to do. This claim of Intuitionism, its

¹² Exhibit #3A, from Dummett [1977: 10]

¹³ Cf Heyting [1980: ch7]

insistence that mathematics does not need to be and cannot be "founded" on logic, that indeed the boot is on the other foot, logic being a branch of mathematics, is a congenial one to many mathematicians. That claim does not itself lead to Intuitionist logic, it was held by Peirce for example. Intuitionist logic requires a further idea: that the truth of a mathematical proposition is to be identified with the effecting of a construction. This permits the *discovery* that some propositions p are neither true nor false, since neither p nor $\sim p$ can be, or perhaps have not yet been, constructed. For the Intuitionist, logic can only be an aid to the use of language in communicating mathematics, which is an essentially derivative and secondary activity. The role of logic for mathematical statements must be to control the adequacy of such reports *in extenso*. It is **not** part of the content reported. The actual mathematics and the construction in the mind are for the intuitionist the same thing, and the mind has no use for logic in making its constructions anymore than Newtonian masses "obey" the laws of physics.

So

[IT4] The correct logic for mathematical assertions, I, must be derived from mathematical experience understood Intuitionistically as in [IT1-3]; Intuitionist logic, I, turns out to be different from classical logic, C, principally in lacking LEM.

There is a positive queerness in Brouwer's Intuitionism, going beyond the negative critique of classical mathematics which it shares with other constructivisms. I mean its theory of the continuum, which relies on the highly nonclassical concept of *choice sequence* with the concomitant "set" theory which Intuitionism calls the theory of spreads and species. This voluntarist aspect of Intuitionism has its attractive side, but here I propose to ignore the distinctive mathematical features of Intuitionism entirely, and to consider only the Intuitionist interpretation of mathematics, its account of mathematical activity. For present purposes this is summed up in these four theses then:

[IT1] A mathematical item is given to us when we have constructed it in the mind. Mathematical entities are mental constructions from basic intuitions.

[IT2] Mathematics does not require language.

[IT3] The form of a proper mathematical assertion is "I have effected a certain construction in my mind" .

[IT4] The correct logic for mathematical assertions, I, must be derived from mathematical experience understood Intuitionistically as in [IT1-3]; Intuitionist logic, I, turns out to be different from classical logic, C, principally in lacking LEM.

In the next four sections I shall argue that the languagelessness thesis of Intuitionism is mistaken, by showing that there are at least four ways mathematics requires signs. Of course what I am really arguing against is an "expressionlessness" thesis. In subsequent chapters I shall discuss the relations of signs and language extensively. For the present I use the word '**sign**' simply to mean a *publically available communication token*.

§25 Signs for intuitions are required in mathematics

I begin by considering more closely just what a construction is. Let us take as an example the sentence (recall that I use '*' for multiplication)

[AS] $3*4=4*3$

which can be proved from Peano's axioms as follows:

$$\begin{aligned}
 3*4 &= (3*3) + 3 \\
 &= ((3*2)+3) + 3 \\
 &= (((3*1)+3)+3)+3 \\
 &= 12 \\
 4*3 &= (4*2) + 4 \\
 &= ((4*1) + 4) + 4 \\
 &= 12
 \end{aligned}$$

(The last inference in each case omits repeated applications of the definition of +, as the other steps unpack *. I omit them because their inclusion adds nothing to the example for our purposes.)

An Intuitionist proof of the proposition presumably reports the separate construction of the left and right-hand sides, and then the apprehension of their equality; its truth consists in the actual occurrence of the last step and only holds once that step is made. Unfortunately my proof seems to work by deconstruction! But perhaps I have presented it in way unfair to Intuitionism, and it should be turned around somehow. Yet although one can see that $3 \cdot 4$ and $4 \cdot 3$ are in some sense "built up" from simpler elements, the proof that they are the same consists in exhibiting them as each equal to 12 by analysis. Is this proof then unacceptable to Intuitionism? It would seem so from IT3. A problem here is the unsatisfactory state of the literature. Accounts of intuitionism skate over the "simplest" constructions which we need to discuss. The treatment of the continuum based on the theory of choice sequences is what most concerns Brouwer, and indeed in his late lectures, when he describes in the "second act of intuitionism", the method of introducing terms for species, he does not consider¹⁴ any need for the introduction of terms prior to these acts. Similarly, Heyting devotes a mere 3 pages¹⁵ to rational arithmetic, in his rush to get to the real number generators. The difficulty is that Intuitionists don't seriously discuss rational arithmetic because they regard rational arithmetic as unproblematic¹⁶, because it is part of a core of mathematics agreed to on all sides as obviously sound - as indeed it is. But the *philosophical* question is, does it conform to their interpretation of mathematics? It is hard to tell from his published works what a Brouwerian version of my proof would be like. Here we can see an example of the "problem of foundations" overshadowing real problems of philosophy of mathematics.

The various mental events involved would seem to be, in order of complexity,

[1] intuiting mathematical items such as 3

¹⁴ Brouwer [1981: 8]

¹⁵ Heyting [1980: 13-15]

¹⁶ Brouwer remarks in passing [1954: 524] "the first act of intuitionism enables us to construct the linear rational grid". This is further justification for my ignoring positive Intuitionist mathematics.

[2] constructing constructions such as $3 \cdot 4$

[3] constructing mathematical truths such as $3 \cdot 4 = 4 \cdot 3$

Let us look at them in turn.

There does not seem to have been much detailed philosophical consideration of what a construction actually is. Most of the literature tries to give suitable formal systems for the notion. Troelstra for example¹⁷ refers one to Nicholas Goodman's work, saying that the theory is "far from being in a finished state". But if we examine it, we find simply more formal systems, no elucidations of what constructions are. As Beeson reports¹⁸ Sundholm saying "What is needed does not seem to be further metamathematical studies, but conceptual analysis."

When I think the proposition [AS] to myself I find myself using the signs '3', '*' and so on. This is a version of the proposition which has been "overconstructed" for further use: the Intuitionist would perhaps not regard this kind of expression of my thought as accurate to the original. So what thought underlies it? And how are the two related? The item to which I give the name '3' is not itself a construction in Brouwer's stricter sense, though it is achieved by repetition of the basic act of dichotomy. It apparently consists of a primitive two-one-ness which is held in the attention and made itself into a 1, and immediately thereby accompanied by a second. The phenomenology is well reported by the set-theoretic signs $\{\{f\}\}$ or better $\{\{f\}, f\}$, but includes temporality as an essential mode.

Brouwer appears to claim that the achievement of such an intuition is itself a basic operation of the mind, as well as being explicable in the way I sketched above. I do not see how it is possible to claim this without admitting something equivalent to naming. Of course the "names" need not be "linguistic". No doubt for a given number such as 3 it is possible to "store a routine" for the production of a 3; some later forms of constructivism make this computing model explicit¹⁹. Brouwer does not make it clear whether when it "runs", the 3-construction is just like one we

¹⁷ Troelstra [1969: 10], Goodman ND [1971]

¹⁸ Beeson [1985: 408]

¹⁹ Cf Beeson [1985: passim]

might make from scratch. In the absence of an explanation of any abbreviating abilities of the mind, I assume for charity that it is the same. The only alternatives seem to be some extra kind of construction, which Brouwer has not mentioned, or the possibility that the name can be used in place of the construction, which Brouwer can hardly allow: that is the essence of formalism!

Tait²⁰ has also criticised this aspect of Brouwer's phenomenology, at least as expressed in a quotation from Brouwer [1952], which is essentially the same as the statement in [1981] which I quoted above - the "first act of Intuitionism" statement. He restricts his criticism to the Brouwerian account of the natural numbers. He has three arguments: first, he cannot find the "twoity", the "step of time" in his own consciousness; second, Brouwer seems to conflate an experience of 30 ticks with an experience *as* 30 ticks; and third, Brouwer's view seems not to be able to deal with large numbers like 10^{10} . I will briefly discuss each of these; only the third seems cogent to me.

First, I think we must distinguish between experiencing a "two-oneness", and its possible bareness. The first I have no problem with. It seems to me to be characteristic of experience that it does indeed as Brouwer says "fall apart" into before and after. Husserl and Heidegger²¹ provide enormously detailed descriptions of experience which confirm this Kantian claim. But what is not so clear is that there is any sense to such a dichotomy without some content or other. Yet perhaps all Brouwer needs is that in subsequent acts of the mind we may disregard whatever the content *was*. This suggestion is an essentially formalist one, and should therefore be unacceptable to Brouwer; but that would be on grounds of interpretation not phenomenology, so I think we should allow Brouwer his "twoity". Second, it is clear enough that to convert an experience of 30 ticks into an experience *as* 30 ticks requires counting, but I do not see why Brouwer can be claimed to have ruled that out. On the contrary, it seems to be perfectly consistent

²⁰ Tait [1983]

²¹ Husserl [1964], Heidegger [1962]

with his thought. It does raise again the question of tokens, that is, the question is what exactly are the mental items which are counted. Do they have specific content? If so, is it legitimate to ignore it? And are the counting numbers which are paired to those contents in counting to be thought of as old constructions being recycled? Brouwer gives little guidance in this, but let us assume that the answer is yes to each of these questions. Tait's third criticism, that Brouwer's view cannot account for our dealing with numbers like 10^{10} , is one which has been argued in another way by Dummett²². It is the argument that Intuitionism has no principled way to justify its intermediate position between strict finitism²³ and classical mathematics. The use of such items seems to conflict with the basic principle IT1. But even if this is so, it is still possible that such items are not essential to mathematics and the Intuitionist account for small numbers might suffice for a philosophy of mathematics. So we must look more closely at the intuitionist account of [AS].

The problem we have is the problem of how the mathematician is sure about the two '3' s being the same. In answer to just such a question about equality for natural numbers Heyting says²⁴:

*"Indeed this point needs some clarification; it forces me even to revise somewhat our definition of a natural number. If a natural number were nothing but the result of a mental construction, it would not subsist after the act of its construction and it would be impossible to compare it with another natural number constructed at another time and place. **It is clear that we cannot solve this problem if we cling to the idea that mathematics is purely mental.** In reality we fix a natural number, x say, by means of a material representation; to every entity in the construction of x we associate, e.g., a dot on paper. This enables us to compare by simple inspection natural numbers which were constructed at different times."*

This seems to me to be a decisive concession. The interlocutors of INT in Heyting's dialog throw in a few quibbles about 1:1 relations and the surprising complexity of these basic items, the positive integers, but they do not seize on the main point, which is this: INT has here rejected Brouwer's "first act" of intuitionism. He is saying that a natural number is *not* a mental construction, that

²² See Dummett [1975a: 121], also van Dantzig [1956]

²³ That is, a position like that of Esenin-Volpin [1970]. See Wright [1982] for an extended discussion of the question whether Intuitionism as interpreted by Dummett can honestly avoid this slide.

²⁴ Heyting [1980:15], my emphasis.

mathematics is *not* languageless - dots on paper are *writing!* He is admitting that a move of time is *not* the essence of mathematics - comparisons of dots are as much spatial as they are temporal. This volte-face cannot be escaped by appeal to the practicalities of human life, the difficulty about equality is one of principle. Dots on paper, or their equivalent, are admitted to be *necessary*, not simply convenient. And we haven't yet had any explanation about how names like '3' are introduced and intuitionistically justified! (Nor do we get any.) Notice also that what INT has conceded here is both more and less than the languageless thesis: certain written signs are necessary; but need they be linguistic? I will return to this point below in §30. What now according to Heyting are the "natural numbers constructed at different times" ? If they are the inscriptions, as "simple inspection" seems to suggest, then they are not mental items at all and Intuitionism verges on a form of Hilbert's position. If the inspection somehow mediates comparison of the mental constructions which the written signs signify, one would like this to be explained rather more! For if the signs are adequate for the comparison, then we don't need the constructions though we do need the signs; and if they are not, well they are not, and we remain facing the comparison problem with the solipsism it implies.

I shall discuss the dependence of mathematics on writing in the next chapters. My first argument against the expressionlessness thesis, and one that Heyting seems to grant, is that in mathematical activity *signs for intuitions* are not merely convenient but necessary.

§26 Signs for defined terms are necessary in mathematics

Let us proceed to '3*4'. What mental item corresponds to this? In particular, what mental item corresponds to the '*' in Brouwer's account? How is the term '3*4' made from 3 and 4? Operations in the mathematical sense like * are not cited by Brouwer as basic elements, so I conjecture that they must be composed somehow out of the primitive juxtaposition of succession in time: we know that multiplication

can be defined from addition, and it is certainly reasonable to interpret $3+4$ as the temporal juxtaposition of a 3-construction and a 4-construction in that order; similarly for $4+3$. Now mere juxtaposition will not do for $*$, since that is what we call '+'. Taking our clue from the usual definition of $*$ in terms of '+', let us not allow ourselves to be distracted by the kind of geometrical aid used in actually teaching the concept of multiplication to people. This cannot be legitimate since for Brouwer geometry has to be reduced to arithmetic. But we can interpret signs involving $*$ thus:

' $a*1$ ' is another sign for a ;

' $a*(n+1)$ ' is another sign for $a*n + a$

Actually the first of these is rather puzzling - the same puzzle as what zero is for bleak Intuitionism. What is the value of having various signs for the same thing? Surely if language is dangerously untrustworthy, as Brouwer is always saying, one should avoid such things? (Any such response as Frege's to the same conundrum relies on platonic referents referred to in different ways; no such move seems possible for the Intuitionist.) The obvious defect of this whole device, though, is its seeming pointlessness. In order to **make** the construction $3*4$ for the "left-hand side" of my mental construction of [AS], I shall have to use not $*$ but, in the manner shown by the recursive ladder used in my proof, to effect the actual construction of 12 (as it turns out to be) by successive additions of 3. In which case there is no actual mental equivalent of the sign $*$ - the original item ' $3*4$ ' is a linguistic item which *abbreviates* but does not *report* a mental content distinguishable from that designated by '(((3+3)+3)+3)'. There just is no *separate* item referred to by ' $3*4$ ' - how could there be? There is no room for another construction, it has to be the same one. So we must ask again the question, (already raised in reality by our use of signs like '3'), how definitions can be understood on the Intuitionist view. For on this account we do not have a theory of multiplication. It has been *literally* reduced to the theory of addition. There is nothing in the construction corresponding to $*$. Ontologically this may seem excellent news; but

from the point of view of the philosophy of mathematics it seems to me to be disastrous. Mathematics just is about multiplication as well as addition. After all, Presburger arithmetic (+ but not *) is decidable, but Peano arithmetic (+ and *) is not²⁵! Moreover, this argument could be repeated to show that addition is reducible to iterated successor application. So my second argument against Intuitionism's expressionlessness thesis is this: most of mathematics is theory about defined items, like *, but if that thesis is taken seriously there is no such thing as definition in real mathematics, only in its inessential expression.

§27 Signs for theorems are necessary in mathematics

Now let us consider what construction corresponds to the whole statement [AS], i.e. ' $3*4 = 4*3$ '. Brouwer spends no time explaining how equality occurs in the mental life of the mathematician; I suppose he thinks it obvious. Let us first consider the simpler case of ' $3+4=4+3$ '. I presume that the mathematician has constructed the two items on left and right side of the equality, and holds them both in intuition for comparison. Now this examination cannot find them identical, for they are clearly not - else they could hardly be compared. Nor can it find them to be simple replicas of one another, as it might find two successive 3-constructions - else the statement would be trivial. In this example we might claim that the comparison in question differs from such a simple one only in the *order* of construction of the two items, that they are replicas if viewed in abstraction from that order. The argument of the previous section seems to force this concession anyway. Indeed Brouwer argues rather like this in his thesis²⁶, though at that stage he was still describing mathematics as done in writing. We might grant this point, but only point out that abstraction is a notoriously mysterious "faculty" to attribute to the mind. How are these "abstractions" made and how does the mind go about "comparing" them? I am not denying that this can be done, I am suggesting that the

²⁵ Boolos and Jeffrey [1980: ch21]

²⁶ Brouwer [1907: 15]

process is complex and hard to explain (and, I shall argue below, *requires* writing). Brouwer and Heyting hardly try. Notice that the rather similar explanation²⁷ which Hilbert might give of this equation in terms of sequences of strokes does not require any extra mental repertoire, because each side of the equation actually *is* a sequence of seven strokes. The 'order of construction' is no longer present in the stroke sequences for which we ordinary mathematicians use the signs '3+4' and '4+3' - but only because an axiom of associativity permits brackets to be discarded in iterated summations. The question for Brouwer is this: does abstraction in the mind have this effect of writing, and if so how? How does the mind remove mental brackets? And indeed put them in? What is the status of such an axiom? Not all algebraic systems are associative, after all.

This is not the end of his problems either. Even if we allow the prop of paper to effect the comparison which results in a judgment of equality, we still must ask what is the nature of the mental construction which constructs what corresponds to '3+4=4+3' out of the two terms. Is a new mental element corresponding to '=' introduced between the original term-constructions? Or is a new item corresponding to the equation created from copies of them with such a new item? In either case we have a difficulty. In the first case, we have no distinction between term and proposition. In the second case, we find that the supposed mental construction becomes a reflection of the expressed proof rather than the other way round. 3 corresponds to the term '3' and the construction $3 \cdot 4$ to ' $3 \cdot 4$ ', but what corresponds to the equation ' $3 \cdot 4 = 4 \cdot 3$ '? Is it a construction from constructions? If so, what makes it different from a term-construction? Notice that we cannot call in here linguistic elements like Frege's unsaturated expressions because they are not mathematical constructions from the Brouwerian elements. There is no separately identifiable mental correlate of a sentence, unless there can be constructions of at least two kinds corresponding to names and sentences in linguistic terms. But

²⁷ An example like this is discussed in Hilbert's [1928], but of course for Hilbert no such identity needs proof, it can be verified by "contentual considerations" - i.e, at a glance. I shall discuss this idea in Chapter 12.

Brouwer can hardly admit this without granting some kind of "language of thought" thesis which would reduce [IT2] to the claim that mathematics *needn't* be spoken out loud.

My third argument then is that further signs are needed in mathematics for making statements, but the Intuitionist account cannot explain this.

§28 Signs for proofs are needed in mathematics

I turn to now to the question whether written proofs can be taken to report constructions as Intuitionism claims. Let us first consider the general nature of proof in Intuitionism. We see from IT3 that correct proofs report veridically the completion of mental constructions. Proofs are linguistic descriptions reporting constructions. In order to understand what Intuitionism admits as a construction corresponding to a proof we must consider the logic which Intuitionists generally admit to describe the proofs they accept. I think the idea is that constructions from prior constructions which "follow" an intuitionistically acceptable proof will effect the construction which its conclusion expresses. The logic which makes explicit the canons for such proofs, *H*, has been formalized on the basis of [IT4] by Heyting and others. It includes most of the usual rules of classical logic *C*, except for certain well-known exceptions. In the propositional calculus these are basically the lack of the laws of excluded middle and double negation, and in quantified logic, the interdefinability of the existential and universal quantifiers using negation. (The differences all arise from the Intuitionist notion of negation.) Intuitionists have to have a logic for practical purposes, in particular in order to criticise classical mathematics.

But what do Intuitionists think a proof does? Assume sentence *S* accurately describes construction *C*. When is the assertion of *S* justified, according to Intuitionists? When it reports correctly a construction, that is when *C* has been effected by the asserter. In that case it can be supplemented by a proof each of whose elements corresponds, in the way which we have been trying to get clear, to

constructed items in the mind of the asserter. Now this view seems to have some obvious difficulties. If you justifiably assert S, can I then do so? Not unless I effect C too, if we are to take seriously the repeated claim [IT3] that a mathematical statement has as a correct translation "I have effected.." This then perhaps shows us the role of proofs, proofs are linguistic means to aid this *copying*. A proof of S is sound if I can effect C by using it. If I have done C, can I assert S anytime later? What if I sleep, so that the construction is removed from my consciousness? Must I effect C again to be justified then in asserting S? If not, how is the certainty of mathematics to be explained, given the defects of memory? I could remember the construction wrongly, and "effect" it differently, that is make a different one. But then presumably I would actually thereby be justified in asserting a different statement S*. Would S still be assertible? Is S assertible by just *anyone* once I have effected C? Or must they obtain some kind of warrant from me? If this must be a proof which they use to construct C, there is no way to use anyone else's results without repeating them.

These difficulties might perhaps be circumvented if we restrict our attention to a single idealized mathematician who never interrupts her activity, so that every item constructed remains part of the growing construction in her mind²⁸. She will be constantly justified in asserting any S which expresses a C which is part of the present construction (but what would be the point of assertion for such a being)? Yet there will occur, say, the number 3 in various places in this construction. It cannot be the same item, so we must accept that there are ways for the mathematician to recognise the "same number" in different instances. But then we may invoke Wittgenstein's argument²⁹ against essentially private languages to throw doubt on the coherence of the idea that there can be only 1 mathematician, else there will be

²⁸ Brouwer says [1933: 443] "for a human mind equipped with an unlimited memory, pure mathematics, practised in solitude and without using logistic signs, would be exact, but the exactness would be lost in mathematical communication between human beings with unlimited memory, because they would still be thrown upon language as their means of understanding."

²⁹ Wittgenstein [1953: §§258ff]

no distinction between thinking that 3 has been constructed, and its being so³⁰. The mathematician would have no genuine criterion for the identity of 3s.

Let us disregard this doubtful solipsist response then, as being in any case too much of a distortion of the obviously and necessarily social character of the actual mathematical process we know, since it is real not ideal mathematicians we wish to understand. Then we must return to the question of when I can accept assertions of yours, and when I may assert them too. I can accept your assertion if it enables me to copy your construction. But now how is it possible for two different copies of the same item to exist? A '3' constructed by me cannot be the same identical item as one constructed by you. But then we must have some way for us to establish that the two are indeed the same. This cannot be purely mental. **On the contrary, we decide that they are the same by reference to the proof.** When I read a proof I do go through some sort of process of construction³¹; but I continually attempt to reconcile it with the proof. If I cannot eliminate conflicts, I must decide that either I don't understand it or else it is wrong. In the first case I am blaming my constructive capacities for the problem, but in the latter I am judging that our constructions differ.

So my fourth argument against the expressionlessness thesis is that proofs play a necessary role in mathematics since we don't want to have to invent it all for ourselves each day; for the function of pooling our knowledge, proofs must be public; and this pooling is necessary for the development of a cumulative social tradition like mathematics.

§29 Some difficulties raised by Dummett

Dummett has rehearsed³² a number of difficulties about the Intuitionist concept of proof arising from its admitted logic of the connectives 'or' and 'implies',

³⁰ We want someone's knowing the theory of groups explained, not what she knows about it.

³¹ Cf Chapter 12 for more on this.

³² Dummett [1975a], [1977]

and the existential quantifier. Intuitionists certainly admit hypothetical constructions, that is they admit proofs of sentences $A \rightarrow B$ where A has not been proved, such as in exhibit #3B³³. Here $a \# b$ (a is apart from b) is taken as a fact, A , for the proof and thereby justifies the use of the signs 'n', 'k' and 'm' as referring to determinable numbers in establishing the theorem B . Such a hypothetical proof describes a construction such that given a construction of A we can then produce a construction of B , and a proof that this is so³⁴. Now what can it be to be given such a *construction*? I can understand a *proof* which describes how to operate on a *proof*, since a hypothetical proof can be given a linguistic label and described, perhaps by cases, or in terms of the number of connectives involved. But how do we gain purchase on a hypothetical *construction* in this way? If we don't have the construction itself to work on, as by hypothesis we don't, we must have something to stand for it: a symbol of some sort. But this breaches the languageless thesis - unless Brouwer has some idea of nonlinguistic labels for hypothetical constructions. Now labels aren't names, as I shall be arguing in Chapter 6, but they are certainly signs. So this response would not help the Brouwerian position.

Clearly Intuitionists do accept assertions based on proofs that we have not actually carried out, but which we know that we could "in principle". This seems to require a distinction between proofs and "demonstrations", or "canonical" and "ordinary" proofs. But can the true Intuitionist accept any such distinction? Surely a "demonstration", by adverting to a construction possible in principle but not carried out, at once breaches not only the thesis of languagelessness as I have suggested, but also the "first act" notion of what proof is for Intuitionism: if mathematics is mental constructions, a "report" of (i.e. label for) a construction I have not actually carried out must be for the Intuitionist not mathematics at all. Yet I take it that Intuitionists have in fact accepted this distinction, else proofs like that of the exhibit cannot be part of Intuitionist mathematics. This once again seems to

³³ From Heyting [1980: 25]

³⁴ This sounds rather surprising! A proof of what we can do as part of mathematics?

refute the bleak Intuitionist position on language if there is no possible rationalisation for this "eclecticism". Dummett³⁵ argues not, and Wright agrees (after a long discussion).

Further difficulties, which I shall not go into here, are raised by Dummett against any attempt to interpret "having a proof" for the Intuitionist by some temporal parameterisation. My view on Dummett's whole imbroglio about Intuitionist logic is that it depends on being too charitable to Intuitionism, and permitting a continual ambiguity between "I have a proof of p", "we have a proof of p" and "there is a proof of p". We must take Intuitionism at its word and try to make the first locution the basic one. If we do so then there is a possible consistent treatment of the logical constants, but it makes Intuitionism solipsist. This is entirely consistent with everything Brouwer says. If we reject it as a travesty of the reality of mathematics, and greatly doubtful as a possibility even in principle for Wittgenstein on reasons, we must then ask whether there is a way to make a logic of "we have a proof" or "there is a proof" which conforms to both the Intuitionist canons of inference *and* conception of proof. My argument in this chapter is that there is not and cannot be, as long as the fundamental Intuitionist idea is retained of mathematical items as mental constructions. I shall go further in the next chapter, and argue that speech cannot overcome this difficulty either - only writing is a sufficient substratum for proofs to support a logic of the kind actually used. I can accept that first person proofs are mental constructions (but not "languageless"), but second person proofs are at least talk and third person proofs must be written: constructions in intuition occur mostly on paper! Mathematics requires third person, objective proofs.

³⁵ Dummett [1975a], Wright [1982]

§30 A doubtful argument

Tait has an argument³⁶ against the Intuitionist thesis that language is not essential to mathematics. It is that constructions make use of rules, and that Wittgenstein has showed that rules are linguistic. This is a bad argument and it is important for my purposes to see why. The point is that rules need not be linguistic, in the usual sense of linguistic. I shall be discussing sign-systems in the following chapters which are not linguistic, in particular diagram systems. There are rules involved in such sign-systems, but they are not linguistic rules, they are *semiotic*, in a broader sense: rules for the use of signs which are not necessarily words. Not all signs, I shall be maintaining, should be construed as attempts at speech. So the argument used by Tait, granting its premisses, might establish only the essential *public, expressed* character of mathematics - which Tait himself elsewhere identifies as the crucial thing about proof, rather than its linguistic character. My point is that the rules in question are rules for mathematical signs, which are not necessarily words, though they are written. I will argue later in Chapter 7 that such rules cannot subsist except in a linguistic context, but that is a different point. Tait says "I can't do my mathematics if I am not disposed to use symbols in a certain way, i.e. if I have no language." The identification of symbols with linguistic symbols here is blatant. But if we examine the explicit concession by Heyting cited above, that symbols are needed for the comparison of 3 and 3, we see that iconic written symbols are what he allows: a mark on paper for each element in the construction of 3. So the rules which are necessary are rules for the manipulation of diagrams (or notation, which I shall show below to descend from diagrams). I don't mind whether we classify diagrams into language, or admit nonlinguistic rules: but we must do one or the other. I shall continue to use 'language' in the narrower sense which includes Swahili but excludes BASIC or the language of flowers. This is the one usual in the philosophy of mathematics - except when it is in difficulties like the present one. I shall use 'sign-system' in a more inclusive sense, so that we can recognise that

³⁶ Tait [1983: 76-7]

certain sign-systems, languages like English, are neither good models for understanding others, such as the notation of Principia Mathematica - nor vice versa.

In this chapter I have argued that for a variety of reasons, the Intuitionist idea of mathematics as an essentially languageless activity of pure thought is not acceptable. I appear to have left open the possibility that mathematics could be a languageless but not signless activity, but the next several chapters will show that this is not so. In any case, I have shown that a necessary feature of mathematics is its public expression in signs. In the next chapter, I turn to the question of what form such expression must take, and in particular what role notation plays in it.

Chapter 4

Text

§31 Examination of texts

In the last chapter I argued that, contrary to the view of bleak Intuitionism, mathematics must be publicly expressed. In this chapter I consider the question how mathematics is actually expressed. I will begin by examining in detail an example of the expression of mathematics. The example I shall look at here is part of page 174 of Curwin and Slater's *Quantitative Methods for Business Decisions*, exhibit #4, which I will sometimes call CS174. This is not an example of pure mathematics, and might easily be rejected by many writers as not really mathematics, or as involving confusing complications. I shall explain in section §39 why this would be a mistake, but in any case it will become obvious as the discussion proceeds that any peculiarities of this particular example can be explained on my account.

In the discussion which follows I shall use some terms frequently, so a few preliminary remarks about them are in order. By *sign* I mean any kind of item which carries significant force, be it a word or a roadsign, a burst of morsecode or a file on a floppydisc. By *discourse* I mean significant production in the broadest sense. A *text* is any piece of discourse which it makes sense to consider on its own for some purpose; I do not intend that *mathematical text* shall have any didactic implication, 'text' does not mean textbook; it means any piece of mathematical discourse. I shall say some explicit things about how I understand these terms later, but for the present they are intended to be theory-neutral. In particular, I wish to explicitly reject the identification of signs with words; words are signs, but many signs important in mathematics are not linguistic and, I shall argue, cannot be replaced with linguistic equivalents.

The main points about mathematical texts which I make in this chapter, and which are enlarged on in the following chapters, are as follows¹. First, any text has a context in which it is used, and signs off the page are part of this context: mathematics is always assumed in the presentation of mathematics. Second, the mathematical text is *written*. Third, the signs in a mathematical text are of several different kinds: Written words, Diagrams, Notation and Paragraphy. Fourth, these correspond to different kinds of sign-function: the written words anchor the sense of the text in everyday experience, the diagram shows the idea being exemplified, the notation clarifies the inferential structure of the text, the paragraphy ties the text together. Fifth, texts have various functions, for example what CS174 does is illustrate a technique. Sixth, signs can always be replaced by others of a different kind to the same end, but there is always some cost, so there is a limit to the re-expressibility of texts.

I next ask, what has the literature to say about the observations I have made about this text? I consider first the logico-formalist account of mathematics and reject its idealisation of the text: it is *this* text I wish to understand. I turn then to Nelson Goodman's account of "symbolsystems", but find it wanting. To explain why, I digress slightly to consider the problem which he makes the motivation of his work, the explanation why some arts have forgeries, others not. I argue that his explanation is not satisfactory, and for reasons which ensure that his framework cannot account for my observations, because he conflates all the many functions of signs to naming. Finally I consider briefly some recent attempts by writers on mathematical didactics to conceptualise text. These seem to me to be too *ad hoc* to provide any illumination of the problems of philosophy of mathematics.

¹ Despite my examining real texts, I am determining the effects of the signs in them by consulting only my own experience as a reader and those of people I know. There is some room for methodological finesse here in subsequent work. Some literature in mathematical didactics which attempts more objectivity on this point but it is too fragmentary to be much use for the present investigation. I discuss some of it briefly in §46.

§32 Any mathematical text has a context of use

The intended use of the sample page CS174 is easy to discern if you examine the book it is taken from: it is part of a chapter teaching about normal distributions, and on this page the idea is being explained of standardising a normal variable. The book from which it comes is aimed at first year business degree students in Britain and claims to make little demand on prior mathematical experience. It contains mathematics all the same, as any student enrolled in "Quantitative Methods" will quickly tell you. So this text is what we might call *lower level undergraduate service mathematics instructional text*. Of course it can be used in all sorts of ways - as an example for the philosophy of mathematics, or as extension material for overactive school students, or for learning English - but the typical and intended use is by a certain kind of student to learn certain pieces of mathematics, in order to meet a requirement of an award which the student wants to gain. We will not understand mathematics if we ignore the fact that in this entirely typical example the transmission of mathematical knowledge is only a small part of what is going on. The knowledge is commonly little valued, forgotten soon after the end of the course, and not put to much use anyway by most of the people using such a text. What mainly endures is an attitude towards mathematics and the academic record of the student. Most mathematical experience, unfortunately, has this flavour, of compulsion, alienation, mystery and social formation. Let's not delude ourselves, mathematics is supported in our culture as a necessary evil.

Before leaving till Chapter 9 this question of context, notice that the example text has a textual context, as well as a wider social context towards which I have gestured. If we examine this page we find that its sense is not self-contained: it continues from the one before and goes on to the next, it uses various signs which have been introduced earlier in the text, it uses various signs which have not been introduced in the text at all, it makes many assumptions about the reader's knowledge and skills, and to use it requires the reader to refer to other parts of the

text, specifically the table of the standard normal distribution at the end of the book. A new piece of mathematics is being illustrated (namely, $X \sim N(\mu, \sigma^2) \rightarrow (X - \mu)/\sigma \sim N(0,1)$), but old mathematics is drawn on in the text. The reader must already understand (at least) subtraction, division, addition, square-rooting, reading graphs, looking up tables, proportion and percentage. Signs relating to all these are used in the text, but not explained anywhere in the book. This use of old terms is a general feature of discourse, but has special effects when texts are written. How should we think of the relation of this text and the table of $N(0,1)$ which it calls on? Is the table a virtual part of the text? And what kind of a sign is a table anyway? A table is not equivalent to a sequence of assertions; a look-up table is a two-dimensional resource, a kind of archival diagram. This will be considered further below in Chapter 6.

I put aside temporarily² this question of the context of mathematical text, and consider how this text can be used successfully in learning mathematics. For the present I consider only the text from "If a population..." to the end of the page. I have highlighted and labelled various parts of the text in yellow and red to assist my discussion.

§33 The signsystems of mathematics

Something which seems obvious but will actually turn out to be extremely important is that the text we are examining is written³. I shall argue in the next chapter that mathematics depends on writing. The signs to be found in this example are of four kinds: **Written words**, a **Diagram**, special mathematical signs or **Notation**, and some auxiliary signs which exemplify what I shall call **Paragraphy**. The fundamental observation is that these different kinds of signs have different kinds of function in the text. Roughly speaking (I shall improve each of these claims in the chapters to follow) the Written words enable the author to say things, the

² I return to it in Chapter 9.

³ I include the printed in the written for the present. At various points below I shall comment on important differences which I have no space to go into.

Diagram to show them, the Notation permits the plausible construction of the exhibited whole, the Paragraphy enables the text to be comprehensibly fitted into its context. Thus Notation plays one of four interdependent function in the text. Notation itself is only a counter system like an abacus, but a visual one; the mathematics is its employment in the whole text. By construction I mean that the numerical answer is obtained by legitimate manipulation of the notation (along the lines of the logico-formalist account of mathematics). By plausible I mean that the legitimation is evident.

The text consists of fourteen main signs which I have labelled W1-6, N1-4, D and P1-3. The Written words are the six sentences, W1 - W6. The diagram D is obvious enough. The special signs, Notation, are the equations in the lines I have labelled N1 - N4. The Paragraphic signs are the labels 'Example' and 'Figure 11.3', and the footer '174...'. This fourfold division of the main signs reflects their different functions in the text. But notice that these four sign-functions are actually intermingled in a complicated way. The Written word 'Figure' is part of the Paragraphic sign P2, the Written words of W4 are essentially Paragraphic, the Notation N1 retains some words, the Diagram uses words, the sentences include Notation. We could, in logic-formalist style, identify all the irreducible signs in this text; for example the sign '0' occurs in many places, and could be found here or in related texts as part of signs which I would allocate to each of the four sign-systems. But it would be misleading to proceed so, because the different sign-functions are what determine the use of the text and therefore its content and cognitive significance. Later⁴ I shall go some way to analysing the kinds of signs typically found playing these various roles, but at present the roles are what I want to concentrate on. In the next four sections I will describe these functions, and in the next four chapters I shall develop these points. The four signsystems which I distinguish consist of functional roles in text, not of sign shapes. The sign 'Example' is not intrinsically paragraphic, but in this sample text it is a

⁴ In the Chapters 5 to 8 on the four different signsystems.

paragraphic sign because its role is neither to say, nor to show, nor to display calculation.

§34 The Words tell the tale

My view of the role of *Written word sign-system*, that is of the signs which are roughly identifiable as transcriptions from speech, is hardly unorthodox, it is that it is used mainly to make statements, and to a lesser extent for any other of the functions of spoken word. In the exhibit, W1 makes a statement which is illustrated by the rest of the text. One can imagine a teacher making exactly this statement in speech and there is no extra mystery about it beyond those which the general philosophy of language must address. W4 is not a statement but a direction, which again could readily occur in speech. It is typical of the fact that mathematics is not, contrary to the logico-formalist idealisation, just a stream of statements but includes other kinds of speech act. Equally importantly, W4 explicitly refers to the text itself, both by using Paragraphy to refer to the figure, and also by alluding to its own point: to say that the two scales may "help in understanding some questions" rests for its sense on the reader's expecting to try to do some questions on this idea, and expecting to have trouble understanding them. It assumes that the book is aimed at her learning. These two pragmatic features of this text, self-reference to its expression and to its point, are not peculiar to this one, they are commonly found in text just as they are commonly found in speech. I draw attention to them to emphasise that the function of the Written word in mathematical text is no one-dimensional simplification of speech.

However in the next chapter I shall argue that statements in writing are cognitively quite different to those in speech. Consequently, although many functions of speech are carried over into the Written word in mathematical text, and many more than logico-formalist accounts recognise, there are new functions and new forms of old ones to be seen too. In particular since ostensions are not available to the writer as they are to the speaker, in written text one finds more explicit

grounding (or "scene-setting", as in an important subspecies, drama). This is particularly obvious at the beginning of a complete text such as the book from which the example is excerpted. The written words anchor the sense of the text in everyday experience. Books, particularly didactic ones, often start by relating what the reader will find within, to her previous experience, her aims and so on. And even in the middle of a text, as in the example, general knowledge is being assumed: the ideas of rock concerts, of promoters, profit, breakeven and loss are assumed in CS174. The knowledge gained by the reader by use of this example (assuming it works) is that standard normal theory can be applied to something like promoting rock concerts by the technique of transformation. Actually this is an exaggeration, since it is a very phoney example - how could you know that the distribution was (approximately) normal without having data which could supply directly the proportion of losses? - but that is the intention. It is typical of examples to be phoney like this - and that is a significant fact which I shall discuss later⁵. In this short text the grounding is verbal, as it is for any mathematical text, though again typically it is written words. The words involved provide anchorage for the sense of the text in everyday experience ('rock concerts', 'profit and loss'). Spoken narrative will also have this grounding in it of course, and I am not suggesting that there is a complete contrast here. Written literature takes over and transforms devices present in speech at the level of single sentences and above it. But it is their being written that allows the peculiar mathematical character of such texts as the example to be created, through their supplementation with signs from the other written sign-systems. The next chapter is devoted to the Written word sign-system.

§35 The Diagram shows the idea being exemplified

The Written words say what's what about transforming normal variables. The diagram, on the other hand, illustrates the example, by *showing* the idea of transformation. It does this by showing the division of concerts into profitable and

⁵ In Chapter 9, on readers.

unprofitable, representing them by the areas under the curve L and R of the vertical at -1.25. Notice that to see that you must know how to look at it. The diagram also *shows* the method of solution (the two axes). Finally it *shows* the basis of the solution, the normal curve. The vertical at -1.25 divides the spaces enclosed by the curve and the axis; the tickmarks indicate the scale; the axes provide a baseline for the definition of the total space A. These enable the diagram to show the division of the area.

Notice that the same diagram could be used for other problems, for example for one in which we are given the proportion 10.565 and want to know the related normal deviate -1.25. It would illustrate the same state of affairs, but would have a different point. This is an observation commonly made to distinguish pictures from discourse⁶, but in fact it applies to all signs. Any sign can be taken in a multitude of ways by someone determined enough⁷. But in any case it is in general true that the diagram shows, not says. The point of having both Written word and Diagram is precisely to make use of their different but complementary functions as signs. The diagram through showing assists the words to say more than they would without it. Chapter 6 is devoted to Diagrams.

§36 Notation makes the inferences evident

As a first approximation, we can say that Notation plays the role of visual counters for inference and calculation. This is an oversimplification which I shall try to correct, but I am trying at the outset to make clear that Notation has a function which is quite distinct from those of the other three sign-systems. The inferential structure of this text is in some ways fairly simple: we get -1.25 from the data, and then 10.56% from the table and -1.25. The "pure mathematics" to be found here is not great, merely the obtaining of the number -1.25 from numbers 15000, 4000000 and 12500 in the given information. The numbers must be bound to the

⁶ Block [1981] *passim*.

⁷ Cf Wittgenstein [1953: §86ff]

letters X , μ and σ (in the formula for standardising a normal variable given on the previous page); in N1 and N2 and N3 this is half done, (using in N3 an arithmetical inference and the fact that variance means σ^2); then N4 produces -1.25 using arithmetic and the identity $Z = (X-\mu)/\sigma$ stated on the previous page, replacing 'mean' by ' μ ' etc. Then 10.56 is inferred from -1.25, and the fact (which can be read in the table on p327 of the book) that 0.1056 of standard normal variates are less than -1.25. Actually, this last is a rather complicated and impure inference. First, there is the logic of looking a value up in a table. Put that aside for now⁸. Second there is the inference from what the table "says" to 10.56. W5 effects this latter. Finally the answer to the question is inferred from the intermediate result of W5 and the facts that the concert attendance is a normal variable (W2) and that (W3) any attendance less than 12500 counts as a loss.

Notation is used in this text for various purposes. Though they are all related to its inferential structure, this relation takes very different forms. The repeated occurrences of 'Z' provide a clearer form of anaphora than would words; on the other hand the use of % is only an abbreviation for 'percent'. The use of the numerals in place of words for two thousand and so on is more special than its unremarkableness might lead one to think. How do we grasp that Z is indeed -1.25? Because we can see at a glance that the arithmetic is correct. This would be less obvious in words, and if the figures were more realistic, impossible to see at all. The algorithms for subtraction and division are implicitly used here in a way which the notation of numerals makes possible. The inference is of course fairly informal and inexplicit. If we made it formal and explicit we would obscure the essential simplicity of the idea of transformation, though we might gain some "rigour".

I have not made "the" function of Notation very clear in this section, because I do not think that there is a single function for which it is used, anymore than Written word is used only to make statements or Diagram only to illustrate. In

⁸ I discuss this further in Chapter 6, especially §72.

Chapter 7 I will discuss these various functions in more detail and relate them to the effects of notation identified in Chapter 1.

§37 Paragraphy ties the text together

Paragraphy connects parts of a text together, assists its comprehension by "chunking" it, and relates the text to its context, particularly the textual context which I call the archive and which I will discuss at length in Chapter 9. The label 'Example' connects the text we are considering to the preceding claim in a specific way. The label 'Figure 11.3' connects the diagram to the rest of the text via the related sign 'Fig.11.3' within W4. The sign P3 connects this page to the others of the text, the '174' indicating its place in the sequence of pages, and the label 'Normal distribution' locating this page as one of the chapter 11. The role of the '/' is basically aesthetic, whereas that of the the long sign

' _____ ' is to make clear that P3 is not part of the main stream of text. There is some more subtle paragraphy here too: the labels 'z' and 'x' on the axes are variables, a form of notation which grew out of paragraphy. The spacing of the signs on the page, particularly the use of the margin for 'Example', is designed to contribute to the text's comprehensibility. Without these paragraphic effects it would be much more difficult, perhaps impossible, to grasp what is the point of the example - why it is *here* on *this* page. Chapter 8 will discuss Paragraphy in detail.

§38 What this text does is illustrate a technique.

Texts can have a great variety of functions. Since the advent of the theory of speech acts (which I discuss in the next chapter) it has been widely recognised that speech can have many functions - ordering, requesting, promising etc need not be seen as disguised forms of, or derivative from, statements. The logico-formalist hegemony however includes the view that in mathematics (and science) the essential locutions are statements. I would not deny that statements are essential,

though I would deny that no other kind of speech act is necessary. But in any case there is a point which most philosophies of language, and linguistics, ignore. Texts are created from sequences (or more complex structures) of sentences and other signs⁹. But it is not the case that a sequence of statements is itself a statement. In first-order logic, it is; but so much the worse for first-order logic. A sequence of statements might be a fairy tale, a joke, a testimony, a declaration of love or a news bulletin.

It would be a mistake to think that the point of the sample text is to state certain facts (W2,W3) and what follows from them (W6). They are almost certainly untrue anyway. What follows "Example" *illustrates* the claim in W1. The example also *illustrates* the hint in the previous paragraph, and the embedded diagram *illustrates* the situation under discussion. The whole sample text is illustration. It is not evidence or justification for W1, since that can be proved, and would be in a different kind of text with a different purpose. It is not any kind of justification or proof by example; it is an *instance* used to give the reader some concrete content for the claim of W1, and thereby some inclination to believe it. Thus its point is entirely "psychological" - as might be expected in a didactic text. The good student grasps (the truth of) W1 by seeing that the particular question is answered and that the whole sample text is indeed an example. Proof by example is *pedagogically* sound because the aim of the text is the grasping of the technique by the reader. Exercises to reinforce the example follow on the next page of the book. These permit the student to practise the technique which has been illustrated, by emulation.

§39 Local and global eliminability of signs

A given text has a specific purpose, but by the nature of human purposes it can be achieved in many ways, like skinning cats. In particular, none of the signs actually used in this example is necessary, all kinds of variants would do as well.

⁹ There is some linguistic literature on texts, for example Dressler [1978], van Dijk [1977], but it seems to be firmly in the grip of the logico-formalist hegemony.

The authors no doubt considered some in composing it. I don't even claim that because this text contains a diagram, it could not be replaced by one just as good without a diagram, or that one could not have an "equivalent" text with no paragraphy (though this is rather doubtful). But I want to stress the following point because it will be very important later: any replacement, no matter how small, will have some costs as well as benefits. There will be significant cost in removing all the signs of one of my four *codes*¹⁰ from a text (and adding others from another code to compensate). And there is no reason at all to suppose that mathematics as a whole would survive the elimination of one of these codes. In fact I shall show in the following chapters just how implausible such reductive ideas really are.

It is rather implausible that the words could be eliminated from this text. Even the great logico-formalist treatises at least *begin* with a great deal of verbiage to launch themselves¹¹. The cost of doing without it would be prohibitively high in a didactic work, for it would make quite unrealistic assumptions about the mathematical experience of its readers. In any case, I shall be arguing in the chapter on Notation below, Chapter 7, that statements are not made **in** Notation but **using** it, from which it will follow that Written word (or word anyway) is not entirely eliminable.

The cost of doing without the diagram would be to weaken the sample text. The diagram illustrates the example; as such it is "merely" illustrative. But if the text failed because of its omission it would fail entirely. (It could not be *just* omitted of course without altering the text considerably since it is referred to in the Words.) Although *this* diagram is "only" further illustration, not all diagrams are so easily dispensable. Those in the problems of Descartes and Euclid are typically the way the

¹⁰ This term is often used with theoretical overtones. I am adopting it here without any prior theoretisation. For the present I just want an alternative term for referring to my four sub-signsystems - 'signsystem' can be rather awkward in some places. 'Code' and 'signsystem' are intended to be synonyms.

¹¹ I mean *Principia Mathematica* and the like. This point is taken up again in Chapter 7.

problem is posed, and "equivalent" algebraic problems are different problems¹². The practical problems for which mathematics is used very commonly have a spatial specification for which only a diagram will do. For example, the brachistochrone problem of the 18th century, or the travelling salesman problem of this one. These have abstract formulations in notation, but they would not be comprehensible in these forms unless already understood through the use of diagrams.

The cost of doing without the notation would be to cut off the text from the mathematical capacities of the reader. This text, like any other, could be expunged of all notation. The word 'mean' could be used instead of ' μ ' all the way through, numerals be replaced by verbal forms throughout, and so on. The text *might* still work. But it would do it less reliably, less perspicuously, less evidently. In cases where it still succeeded the reader would have to give special meanings to written words such as 'plus' and 'equals' to make them equivalent to those of '+' and '='. This can of course be done - you can do anything with enough effort. But it would be difficult, and the results unstable, because the words already have meanings which are *not* confined to mathematical contexts. Since *we* have the N signs as well as the words, we can always take written words 'plus' and 'equals' as "long for" the special signs in a particular text if we are told somehow that that is required. If we did not have these extra signs already in use it would be much more difficult, since we would have to establish somehow else the special mathematical meanings which we wish 'plus' and 'equals' to bear here. The fact that the signs '+' and '=' evolved out of verbal mathematics shows that there is no simple impossibility here; but the length and difficulty of that evolution, and the restricted achievements of mathematics which does not introduce such further signs, should warn us that there is, equally, no simple possibility of complex verbal mathematics either. The fact that

¹² Knorr [1986] shows the great extent to which the geometry of antiquity was concerned with just such "problems". This point is taken up again in § 73 of the next chapter.

[12500 - 15000] / 2000 is -1.25 would no longer be *perspicuous* as it is to us who are aware of how the arabic notation for numbers supports algorithms for arithmetic, and a great deal of extra text would have to be added to make it so¹³. For these reasons and similar ones, if notation were eschewed the text would actually have to be much longer and it is at least doubtful if it would often succeed in getting the idea of transformation across to its intended audience (who are not, let me repeat, infinitely capable and patient pure mathematicians). It would be longer but cut off from the prior mathematical experience which gives legitimation to the inferences made in it. One might say that all mathematics could be verbalised, so that the contact could be re-established with it that way, but I deny that this global verbalisation is possible. Any individual text may be verbalisable without texts collectively being so. In Chapter 7 I give several arguments for doubting that mathematical texts could ever be *typically* only verbal. The thought experiment of changing this text presupposes that the same audience would remain available; but this could not be simply assumed if all texts were to be verbalised.

The cost of doing without paragraphy in the example text would be to disable it. In some cases it is virtually impossible to separate the Paragraphy from the Words anyway. The sign W4 is really a paragraphic pointer to a feature of D, together with some indication of why it is there. Even with this kind of 'advance organiser', instructional text is very poorly understood by most of its intended readers, and succeeds all too little in its aim. Without Paragraphy it would fail almost universally. The present kind of social penetration of mathematics would be impossible without paragraphy. Impossible in the sense that creatures such as we are, simply would not be able to spread the knowledge to the thickness that we do. There is a cognitive economics which arises from our cognitive limitations of grasp and attention and so on, just as there is a material economics deriving from

¹³ I have adjusted the format of the first line of this sentence to keep the N-sign together to show this.

the limitations of the environment. I shall return to this point repeatedly, in particular in Chapter 9.

To summarise, the point made in this section is that certainly, in a given text signs from one of my four codes could be replaced by signs from another, to produce a different text which might be roughly equivalent to the original - but that this does **not** show that any of the codes is entirely eliminable, that such a global replacement could be made on all mathematical texts at once. For each such local replacement has costs which would collectively pass beyond our means if generalised.

§40 What this example indicates about mathematical text

Summarising what we have seen in this example of mathematical text then, real mathematical texts are produced for particular purposes and therefore have specific contexts which must be taken into account. Such texts consist of various signs with a great variety of different functions, but which can be grouped into four systems for purposes of this discussion, W D N and P. I do not intend the word 'system' to bear any great weight here. In particular the logico-formalist notion of a system is much too cut-and-dried for real text. What I mean is that the signs found in a mathematical text can be assigned to one of these four groups by reference to their function in the text, and that the signs in each of these groups show considerable collective organisation in their use - which I shall describe in the next four chapters. Mine is a functional classification and equiform tokens are found sometimes in one system and sometimes in another; but we shall see in the following chapters that signs typically have a main allegiance to one of the codes. I have also pointed out that these codes are written ones. These points suggest the following reasons for looking at some more actual texts, as we shall.

First, it is clear that we need to examine other pieces of text and make comparisons of the way different kinds of signs function in them. Situating this first text vis-a-vis others will further justify my claim that it is in fact a good

example. Second, we need to consider in some detail the kinds of sign-function which I have pointed out and try to get clearer the differences between the various kinds of signs used in mathematics - Written words, Diagrams, Notation, Paragraphy, and their relation to speech and other signs. Thirdly, we need to consider in more detail the uses of mathematical text so that we can say more about the necessary preconditions for its successful use. Unless we do all this the fact that experience with texts and a mathematically formed mind are needed to use a text will be lost through looking at texts in isolation from one another. And we shall be in danger of misinterpreting mathematics because we concentrate on constative aspects of text to the exclusion of many others which are equally important.

I can now explain, as I promised to in §31, why the sample is actually a *good* example of mathematical expression. Most of what I have said about it is true of the other examples of texts in the exhibits, but many texts do not contain all four kinds of sign. This example shows all four of the kinds of signs which are important in mathematical text; some others do not, but I do not accept that because individual texts work without, say, Diagram signs, that mathematics could exist entirely without them. I shall be arguing that individual mathematical texts can only be **cognitively effective** in a context which includes texts utilising all four of the sign-systems I identify. Many texts from advanced pure mathematics are misleading as samples of mathematical expression because they exclude diagrams. Another way in which they can be misleading, is in tending to imply that examples in which mathematics is applied are not really typical, but as involving further and irrelevant issues. This is a misleading viewpoint, because it disguises as a problem of philosophy of mathematics what is really an artefact of a redefinition of mathematics. Mathematics includes texts which are more or less applied; it is misconceived as something which is approximated in pure texts and applied in others.

In the next five chapters, I will examine in more detail Writing, Diagrams, Notation, Paragraphy and the mathematical context, expanding on the points

made rather briefly in the present chapter. But first, I turn to several accounts in the literature which might be expected to throw some light on the four codes I have identified. Unfortunately, they don't. First, I look briefly at what the logico-formalist hegemony says about my example; then I consider Nelson Goodman's account of sign-systems; finally I consider briefly some recent attempts in mathematical didactics to conceptualise text.

§41 What the logico-formalist hegemony says about the sample text

This section will be brief because the logico-formalist hegemony says very little about a text like mine (which is one of the reasons I chose it). Since it contains little formal deduction, in the logico-formalist view it contains little pure mathematics. The arithmetic it uses is an application of some pure mathematics which can be found expressed "rigorously" in *Principia Mathematica* and related systems; for the rest, particularly the Paragraphy and Diagram, it is psychological padding which is of no relevance to mathematics proper. The application of the arithmetic is secured by the interpretation of arithmetic signs, for example the term '15000' of pure arithmetic is applied through the sentence W2 which (somehow) makes this term apply to the mean of attendances at concerts. To make this clear would require a considerable reformulation of the text since the 'mean' attendance is itself an application of a highly derived term in another mathematical theory, that of probability distributions. But the logico-formalist hegemony is largely content to leave it unclear, since it is not really relevant to the characterisation of (pure) mathematics, rather it can be labelled as the application "problem".

The reason why I cannot accept that this line of thought can provide a satisfactory philosophy of mathematics is that it implies that there is no mathematical knowledge to be accounted for which this text is used to create. The pure mathematics applied here is simple and already known to the reader, the application is not really mathematics. Yet it seems to me that when this text works, as it sometimes does, the reader has learned some more mathematics, which is

roughly stated in W1 as "If a population is known to have a normal distribution, and its mean and variance are known, we may use the table to express facts about this population". I say 'roughly' because the **method** of doing it is the novel knowledge, and this is not conveyed by this statement. Here we have an example of mathematical knowledge which is not well thought of as a proposition which the reader has come to believe, because the test of its being known is whether the reader can do the exercises.

In any case, the logico-formalist hegemony makes no attempt to account for the role played by the different sign-systems and the context in the working of such a text, so I will go on to some alternatives which might. I shall return in the following chapters to some of the assumptions underlying the logico-formalist hegemony, in particular the assumptions that Diagrams and Paragraphy are dispensable, and Written word reducible to Notation.

§42 Goodman's concepts of notation, character and score.

There are few recent works by philosophers in English which consider rigorously and in detail signifying systems other than speech. There is a good deal of literature on semiotics, semiology etc, most of it continental in origin or inspiration, but even the most valuable of¹⁴ it is difficult to relate to the usual concerns of the philosophy of mathematics. In the following chapters I shall present a framework with which I can satisfactorily discuss mathematical text context and use; in this chapter I shall consider the framework put forward in Goodman's *Languages of Art*¹⁵, because although I largely disagree with its theses, and its major interests are irrelevant to my concerns, all the same it provides the best way to introduce some important issues in short order. First, its clarity and rigour provide definite theses to deny or support. Second, Goodman's framework is essentially the logico-formalist system framework which dominates thinking

¹⁴ Cf the works by Eco, Barthes and Sless in the bibliography, and the collection Innis [1986].

¹⁵ Goodman [1981]

about signification in mathematics. Third, Goodman is concerned with some of the same items, such as buildings and their plans, which I place in the mathematical world¹⁶, though his interest is in their being in the aesthetic world. These worlds overlap of course - unlike Goodman I believe there is just the One Real World.

Although Goodman makes "notation" the central concept in his discussion, he does not have the same quarry in view as I do here. There is no doubt that a formal system should, if anything does, count as a notation in his terms, but the paradigm example for him is surely musical notation. His discussion is presented as a general discussion of "symbolsystems", so it should certainly be usable for the system which mathematics uses, and he does make some remarks about diagrams and texts. But I shall argue that his approach cannot help us, because despite his (over)emphasis on the conventional element in signification, he takes no account of the variety of purposes which lie behind the conventions we use, and the consequent variety of functions which signs have. In the end his account of "symbolsystems" is another version of the picture theory of meaning; which is highly ironic considering the aesthetic applications he makes of it.

Goodman's framework can be described briefly as follows. A *symbol scheme* is a set C of classes of *marks* and rules for compounding them¹⁷. Elements of C are called *characters*. A *symbolsystem* is a symbol scheme *correlated with a field of reference*. *Correlation* is used here to mean that there is a relation of *denotation* of objects in a certain class by characters, whose converse is called *compliance*. Denotation is "taken somewhat more broadly than usual"¹⁸. *Compliance* is a primitive relation between an object and a mark. A *field of reference* is simply a set of objects (one infers - there is no discussion of this), such as performances or pronunciations.

A *notation* is a symbol system satisfying five desiderata:

¹⁶ See Chapter 9 for a discussion of this.

¹⁷ Goodman uses 'symbol' where I use sign. While discussing him I will use 'symbol' and revert to 'sign' afterward.

¹⁸ Goodman [1981:143]

unambiguity: no mark may have different sets of compliants at different times or contexts; no character may have marks with different compliance classes.

syntactic disjointness: characters must be disjoint sets of marks under the equivalence relation of same-character.

syntactic differentiation: there must be no mark *m* and pair of characters *C* and *C'* for which it is theoretically impossible to determine that either *m* does not comply with *C* or else that *m* does not comply with *C'*.

semantic disjointness: the compliance classes of any two marks must be disjoint.

semantic differentiation: there must be no object *h* and pair of characters *K* and *K'* for which it is theoretically impossible to determine either that *h* does not comply with *K* or else that *h* does not comply with *K'*.

A *score* is a character in a notation. A system is called syntactically (semantically) *dense* if syntactic (semantic) differentiation fails to hold for some characters. If dense nowhere, it is *articulate*.

To distinguish them from what I would call notations, I will call sign-systems which satisfy Goodman's criteria for a notation "Goodman notations".

§43 Goodman on diagrams and icons, and language

Goodman briefly¹⁹ applies the machinery which he develops to the mode of signification of diagrams. He replaces the usual contrast between analog and digital which is often used here, by a symbolsystem's being dense or articulate (in his technical sense), syntactically or semantical, giving four possibilities. A symbolsystem is diagrammatic if syntactically articulate; diagrams are pictures in a syntactically articulate system. But even if this permits him to say that only some diagrams are notational, because such a system may still be semantically dense, this does not seem to help with the questions which I wish to ask: what role (functional, pragmatic, social) do diagrams play? Are they necessary for the texts in which they occur to play the cognitive roles they are constructed for? Can they be

¹⁹ Goodman [1981:170-3, 218-221 and 225-232]

replaced by formulas without cognitive loss? Even where they are notational and so **might** be used to denote, we might still wonder whether in fact they **are** so used - or whether some other logically anterior office is being filled by them, or even some illogical ones. And of course, most diagrams cannot be classified in this fourfold way at all since they are what he calls mixed - they involve signs from different systems such as lines and letters together (the diagram in the text which I discussed above is an example, as are most of the others in the exhibits). Moreover most diagrams have some parts that denote, others that depict - is that "mixed"? Surely only in a different sense. Line segments can be significant without referring as do their endpoints. Some diagrams have no denoting parts, despite being digital: as in Euclid, perhaps. But we would like to know how mixing is possible, and what the total effect of such mixed signs is. Indeed, my view of mathematical text makes just this a crucial issue. How is the function of the text related to those of its constituents?

Goodman deduces from his analysis that the oft-cited distinction of Peirce between indexes, icons and symbols has no force. The icon is distinguished from the symbol, according to Peirce, by its sharing properties with its referent. Goodman's rejection of this follows from his earlier argument that picturing does not require resemblance (which has been disputed by many, and I think refuted by Novitz²⁰); he reiterates²¹ the distinction between the notational and the dense scheme, but argues that within the latter class there are only differences of degree - i.e. of diagrammaticity. And the former sharp distinction does ~~not~~ depend on the internal structure of the individual sign, since this may describe in one system but depict in another. Thus Goodman relies entirely on the discreteness of a descriptive system to differentiate it from depicting ones. Yet one must ask whether the holding of his criteria can really be independent of the internal constitution of the symbols used and their users. I shall argue below that the nature

²⁰ Cf Novitz [1977]

²¹ At [1981: 229 ff]

of its users impose constraints on what marks can be used in a sign-system, and how. Goodman is right to emphasise the system over the individual sign, but it seems implausible at least to say that when a depiction has structure that this is irrelevant to the manner of depiction, or to just what is depicted.

It is not always clear whether Goodman is considering diagrams primarily as objects or as signs; paintings clearly are objects and one can discuss whether a notation could be invented for them as a field of reference; but Goodman seems to vacillate between considering whether a diagram **is** a sign in a notation, and whether it could be annotated, i.e. **given** a sign²². Goodman seems to shift uneasily between describing the semiotic system of painting and scouting the possibility of notations for paintings. But as a sign, a diagram is for Goodman a character in a descriptive system. He does not seem to have anything in his account to cover the contrast between the functions of diagram and word which I have remarked in the sample text, and which is obvious in many others. But this is not surprising given Goodman's assumption that the function of a notation is identification.

Since Goodman takes the "prior logical office" of a notation to be identification, it is not surprising that his concept of notation is really a concept of a naming system or nomenclature. What is surprising though is that he should suppose that such a system even *could be* a language - though he is quick to deny that his book's title is serious, he equally plainly thinks that notational systems are languages²³; it is the non-notational systems which don't merit this title, being too dense. It seems, though, that many of the features of language are lacking in other symbol systems: we cannot state truths in diagrams, ask questions in musical notation, give definitions in Labanotation or orders in decimal numerals. We cannot warn or greet or inveigle or shock or castigate or slight someone using musical notation. We can use these systems in conjunction with speech to do these things perhaps. But I deny that any speech act other than naming has been provided

²² Cf Goodman [1981:116]

²³ See p xi. On p194 music notation is called a language, on p200 "notational systems" and "languages" are used interchangeably.

for when we have a notation in the sense of Goodman - and even that has not been properly explained, since we have no axioms about conformity of the notation's classification with preexisting ones, nor about a user's (conversational) duties to observe prior uses in extending it, or even just in using it. I will try to substantiate these various claims now by examining Goodman's argument about forgery.

§44 Goodman's forgery puzzle and his solution of it

There are many cases, some very celebrated, of forgeries of paintings. There are not, indeed cannot be according to Goodman, cases of forgeries of musical works of art. "Of the London Symphony [of Haydn], unlike the Lucretia [of Rembrandt], there can be no forgeries." ²⁴ Why this difference? It is to explain this difference that Goodman makes his analysis of notation. Goodman claims that notations are those symbol systems whose "logically prior office" is to permit identification of a work²⁵. His analysis of symbol systems and criteria for notation systems are aimed at distinguishing such systems from other kinds of symbol system.

Two purported instances of a painting cannot both be genuine. Only one can be that painting made by painter A at place and time B, for example, the Lucretia. If the other is a painting made by A but not at B then it is a different work. If it was not made by A at all but is represented as being so then it is a forgery. The difference between P (the Lucretia, say) and P* (a forgery) is in the first instance that they are different works. (Goodman spends some time persuading us that this difference is an *aesthetically* relevant difference between two objects which may be perceptually, though not historically, at present indistinguishable; but this is not a *logical* problem and I do not consider it here.) By contrast, two purported instances of the London Symphony might well be both genuine. "All that matters is what may be called *sameness of spelling*..." ²⁶ That is, we decide that a purported instance is

²⁴ Goodman [1981: 112]

²⁵ Goodman [1981:128]

²⁶ Goodman [1981:115]

genuinely an instance if it *complies with* a certain specific *character* ("the score") in a certain *notation*.

Goodman's solution to his puzzle about forgery is this: a work can be forged if and only if it is a unique object of which there cannot be a true copy because its origin partly constitutes it for us. A forgery is then another object which may be fraudulently represented as the real one in virtue of some similarity²⁷, if their provenances are ignored or unavailable. By contrast, a work cannot be forged if there are multiple instances of it which are identified as such by determining whether they comply with a character in a Goodman notation. In this case the instances are *in the relevant respects* identical; in the other, two purported instances cannot be. The fraudulence aspect is irrelevant to the logic here; the essential point is that a work has instances if and only if there is a Goodman notation for which it is a compliance class of a character.

§45 Difficulties with this solution to the puzzle

Even in its own terms this solution is seriously inadequate. First, surely it is clear that only certain possible "notations" can fill the "logical" office Goodman has identified. The idea of a spoken musical score is absurd, as is that of one based on smells or taste, or one using sculptured blancmange. We could not use such scores to produce works - though Goodman declares that to be irrelevant. But in fact we could not use them to identify works either, which he regards as essential. He implies that we could not identify with such systems by his (rather inconsistent) use of 'inscription' as a synonym for 'mark'. A real notation needs some of the properties of writing even for its identificatory function.

Second, the "same" chess game can be played on more than one occasion, each such event complying with a certain score in chess "notation". But these games are not instances of a work; on the contrary, the players involved would surely claim that each game was "all their own work"! Thus here we have a "notation" in

²⁷ There's something awry here, given Goodman on similarity!

which different works comply with the same character; must we conclude that chess "notation" is not notation? This seems to follow, yet all the criteria are satisfied. Again, in mathematics most mathematicians would agree that what is the same proof can be specified by more than one character in a given notation (or formal system as it is usually called there.) Finally it worth mentioning the parable of Jorge Luis Borges, "Pierre Menard, author of the Quixote" about a modern writer who with immense labour writes a work which is word for word the same as Cervantes', but clearly much deeper, more ironic etc etc. Here we have the same score and different works again. There is no forgery of chess games or proofs or the Quixote in these examples as described. (But there might be if the originals are used to produce the copies!) In these examples we have a Goodman notation in which the members of a compliance class are *not* instances of a work.

Third, cases where performances of the same work really relevantly differ pose some problems: cadenza and jazz for example. These are cases where scores of two performances can be classed as the same despite being not samespelled, if other factors outweigh this one. This is possible because not only can performances be copied directly, but one can transform a score in using it for performance: interpretation cannot be confined to inessential uses as Goodman implies²⁸. On the other hand, any piece with a cadenza gives us instances of the same work which are not samespelled in the notation - though they may be in some logician's "real" notation (here we can see the logicist rewriting of mathematics lurking!) To deny as does Goodman²⁹ in response to this problem that musical notation is really notational is to get the cart before the horse: performances of a folk song cannot be forged even though no notation is in use for such works. Here we have instances of a work which could not be members of a compliance class of a notation which actually described them. (Names of pieces don't count as a Goodman notation - but why not?)

²⁸ Wanda Landowska is reported to have commented on someone's playing of a Bach piece, "You play it your way and I'll play it Bach's."

²⁹ Goodman [1981:184]

Fourth, suppose two children playing a secret game devise the following code to assist them: any kind of whistle is to indicate the approach of an adult or other source of interruption; a shout is to indicate that the shouter is tired of the game and is now doing something else; raising a flag requests and thereby effects a reversal of roles in the game. On the Goodman view there would appear to be three "works" here (interruption, relinquishment, reversal) which can be tokened and identified when they occur readily enough; but would anyone really agree that there are instances here in any real sense? The items grouped together as compliants have little in common. Indeed we can suggest far more bizarre codes now that we have the idea: let any A stand for my left foot, Aldebaran or Nelson Goodman's salary; let any B stand for the Duino Elegies of Rilke, the crushing of a certain ant in Ecuador on 1/1/1001 (details from the author), or Rembrandt's Lucretia. This assignment satisfies all the criteria for a notation according to Goodman, but only in the most strained logical sense is there any kind of tokening involved here. Worse, it would seem that now it is still possible to forge the Lucretia yet it has genuine copies! Here we have the possibility of Goodman notation where we feel that we simply can't accept that the instances are instances of anything real.

These considerations show that the biconditional (Goodman notation \leftrightarrow instances) doesn't really hold up. Goodman's solution rests on an insufficiently clear notion of identification of a work, and does not sufficiently explain why forgery can or can not occur in less usual cases. So we must look closely at his concept of identification.

§46 The error underlying Goodman's conception of symbolsystem

There is a basic error in Goodman's analysis, which underlies the difficulties I have been raising. There are several ideas conveyed by the term 'identification', and the one which Goodman captures is not the only one for which real notations are actually used. The correct sense of identification cannot be separated from that of production and other uses of notation as he demands. The

multiplicity of *purposes* for notations is part of what they are, and Goodman's uniform interpretation of their semantics is inconsistent with this basic pragmatic fact.

Goodman claims that notation has a unique logical office, identification, and that other roles it may play, such as in production of works, should be ignored for logical purposes. But there are at least two different concepts of identification which must be considered. Names are used to identify individuals, in the sense that the name and the individual are made in some sense one, and the name can in some places do duty for, stand for, the named. The ideas of picking items out and of re-identification are involved in this sense of identification. The logic of Goodman's system is meant to justify this inference: if **a** and **b** comply with **c** and **d**, and **c** and **d** are marks of the same character in a notation for things of the kind of **a** and **b**, then **a** and **b** are tokens of the same type. (But to put it like this disguises the volition and interests at work in the network of praxis in which characters are recognised as the same, and objects recognised as complying with them. In identifying **X** to you, I pick it out and label it; both our intentions must be engaged for this to happen.) In another sense of 'identified', however, we can say that the performances classified together as of a particular work are identified, meaning just that: they are accepted or taken as the same *qua work* - we are dealing with relative identity or identity under a sortal now. The aim of *identifying* in this sense is information: we wish to locate the item in question vis-a-vis others. In this sense performances are not picked out or re-identified by scores, they are *classified*. There is a parallel to this too in mathematics in that precise definitions are needed for items which are to be used via a name or via a predicate in subsequent constructions. To identify **A** as **W** is to certify that **A** is the object **W**, but to define **A** as a **W** is to certify that anything called **A** is an instance of **W**. In short, Goodman's system conflates the ideas of name and predicate.

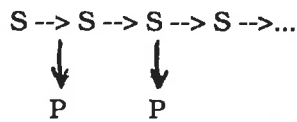
Goodman claims that the prior logical office of notation is to enable identification, and warns against being misled by the use of scores in production.

But contrary to Goodman's view, museum curators have perfectly good schemes of notation for identifying the art works in their charge (they better have!); that is, if names is really all that is needed. Labels with arbitrary numbers will do. But such schemes are of no use for reproduction (just as well for the economics of the artworld!). Goodman might reply to this that such "notations" do not *guarantee* identification as do the ones he has in mind. Indeed; but here we have just arrived at the old controversy about identifying descriptions and proper names in a new guise. Only logically proper names give such guarantees, one old view had it; Goodman notations are intended to be systems of just such nomenclature. But it is really because such a scheme can guide construction that it may give such guarantees as we have in this life: because it is **descriptive**. So I think that despite what Goodman says, it is not possible to have a notation which will fill the logical office of identification without its being used for productions.

It is clearly not enough that there be a possible notation for works to be identified by it; it is necessary that it be *used* to make identifications. It is clear from Goodman's discussion³⁰ that the notation must actually be used for the works to be constituted. But then it must be actually, not just theoretically, possible for the users to use it, at least for identification, as he says, but also for production which he denies. It is misleading to say that a mark denotes a performance even if no such mark was used to produce it; it may be true, but it does not follow that this could apply to all the performances of the works in a realm of the kind we are concerned with. Indeed, I would suggest a strong counterclaim: we only recognise a performance as an instance of a score if it fits into the causal tree of performances linked together by marks of the score (and perhaps some related scores). A symptom of the way Goodman is distorting the role of a score is that he gives the impression that chains of copies go like this:

³⁰ Goodman [1981:197]

$S \rightarrow P \rightarrow S \rightarrow P \rightarrow \dots$ [S score P performance]; the discussion eliciting his 5 desiderata for notation rests on the need to ensure that chains like this don't go awry. But in fact the structure is usually more like this:



That is, copies of a score are made in general from other copies, not from performances; performances are made from scores but not vice versa. There is good reason for this: performances are not correctable, whereas scores can be proofread, edited, definitive versions established etc. And very few performances are reliable enough to use them for writing down the score. A score is used in order to enable the making of an instance of the specific work we are interested in, it *specifies* how the performance must be. It must be used in conjunction with unstated conventions for its use which enable the players to produce an instance of the work. This aspect of a score resembles the Kantian/Intuitionist way of thinking of proofs in mathematics as guides for the construction of desired mathematical objects.

The reason that we don't accept as forming a real symbolsystem the classes of marks and objects in the bizarre cases which I invented in the preceding section, is that the marks have no such causal tree connecting them via uses of the notation. This is required for us to recognise a type; moreover it is only made possible by our use of abilities to recognise similarities in the production of tokens from scores³¹. So, contrary to Goodman, scores must be actually used to produce performances for the scores to denote them; so scores must be apt not only for identifying performances but also for producing them. Therefore a possible notation is not any logically possible one but one which we could possibly use for this purpose.

The problem arises because Goodman makes no restriction as to what may be a mark. He does not tell us what a mark is, but we are told that marks include

³¹ Cf Goodman [1981:198] on the need to accommodate to any preceding classification of works.

inscriptions which in turn include utterances. But also "an inscription is any mark - visual, auditory, etc. - that belongs to a character." All this is on page 131; it seems to be rather confused - surely utterance is a *wider* term than inscription. Yet for scores to be usable for production, marks must be persistent compared with their compliants, be visual objects, and be essentially two-dimensional. In short they must be *written*, in a slightly extended sense of writing, which I will clarify in the next chapter. Goodman ignores the pragmatics of mark use - but it actually follows from his own requirements for a notation that one cannot. This claim may seem to have force only in that useful notation systems for creatures like *us* must be suited to *our* perceptual systems, whereas Goodman is concerned with the "logical role" played by any notation. But this conflicts with the need to give some content to the "possibility" which appears in his differentiation requirements: recall that "there must be no mark *m* and pair of characters *C* and *C** for which it is theoretically impossible to determine..." Theoretical possibilities must be ones we can deploy *in production and identification*. The basis of the individuating power of a notation is the individuation of the characters: so the first thing we need to know is how this is obtained, if we are going to explain the 'can' in the statements of his criteria. It's really the scores that are the same, not the performances! This point is significant for the parallel with mathematics: logico-formalism is an account of how proofs may be identified, not of the alleged objects which they allegedly describe.

Goodman can hardly avoid taking the line he does because he has rejected resemblance as an element in depiction in his first chapter. So he has no *means* built into a score as he conceives it, for its use in production. By excluding the mode of production of performances, by rejecting resemblance as an element in depiction, Goodman restricts symbolsystems to a single semantic relation between mark and object. Though he admits in his preface that his book should be called "symbolsystems of art" (plural), he simply assumes that there is a single general relation of symbolisation or *reference*, denotation in one direction and exemplification in the other, which can be ascribed to all kinds of symbol. He

distinguishes "representational" systems, which permit depiction, from languages, which allow description, as two species of denotation³². The differences largely can be traced, he suggests, to the satisfaction or otherwise of his semantic requirements on symbol systems. Goodman does not systematically discuss this assumption that there is a single relation of reference in any symbol system. I believe that it immediately makes his account inadequate for language, because predicates do not refer, nor do sentences. Goodman does not actually apply his theory to natural languages; when he discusses written texts he treats them simply as complex names for works of "literature" and does not consider such linguistic items as signs themselves. In this he conforms to the standard view of writing as no more than a means to record speech. Nevertheless, as I have argued, he has really only given an account of *naming* systems.

Now for his ostensible main purpose of characterising those systems which can fulfil a certain logical role, it might not matter that these two restrictions (one relation, one role) confine his interest to systems of nomenclature. But for the wider conclusions which he also draws from his account they are important. In particular, although it is true that assertion can be added or subtracted, in a sense, from any system (pictures can be used to assert), this does not prove that assertion is not originally and basically a speech act. Indeed *reference*, as Searle has convincingly argued³³ is a speech act, and Goodman's account fails precisely because he does not take account of this fact. A description is only effected by attribution, so that a score does not in itself describe a performance, though it might be said to depict it; but I **might** use a performance to describe a score. (Goodman admits this paradoxical possibility on p231). In any case, signs with syntactic roles other than naming, for example predicates, need an account if language is to be covered. Goodman seems to me to subscribe implicitly (in this work) to the standard

³² Goodman [1981: 41-43 and 225-6]

³³ Searle [1969: ch4]. After my next chapter, it will be clear that I would say rather that there is a speech act of reference - but there may also be other kinds of acts of reference.

logico-formalist hegemony account of signification, his shortcomings are some of its own. On his account there could be no bar to the the complete translatability of a sign in Words to one in Diagram. Conflating the functions of reference and predication, Goodman is naturally going to conflate those of different signsystems like Word and Diagram; he is not prepared to allow the **point** of a signification to enter into his account.

All this discussion of Goodman is not intended primarily as a major criticism of *Languages of Art* as a contribution to aesthetics. That a discussion of the semiotics of diagrams must try to make something of such a scrutiny of a few pages of a work about aesthetics is a criticism of a whole other literature. Nevertheless, since Goodman's theory of notation is inadequate for describing mathematical text, I shall give a less elegant but more realistic and adequate account in the following chapters.

§47 Text in mathematical didactics

A few writers³⁴ have (quite recently) looked closely at the real features of texts used in mathematics, mostly writers concerned with pedagogy. The essential point on which they have stumbled (very belatedly, it seems to me) is that if learners have to cope with text as text actually is, then theories about learning and teaching must give attention to what the features of mathematical text actually are. The title of Morris's and Stewart-Dore's book³⁵ - *Learning to learn from text* - indicates both the nature of the interest in text which these writers have, and the limitations of their analyses. The point of their work is to help teachers to help students learn from texts. Students, students of mathematics particularly, learn using texts and yet the the specifics of the use of texts in learning have not been much studied and there is little by way of theory about it. Didactics is full of rather ineffective theories, so it is reasonable that writers who are aware enough to have

³⁴ Morris and Stewart-Dore [1984], Shuard and Rothery [1984], Otte [1986], Van Dormolen [1986].

³⁵ Morris and Stewart-Dore [1984]

noticed the role of texts in learning should be more concerned to find ways to assist with the composition of effective texts and strategies for coping with not-so-effective ones, than to address the kind of issues about the nature of signification which I am pursuing here.

Van Dormolen and Otte³⁶ are also writing with pedagogues rather than philosophers as audience, though there is clearly a rather heavily theoretical German Didaktik lying behind some of the work of the group to which they belong³⁷. They are alive to the presence of the various different sign functions in text which I have identified, but the extent of conceptualising they reach is only an adoption of Thom's contrast between geometrical and algebraic language³⁸, or the contrast of the verbal-numerical with the visual which is perhaps the same, deriving from the work of the anthropologist Leroi-Gourhan. This **polarisation** of the text seems to me to rely only on its appearance and to pay too little attention to how it actually works.

A similar inadequacy seems to attend the other work in this group³⁹, in which the "visual language" of mathematics, as they call it, is treated using Twyman's classification⁴⁰. The basis of this "scheme for the classification of graphic language" is the nature of the eye movement required to read the text. This is certainly a useful idea. But it *seems* to me that it rather forgets that many kinds of text, particularly mathematical text, are typically read several times, and not in any given manner either. One enters the page at various points and traverses different areas in different ways on different readings. Moreover it ignores the differences among the cognitive functions of the various sign-collocations to be found on the page. Perhaps this is inevitable in that the point of the text analysis in this work is to provide a guide to improved text construction - improved from the

³⁶ Van Dormolen [1986], Otte [1986]

³⁷ Cf their citations, particularly to the Bielefeld school.

³⁸ Cf Thom [1973]

³⁹ Shuard and Rothery [1984]

⁴⁰ Twyman [1979]

point of view of usability. A main aim of these authors is to make some kind of advance on readability formulas for prose which takes account of the real nature of mathematical text. Their aims are practical.

These criticisms are intended to indicate the relative uselessness of these works for my present purposes, not to decry it - far from it, it is the 0.1% of work on learning mathematics which may actually be useful. But it may need to be revised on the basis of a deeper account of signification in mathematics, even to be didactically useful. For present philosophical purposes it is useful only as confirmation that some other people have noticed the complexity of mathematical text and the learning processes in which it is involved.

Chapter 5

Writing

§48 Role of speech in mathematics

My first main point about mathematical expression is that it needs to be typically in writing. But this does not mean that mathematical discourse is not founded upon speech, and it does not mean that there is no mathematics at all in speech; so that we certainly need to have an acceptable philosophy of language in the usual sense. The most primitive mathematical terms, those for small integers and simple fractions and shapes (three, half, round) already occur in speech. Even prior to mathematics proper, there are in speech terms and concepts which have pre-mathematical content, such as the vocabularies of direction and comparison (up, down, faster, heavier). (This recognition is a healthy recent improvement in mathematical pedagogy over the poor child psychology informing 'new math'.) Finally, thanks to the phenomenon of inversion¹, some of the content of mathematical writing flows out in spoken versions - though I shall argue below that it must be some, not all.

Although I am emphasizing the differences between written and spoken language, because they have been largely ignored, all the same it cannot be denied that the one founds the other in a sense I shall try to get clearer. The written word component in mathematical text is largely to be understood as an extension of speech. However, the issue is somewhat clouded here by the way in which speech is adapted in order that mathematics can be read out, because some characteristic features of mathematics (proof, definition) are thereby disguised. The situation is complicated, but my point is that for mathematics as it now is, written expression is necessary, even though speech was and is required for writing. But speech is not the only parent of writing.

The prevailing image of mathematical discourse is that it is a sequence of statements which happen to be usually written, but that this is inessential: these statements are essentially like those made in speech. In this chapter I will show

¹ The simulation of a later form of expression by an earlier one on which it depends.

that in general, that is apart from special features of mathematics which will be taken up in the following chapters, the speech act framework is inadequate even for "statements" made in writing. I shall distinguish **text acts** from speech acts. Later I shall show that many text acts used in mathematics are not best thought of as transposed speech acts, nor as statements². But first it is necessary to have some preliminary clarification of just what writing is.

§49 Antecedents and development of writing

Some light is cast on the complexity of the mathematical sign system by a consideration of the origins and development of writing. In this section I will point out that there are several independent semiotic sources behind writing, and several different kinds of writing system.

There are two main theories about the origin of writing. The older theory is that the writing evolved from painting by schematization. The painting that evolved into writing was painting directed to particular purposes: the recording of the works of particular individuals, kings mostly, which is intimately connected with the promulgation of laws, the keeping of tallies and so on. But there is a lot of evidence of neolithic recording systems which long predate writing, by several millenia. The various kinds of tally stick are discussed in Menninger³; Marshack⁴ has argued that some of these were actually used for recording calendric information such as days between phases of the moon. (Menninger exhibits the accounting roles of such systems persisting well into this century.) Tally sticks are directly relevant to Schmandt-Besserat's theory of the origin of writing, which is much discussed at

² We must always recall the context of mathematical action; at many points in the complex web of mathematical practice, speech is used in the production, distribution and consumption of mathematical signs - in instructing students in differentiation, in assembly of a building from complex plans, in persuasion that a new proof technique is valid, in explication of cosmological theories on the Science Show, in political controversy about the interpretation of the inflation rate. This point will be taken up in Chapters 9 and 10.

³ Menninger [1969]

⁴ Marshack [1972]

present⁵. Contrary to the painting theory, which was previously widely accepted⁶, she derives writing from these stone-age recording systems. She proposes a five stage process involving "tokens" or counters, and "bullas". First, small moulded and baked clay figurines, which she mostly calls tokens, but which I shall call counters, were made to stand for herd animals etc for tallying purposes; second, holes were drilled in these counters which allowed stringing them together; third, clay "envelopes" called *bullas* were introduced in which a number of tokens would be sealed. These were exchanged in transactions and kept as a kind of accounting record. In the fourth stage, marks of the tokens were made on the outside of the bulla (to obviate the need to smash them and look at the contents in order to verify a particular transaction), at first by pressing the contents against the wet clay bulla before baking it, and later by incisions on it. In the fifth and final stage it was realised that the bullas were all that were needed - the contents had become unnecessary. The marks on the bullas were the first form of writing.⁷

It seems to me that one need not choose between these two hypotheses about the emergence of writing. The painting theory is well supported by Egyptian evidence, the bulla theory by Mesopotamian. Since what counts as the "first" writing is vague and contentious, I prefer to accept both of them as identifying complementary factors in a multicausal process. Notice that on either of these views, the earliest form of writing is the icon, the moulded or drawn substitute for the thing, **not** for the word. Recognising these two different forms of signification - painting and moulding - interacting at the birth of writing means that we can offer a hypothesis about why *phonetisation*, the shift from representing things to representing the words for them, came about. Gelb⁸ calls this the most momentous step in the evolution of writing. One paints *visible* objects, one forms *tactile* ones; when these two ideas are conjoined the possibility of representing in one modality

⁵ Cf Sampson [1985: 57-61], Bunn [1981: 83-5], Schmandt-Besserat [1978] , [1979]

⁶ E.g Gelb [1963, *passim*]

⁷ This theory clearly foreshadows formalism, and I will revert to it in Part III.

⁸ Gelb [1952: 12]

items sensed in another *or in none* becomes available. The visible sound, the painted word is possible. The need for it was already present in the need to represent what is represented in language but is not objectual - the problem of painting "the" "is" and "on", as well as "cat" and "mat".

Whether there is a testable hypothesis to be developed here is not my present concern; rather, I wish to emphasise the presence of several different kinds of signification at work in the constitution of writing. There is the picture, clearly lying behind early hieroglyphics and continuing in some ways in diagrams; there is the counter and its aggregation in the bulla, which continues in various calculi employed in mathematics over the millenia; there is the idea of the brand, tag or label involved in the tracing of the counter on the bulla, which I will connect later with variables; there is the ambiguity of enframing, in the transition of the glyph from something to look at to something to direct our looking and of the bulla from receptacle to background. And of course there is the signifying system par excellence, speech, being interwoven with all this.

§50 Characterisation of writing

There are basically three kinds of system which have been used in recorded (!) history for representing words visibly. They are the logographic, syllabic and alphabetic systems. It is often suggested that these three kinds of systems form a progression, but this is a rather 19th century notion, refuted by the continued vitality of systems of each distinguishable kind. Chinese logography is alive and well! This is not to say that we won't learn a great deal relevant to philosophy of mathematics by considering why transition from one kind of system to another might have occurred, since western writing did evolve through these three forms. (Logography is hard to use on present computers, or on typewriters.) Writing proper begins with the invention of the idea of representing a word rather than what it stands for; the advantage of this plan is that words without physical referents can be included. The first form of this idea rests with an arbitrary mark for each whole

word, i.e. logography. The second form of word representation links marks and syllables, that is words-that-might-be within words. The third form is that of the alphabet, in which each sound segment in a word is given a sign. It is also possible to represent the *features* which collectively distinguish a segment, but practical systems of this kind are few and recent, such as Pitmanscript⁹. So, the written speech sign may represent a sound (a phonograph), either a segment a syllable or a feature, or may represent a word (a logograph). The base repertoire of graphs may be a small alphabet, a medium sized syllabary or a large character. In the latter case there may but need not be some syntactic structure within the character. As ways to represent speech, all systems include the basic formation rules that graphs may be built up by linear juxtaposition, to mimic the "line" of speech. The simplest system, of a single winding line, as in the Phaistos disc or in Greek boustrophedon, is generally supplanted by the principle of the standard direction of writing (though the direction may be any of the four orthogonal possibilities). This introduces the "line" as an element of text quite absent from speech; the interplay between line and speech contour provides material for written poetry, and leads on to the ideas of the paragraph and the stop.

The fact that mathematics involves signs with no counterpart in speech (an important example is the bracket) suggests that we should consider just what is the relation of writing and speech. Until recently linguists denied that writing was language, seeing it as merely transcription:

"Writing is not language, but merely a way of recording language by means of visible marks." ¹⁰

*"...the spoken forms alone constitute the object....People attach even more importance to the written image of a vocal sign than to the sign itself. A similar mistake would be in thinking that more can be learned about someone by looking at his photograph than by viewing him directly."*¹¹

⁹ Cf Sampson [1985: 40-2]

¹⁰ Bloomfield [1933: 11]

¹¹ Saussure [1959: 24]

The general insistence on this view in linguistics, excepting the Prague school¹², has been attributed to its need early this century to establish itself as a scientific discipline, distinct from grammar, with actual usage as data.

In opposition to the long neglect of the specifics of writing by linguists, a number of writers¹³ have recently insisted that writing should not be thought of as merely a method of transposing speech. That written languages exist in their own right and are not mere transcriptions of spoken language is argued in several ways. First, numerous small differences can be pointed out between the features which are "marked" in the spoken and written forms of specific languages such as English¹⁴, grading to cases where the written and spoken languages are clearly entirely different, as in Chinese - where until recently written texts read aloud could not be understood¹⁵. An even more extreme example of this kind is learned Latin - which persisted without a spoken form for 500 years¹⁶. Second there is overwhelming evidence that the first uses of writing were for lists and related signs not found in speech¹⁷; writing was never used *only* to render speech visible, and at first it may not have been used for that at all (see next section). Third, it is possible, and there are well documented cases, of persons learning to read without speech, for example, Helen Keller; so there is no logical dependence requiring writing to be acquired after speech. I take it that the minority viewpoint is now sufficiently well-established to be adopted, but in any case I do not need the strong thesis that written English is a distinguishable language, only the weaker thesis that it is not a mere transposition of spoken English. Nor do I actually believe that any of these arguments shows that writing could function as a *language* except in a context of speech. By *language* I mean any sign system adequate for communicating the full range of human

¹² Cf Vachek [1973].

¹³ Cf, for example, Gelb [1952: 11-20], Derrida [1974], Stubbs [1980: 21-42], Ong [1982: 78-115], Sampson [1985: 26-32]

¹⁴ Stubbs [1980: 32-40]

¹⁵ Sampson [1985: 27]

¹⁶ Ong [1982: 112-115]

¹⁷ Goody [1977]

concerns; on this view all human spoken languages are indeed languages, though some are less generally useful than others. One reason for this is lexical; English has a much larger vocabulary than many unwritten languages. The other is the extensions to speech which support English. Computer "languages" are not languages at all in my view; they should be regarded as notation systems.

Recognising that the mirror view is wrong, Gelb and Sampson try to give more inclusive characterisations of writing. Thus Gelb says ¹⁸:

"Writing is ... a system of human intercommunication by means of conventional visible marks ...";

Sampson says ¹⁹:

" [to write is] ... to communicate relatively specific ideas by means of permanent visible marks."

Formulations like these do not assume that the sole purpose and function of writing is to transcribe speech. However they are by no means unproblematic and I shall try to improve on them. The ideas they adumbrate are communication, convention, visible, marks, permanence and "relatively specific ideas". Let us consider how necessary each of these is for writing.

I have already indicated that I take it for granted that writing is a communication system. Unlike Sampson, Gelb insists on *conventionality* as a criterion of writing ; this is because while he wants to include early systems close to painting, he also recognises that "fully developed writing expresses linguistic elements" [1952:13]. Conversely, Sampson requires "specificity of idea" (not a very specific idea!) to *exclude* painting. Recent discussions of convention²⁰ show it to be rather a two-edged sword for distinguishing picturing from describing; I shall discuss this issue below in connection with icons; I take it that convention is involved in all sign-systems and will not discuss that any further here. So both

¹⁸ Gelb [1952: 12]

¹⁹ Sampson [1985: 26]

²⁰ For example, Goodman [1981].

conventionality and use for communication I take as characterising *all* sign systems and having no part in distinguishing writing from speech.

Sampson argues for *permanence* as a characteristic of writing to exclude gestural systems; Gelb does not consider permanence as part of his definition, but makes it clear that he is only considering systems which involve putting relatively enduring marks on objects²¹. How permanent is permanent? This is actually a very interesting issue, the advent of VDUs changing the game recently. For now, let us say that permanence is relative to the transience of speech, and that there will be significant differences between speech and writing if this is so. Early writing employed a variety of media (clay, papyrus, walls, stone columns), perhaps partly because it arose from a variety of earlier systems with a variety of purposes. More detailed investigation of the consequences of permanence would need to address the differential permanence of different media.

Sampson, Gelb and I all agree that writing must be visible. Neither of them says so, but I think it is also necessary to point out that writing is (phenomenologically) two-dimensional. The planarity of the writing surface has important properties which I shall discuss later.

Sampson, Gelb and other writers distinguish between *semasiographs* and *glottographs*. A semasiograph represents a thing or idea directly; a glottograph represents a word. This is not a good way to make the distinction if you wish to deny that written languages have independent lives; and the concept of representing an idea, or even a thing, directly, is rather obscure. I will use the following terminology. A **sign** is an iterable object conventionally used in communication. Signs occur in *systems*, being individuated by contrast with other signs. I call a *permanent visible two-dimensional sign* a **graph**. A sign system using graphs is a **graphic system**. I accept the term *glottograph* for a graph which originally represents a word and thereby derives its meaning. Other signs in a graphic system can be called *semasiographs* by contrast. Gelb argues that semasiographic elements

²¹ Cf Gelb [1952: 6]

are always found in writing, particularly in its early forms; he is rather two faced about whether writing includes both, wishing to do justice both to the origins of writing and to the glottographic character of "fully developed" writing. Sampson prefers to include both in writing, because semasiographic systems are important in their own right. He regards mathematics as such a system. (That this is an error will become apparent as my own view is developed; rather, mathematics is a semasiographic *extension* of written speech.)

In any case I wish to keep these different kinds of system clearly separate so that we can examine their interaction; so I am going to restrict the term **written speech** to glottographic systems, those representing speech. I shall use **writing system** to refer to a sign-system such as that of mathematics which extends written speech with semasiographic signs. I shall use **written language** for any language expressed in graphs, of whatever origin. The relation between these three concepts is this: written languages in fact have arisen from written speech, but have obtained considerable independent life in the course of time; a written language need not arise in this way, as the use of semasiographic systems shows; but in fact semasiographic systems play the role of extensions to written speech. I accept that semasiographic languages might exist, though I doubt the emergence of any purely semasiographic language into a world so speech-sodden as ours; but I shall be arguing in Chapter 7 that it is a great error to think that mathematical notation is or could be such a language. I call **Written word** the written speech system employed in mathematics, which for most texts uses an alphabetic system of transcription. The **mathematical signsystem** is a writing system which properly extends Written word with the semasiographic systems of Diagrams, Notation and Paragraphy which will be discussed in the next three chapters below. In the rest of this chapter I consider further the nature of Written word.

Sign systems, such as magnetic tape, which require tools for the writing and reading of their signs, might still be called writing in a sense broader than mine. They are excluded by my characterisation because the signs of such a system are not

visible, and because they are not languages: they are mark systems but not graphic ones. Braille is an interesting case: the marks are visible but not as signs; as signs they are tactile. So it is not a writing system in my sense. We see that my characterisation is in fact somewhat restrictive; this is compensated by its giving emphasis to systems excluded by other conceptions. It will be apparent from the discussions which follow why I insist on visibility.

§51 The cognitive significance of writing

Over the last twenty years, a number of writers have pointed out the great cognitive significance of writing. Goody, an anthropologist, has written a number of books and papers on the subject²². Derrida, a philosopher, has written extensively about the role played in Western thought by the "repression" of writing²³. Ong²⁴ has extended a tradition in literary studies stemming from Parry and Havelock making similar points. One of Ong's chapter titles puts it well: *writing restructures consciousness*. A more general thesis about the effects of technology on thought became briefly fashionable through the exciting but somewhat incoherent writings of McLuhan. He was a student of Harold Innis, whose *Empire and Communications* was a pioneering work of this kind²⁵. Recently Eisenstein²⁶ has written a wonderful and extensive historical study of the cognitive consequences of the introduction of print. Earlier historical studies like Lefebvre's²⁷, while fascinating, make too little of the cognitive aspects. I have no space here to even outline the ramifications of these very wide-ranging works. Rather, I will demonstrate this general thesis from a point of view which will enable me to

²² Goody [1977], [1986]. Goody and Watt [1962] is a seminal statement of the general idea.

²³ Derrida [1974], [1982].

²⁴ Ong [1982]; Havelock [1963] identifies the formation of Greek philosophy and the emergence of alphabetic writing.

²⁵ McLuhan [1964], Innis [1972]

²⁶ Eisenstein [1979]

²⁷ Lefebvre [1976]. The last chapter, on the book as a force for change is thin, and compares poorly with Eisenstein.

conceptualise mathematical text. Consider the exhibit #5, an extract from Euclid book I - proposition 47, the famous pythagorean theorem, together with part of the preceding proposition. Notice that not every sentence of this text is a statement, indeed the second is not. So I will next consider speech act theory, which derives from the observation that such things happen. I shall be critical of this philosophy of language however, since it does not recognise that statements in writing are very different from those of speech.

§52 Speech Acts

In the decades before and after the second world war in England, Wittgenstein and Austin tried to turn the philosophy of language away from the formal preoccupations of Russell, Frege, and the earlier Wittgenstein, toward a more functional orientation. The primary focus became not the abstract constitution of the sign, as idealised and codified by formal systems of symbolic logic, but the concrete context of its use. Wittgenstein's investigations²⁸ were the more profound and are still being assimilated into the general philosophical consciousness, though this has gone quite a long way. But Austin's insights²⁹ were taken up and further systematised, some would say brutalised, by Searle, into the "theory" of speech acts, which now has a considerable literature, and has recently been given a formalisation by Searle and Vanderveken³⁰. These developments were important against the background of the positivism dominant between the wars, for example Ayer's *Language, Truth and Logic*,³¹ and did much to rehabilitate various kinds of talk which had been ruled out by linguistic verificationism as nonsense, such as moral and political talk. Nowadays speech act theory has been received into the conventional wisdom. There is a large philosophical and

²⁸ Wittgenstein [1953] and his other writings.

²⁹ Austin [1962]

³⁰ Searle and Vanderveken [1984]

³¹ Ayer [1936]

linguistic literature³² and chapters on it can be found in progressive first year linguistics texts³³. But this acceptance rather masks the fact that these pragmatic approaches to language are really quite opposed to the positivist tradition, which in philosophy of language still dominates, in the reductionist logico-formalist programs associated with the terms "what is a theory of meaning" and "causal theory of reference" and the names Donald Davidson, Richard Montague, Saul Kripke, Hartry Field, Michael Devitt et al.

The main point of the speech act theory is simple to state, it is this: not all sayings are statings. That is, it is not the case that whenever someone, call her S (speaker, she), says something to someone else, call him H (hearer, he), that S is making a statement, conveying a packet of information, informing H, transmitting a message. This may seem obvious, but there was a considerable depth of presumption in the early philosophy of language, that the assertion of a proposition is the *basic* form of speech, with others such as questioning and commanding being *derivative* forms, modalities of the basic one. Austin explicitly contested this idea, pointing out that it is possible to do all kinds of things with words, indeed that some things are actually done simply by the utterance of suitable words, without any fact-stating or reporting being involved. The importance of the speech act viewpoint is that it reasserts what some scholars are prone to forget, that language is a survival skill used by certain social animals to get things done³⁴.

One of the standard speech act examples is promising. If S says to H

[1] " I promise to remember your birthday this year."

then S has not so much given information to H as *committed herself*. This commitment is the main *point* of the utterance. Another typical example is apology. If S says to H

³² Cf Bach and Harnish [1979].

³³ For example, Hurford and Hearnshaw [1983].

³⁴ That viewpoint was asserted in a famous lecture by Brouwer, the founder of mathematical Intuitionism, which might be romantically seen as the origin of the present tradition, since it is supposed to have been the stimulus for Wittgenstein's realisation that the Tractatus had flaws and for his return to philosophy.

[2] " I apologise for forgetting your birthday last year."

then the point of this utterance is not to assert the fact, if it is one, that S is in a certain psychological state, that of being sorry; rather, the point is to *express* this in this way. If S said to me

[3] " I am sorry that I forgot H's birthday last year."

then S would not be apologising to *me*, but I would be just as much informed as H would be were that said to him instead. The central idea of speech act theory, then, is simply that different kinds of utterance have different points, *illocutionary points*, as they are called. The illocutionary point of promising is to commit the speaker. The illocutionary point of apologising is to express S's attitude to something bad for which she is responsible. Promising contributes a new strand to the social fabric, apology repairs a rift in it. Austin created a very tentative classification of performative verbs, that is those verbs which like "promise" signal explicitly the illocutionary force of an utterance. He divided them into five groups, to which he gave the rebarbative names verdictives, exercitives, commissives, behabitives and expositives. He claimed little for this grouping of the 1000 or so such verbs to be found in a concise dictionary, except that it would allow him to play Old Harry with the true/false fetish and the fact/value fetish. Protootypical examples of these five classes would be assess, appoint, adopt, apologize, affirm. If I assess I give a verdict, if I appoint I exercise my role or position, if I adopt I commit myself to whatever is adopted, if I apologise I express my attitude to behaviour, if I affirm I indicate how I think the proposition in question fits into discourse.

Searle's book *Speech Acts* provided a detailed attempt to describe speech acts in a *systematic* even "scientific" way, invoking an apparatus of rules to govern the employment of various "illocutionary force indicating devices". In that book he offered no improvement on Austin's classification, but rather delved inside speech acts to examine how reference and predication may be thought of as speech acts too. Now this involved him in regarding any speech act as made up of a proposition p and a force F, so that the general "formula" for a speech act is F(p). This is already a

debatable move, one which I believe is actually quite contrary to the spirit of Austin's attempt to escape the idea of statement as basic. In the most recent "authoritative" version of speech act theory by Searle and Vanderveken, this treason becomes plain. In this work there are claimed to be *precisely* five basic kinds of speech act, because there are five possible basic illocutionary points: assertion of p by S, committal of S to p, direction of H to p, declaration that p, and expression of p, where in each case the proposition p is the content of the act, e.g. that S forget H's birthday. Promising is commissive, apologizing an expressive speech act. (The argument given for why there are precisely five basic points is one of the most feeble I have ever seen in a serious philosophy book. It is that there are four possible "directions" of the world-word fit. You can produce words to fit the way the world is ("assertives"), or change the world so that words fit it ("commissives" and "directives"), or change both ("declaratives") or neither ("expressives"). Thus 2×2 gives five, because either S or H can be responsible for changing the world - as in promising versus ordering. I cannot bring myself to marshal arguments against this pathetic transcendental deduction. Its invalidity will become evident. Anyway, utterances have various illocutionary points even if their classification is not yet finalised.)

Because speech acts are attempts to do more various things than merely state truths, there are more ways they can go wrong than to be false and so there are a variety of conditions for them to be *felicitous*, as Austin puts it. Austin rejects the primacy of the true/false dimension of evaluation of speech acts. He gives six conditions for felicity. The first two conditions, A1 and A2, are that there be a suitable conventional set-up, and verbal procedure (paradigm example: marriage ceremony - one of Austin's favourite examples is the saying of "I do" in a marriage ceremony), and that S and H fit into it appropriately (say, celebrant and groom); conditions B1 and B2 are that the procedure be carried out correctly, and completely ("Will you..." "I will"); conditions G1 and G2 are that S and H have the right intentions etc, and that S and H act appropriately afterward (S must be authorised,

H must mean it). If one of the first four conditions fails, the act does not come off - there is no marriage; but if one of the others fails, it is successful but defective. Searle avoids the term felicity, because he wants to mark the distinction between conditions for the successful effecting of a speech act and conditions for its defective but successful performance. However I retain Austin's term because I don't think it is clear that we always want to equate defectiveness with the failing of any particular conditions and not others.

Searle distinguishes seven "components" of illocutionary force in his most recent classification of speech acts. First, as we have seen, each speech act has its own *illocutionary point*, or essential condition, as he called it in 1969: for example, an utterance just is not a promise if it does not commit S. Second, a speech act has a *degree of strength*: to request H is weaker than to insist that H remember my birthday; more is required for me to be able to insist. Third, there is the *mode of achievement* condition. Many acts must bring into play institutional roles and other devices in order to succeed: to command H, S must be in a position of authority and use it; to testify in court S must be acting as a witness. Fourth, there is the *propositional content condition*; only certain propositions can be the content for a given speech act. You cannot promise to have remembered H's birthday, or apologise for the law of modus ponens. Fifth, there are always *preparatory conditions*. That is, certain conditions must hold in the context of the act, before it can be sensibly attempted. If H is sensitive about his age and does not want S to remember his birthday, she cannot promise to remember it, though she might in the same words threaten to: one of the preparatory conditions for promising X is that H want X. S cannot apologise to H for something H did, or for doing H a favour: you can only apologise for something for which you are responsible. And so on. Sixth is the *sincerity condition*. A speech act purportedly expresses a certain psychological state and is only felicitous if that state obtains: you cannot be promising if you have no intention of remembering H's birthday, you cannot apologise if you are not sorry. Searle wants to call an insincere speech act successful but defective. Seventh, a

speech act's *sincerity condition has a degree of strength*: promising faithfully is stronger, commits one more, than just promising. The Queen's pardon is stronger than a simple forgiving (and to make it she must deploy her social role).

This list seem to me to muddle together a revised version of Austin's list of felicity conditions, the point of the utterance and two "degrees of strength", those of the point and the sincerity required. Searle and Vanderveken wish to claim that the ordered set of these seven components classifies *all* speech acts. Their lust for "scientific" classification perhaps explains the peculiarity of forcing these different kinds of things - conditions and degrees of strength - together. In the end though, they deal with only a hundred or so performative verbs, omitting many mathematically interesting verbs mentioned by Austin such as 'define' and 'calculate'. I shall suggest the reason shortly. Moreover the procedural conditions B1 and B2 of Austin seem to have got buried somehow. The preparatory conditions (plural) seem to run together Austin's A1, conditions about the institutions which must exist, and G1, conditions on S or H's mental states - which are also involved in the sincerity condition. Anyway, Austin and Searle propose various conditions of *felicity* for speech acts, though it is not clear to me that we have as yet a definitive classification of them.

The root idea of illocutionary force is that in saying something often I am doing something. By saying certain words I actually promise, that is create an obligation. Very often something further is brought about by a speech act: S may annoy H by promising - *yet again*. Or S may avert divorce by promising, or prevent H hearing the weather forecast, and so on. Such consequences of utterances which go beyond its illocutionary force are called perlocutionary effects - things achieved by, through or as a result of, speech. Austin devoted some effort to finding ways to distinguish them clearly. In his most recent version of speech act theory Searle has tried to limit consideration to illocutionary force on the ground that perlocutionary effects can be achieved not only through speech. This is a bad argument, for illocutionary effects can be achieved nonverbally too - I can promise by "crossing

my heart", or agree by nodding my head, or apologise by facial gesture, or marry by signing my name.

To sum up this section, speech act theory points out that utterances are not all statements, they can have various illocutionary points; that they can go wrong in more ways than to be false, so they have various felicity conditions; and that utterances can have effects, one can do things by saying something as well as in saying something.

§53 Hard cases for speech act theory

If our exhibit text were spoken, we could use the framework just described to distinguish between the first two speech acts in it; the first is a statement, but the second is an order, a kind of exercitive. I do not object to this, indeed I think it a useful advance on the logico-formalist idealisation of mathematics which pretends that it is all statements, and I will come back to the role of such non-constative speech acts later. However if we think about the general characterisation of speech acts we shall soon see that there are some peculiar features to the alleged speech acts involved in the written text we are actually considering.

In this section I shall add to this text, exhibit #5, various other examples with which I think speech act theory cannot cope, and discuss some of its self-imposed limitations. The reason why the examples I give pose problems is that they all involve writing. This is the relevant point for our present purposes. I am not concerned with the (many) other limitations of the speech act point of view.

Example 1: written-down speech acts

Suppose S takes some quite standard speech act and writes it down, for example "I promise to remember your birthday next year", and shows it to H. Now what difference does this make? Has S promised? Maybe; certainly it is possible to make a promise in this way. But suppose S sends it through the post instead. Now has S promised? The post can be rather tardy, or go astray: what happens to the

promise if it arrives after the birthday? Does it apply to the next year's? Or does it just lapse? What if it never arrives at all? In any case it seems much more doubtful that it can effect a promise, since if that is the whole communication, H will have no idea from whom it came. So the *recognition* of the intention to commit oneself, which is part of the condition for successful promising, seems to be absent. S could obviate this problem by writing her name on the message, but to make it stick we really need a signature and a date - and perhaps even a witness to them. For a promise, H must accept that it is made in good faith: this is hardly possible except in a face-to-face speech situation where a whole lot of nonverbal interaction surrounds the talk. Telephone conversations are very hard to use for emotional ends for just this reason. Sam Goldwyn remarked that a verbal promise wasn't worth the paper it was written on, but a simple written promise may be even less valuable. Writing down a promise and trying to ensure its felicity seems to lead toward the notion of a *contract*. What this example shows is that the means of expression of an utterance, say writing, may impose further conditions on its felicity which have no parallel for speech acts: conditions on the temporal structure of the whole communication, conditions on the form of the utterance, conditions on the chain of communication; while some of the conditions of the speech act of promising, those concerned with the mutual recognition of their intentions by the participants seem to fall away.

Example 2: bequest

Now consider an utterance which I will argue *needs to be* written. Suppose the sentence

"I leave everything to the ABC."

should turn out to be the entire content of my will when the executor opens it. How does this utterance fit into the speech act framework? If it is felicitous, the utterance has the effect of reassigning the ownership of certain items. That is its point, to gain this transfer. Which of the *illocutionary* points is that? Clearly not

the assertive or expressive: I do not say that certain things are owned by X or Y, or that I would like it to be so. It can hardly be a commissive - the will cannot take effect until I am dead so that I can hardly be committed by it. Directive perhaps; well, yes, the executor is guided in his actions by it. But nobody specifically is instructed to do something, that is not necessary for the bequest to be effected. How about declarative? Surely the reading has brought it about that I made a certain bequest? Well, no, not the reading, I am by definition not there for that, the executor does it: a declarative is successful when the *utterance* makes things so. So whose utterance would it be, my writing it or the lawyer's reading it? Clearly my writing the sentence on the paper does not by itself effect anything - I have to die too, and then the will be read. So if we consider the utterance alone, it has no effect; but if we consider it completed, *I alone* have not acted. Either way it does not appear that the utterance has any of the illocutionary points as given by Searle or Austin.

Let us consider some of the other conditions. A will is a written document which must be written when in sound mind and under no duress, witnessed by two witnesses and signed by the testator. Moreover it must be read after W is dead; only such a reading can effect the will. Readings of it before death are no good: otherwise, we can't be sure it's the *last* will and testament. So we have various preparatory conditions; but notice that they include the need for the utterance to be written, not spoken. Saying it won't work. There is no verbal equivalent of a will. Then there are clearly some propositional content conditions, you can't bequeath what you don't own, or just write for example "I leave it all gladly, you're a pack of bastards." But the sincerity condition seems quite irrelevant. Whether the bequest succeeds does not depend on my state of mind when the will is read, obviously; nor on my state of mind when writing the will, since that is not available for probate. Writing and not meaning it doesn't vitiate it any more than mental reservations do in weddings! (Indeed I don't believe sincerity is required for getting married.) Notice also that there is no H; the utterance needs no *particular* audience necessary to be effective. Summing up, it seems that some of the conditions for speech acts carry over to this

kind of utterance, but that the illocutionary point of this act is not easily squared with the standard types, that sincerity is irrelevant and that it is *necessary* that the utterance be a written one. On the other hand, there is no H, it *cannot* be done in speech and the intentions of S are presumed to conform with the utterance rather than vice versa. These are not surprising divergences, since they arise from the fact that there is by definition no face-to-face two person interchange here. (It might be said though that I have merely pointed to a kind of speech act which has special "mode of achievement" conditions. One might say this, but its force will weaken as we see more and more examples.)

Example 3: writing a cheque

Suppose W writes out a cheque on one of her own cheque forms in my favour for \$100, having \$200 in her cheque account, I take it to the bank and present it to the teller who credits it to my account. How shall we describe this utterance in the speech act framework? It seems obvious that we have here a directive: W has directed "the bank" to transfer some funds from her account to mine. Cheques actually say "Pay..." But who has been directed? No-one in particular, "the bank" is not a person, though legally this fiction is maintained, and in fact any old teller will complete the transaction. Yet not just anyone either: it's no good my sticking it under *your* nose. Again the sincerity of W's direction seems quite irrelevant here, the transaction is effected on the basis of the cheque, not any mental state in which it was written. There are no insincere cheque-writings! As in the case of the will, the signature is what counts. Indeed in this case, there is not really any presumption about W's intentions at all. Granted, there are various preparatory conditions here, W must have an account and fill in the date etc; once again one of them is that the cheque be *written*³⁵. I can't drag her into the bank and get her to say to a teller "Pay Edwin Coleman \$100". The bank needs a document to record the transaction, apart

³⁵ I am informed by counsel that this is usually part of the customer's contract with the bank (in the fine print!)

from anything else. There are very strict propositional content conditions here too (cheques require a particular kind of verbal form). But I am not claiming in these examples that they are *nothing* like speech acts; rather that they have certain particular differences. In this present case we find as in the last one that there is no particular H, that there is no spoken equivalent because of the details of our legal system, and that S's intentions are irrelevant.

Example 4: roadsigns

Consider a typical roadsign, one saying "30". Is this some kind of speech act? It is certainly an utterance, it could be glossed as "Drive at less than 30kph", someone wrote it and put it there for people to read; though we prefer shorter road signs for very good reasons. So it looks like a directive again. Is it? Who is S? The law? The people of Australia? The signwriter? It is not very clear. And who is H? Well, anyone driving by. In fact such a sign is very unlike a speech act in that no single speech act could possibly correspond to it. To get the same effect from speech you'd need someone posted there yelling out "Drive at less than 30" to any car approaching. But it wouldn't work even then because the driver wouldn't, couldn't, hear. (Some kind of electronically projected whisper from the road might work, I suppose.) Once again we see that the sincerity condition is meaningless here, and the basic dyadic situation of S and H and conversation is a poor model of what's going on. What *is* going on is that the point of this utterance, which is warning or admonition, is achieved not by its being uttered but by its **staying uttered**, by its being continuously available for uptake. There is no kind of reciprocity here, when I read the sign there is no possibility of my taking the matter further with its writer. The idea is absurd. Of course, if I think 40 would be safe, I could write to the Department of Highways and suggest they change it, but this is not possible or pointful at the time of my reading it. And I'd *have to write*, since it is a bureaucracy not an individual which uttered it. We see from this example that there are utterances to which the roles of S and H and the basic conversational set-up are

quite foreign, whose point is achieved not as a kind of transmission of a meaning from S to H but by *being there* for us to notice at need. The intentions of S are irrelevant, the initiative is differently located.

Example 5: mathematical text

Obviously most literature exemplifies what I'm pointing out here, but the example I'm actually interested in is mathematical text. Consider again the example the text from Euclid, exhibit #5. It seems to be a sequence of not only assertives but also directives - which is already in conflict with the logico-formalist view of mathematics as a string of assertions. What I want to point out now is that once more it is clear that the state of mind of S is totally irrelevant, this time because it is the characteristic feature of mathematical text that *it justifies itself*. Who cares whether *Euclid* believed Pythagoras' theorem? The second point about a mathematical text like this is that like any other written text, it can have many different readers. They can be so different that the effect of reading the text can vary wildly from one to another (this remark is enlarged on in Chapter 9). This is a commonplace about reading literature, but it is just as true about mathematics. Once everyone took Euclid as a model of rigour, only the innocent do now. Thirdly, a piece of mathematical text like this one might be used by a solitary reader for learning mathematics or perhaps greek, but is much more likely to be used by someone to teach a third person some geometry.

What this and the last example show is that we can have utterances which work without any assignable utterer, without any particular intended audience, and whose point is largely to be available for reference. We also see that there may be more than two parties to the effecting of some communication acts - we must also remember that Euclid has been passed down to us over a very long chain of copyists, translators and commentators!

§54 Limitations of speech act theory

Some conditions that must be satisfied for "illocutionary uptake" (getting the point of an utterance, as hearer) are swept aside as of no theoretical interest by Searle, such as H being awake, paying attention and speaking S's language. Austin also puts these things aside - together with the possibilities of acting, lying, literature, soliloquy and so on. The basis for these exclusions, which Searle repeats more stridently, is not given by Austin, but it is fair to think that he held some kind of priority thesis of the kind which Searle explicitly endorses: fiction is logically dependent on fact, you couldn't have jokes without serious discourse or metaphor without the literal. Searle explicitly likens the enterprise of philosophy of language to that of science, and justifies idealisation and the exclusion of complicated cases on the basis that it works in science³⁶. The idea is to get the basis understood and then look at the superstructure. Molecules first, then heat.

Stubbs has argued³⁷ that there are many other respects in which speech act theory restricts its purview: apart from one I have pointed to, namely the limitation to two person face to face situations, he mentions the concentration on invented isolated "writing-like" sentences, ethnocentricity, avoidance of the negotiability and indeterminacy of real speech act situations, an underestimation of the variety of possible acts and ways of effecting them, a restriction to explicit cases and ignoring of hedging (a very common and important phenomenon), the assumption of expressibility, that is that any speech act can be given in explicit form (despite the fact that many of them are *meant* to be vague and uncommittal and inexplicit), and a tendency to concentrate on unimportant examples like "Pass the salt!"

I believe this is all very much to the point, but the response from Searle would be simply this: we will get to them, give us time, we have to get the basics straight first. To that, the question which arises is this: how can you be sure that you have not excluded *essential* features of speech acts by ignoring all these

³⁶ Searle [1969: 56]

³⁷ Stubbs [1983]

possibilities? Are these not *necessary* possibilities in the sense that it is of the *nature* of a speech act that it can be pretended or dissimulated or vague or disseminated or written? And if so, won't your description of speech acts be hopelessly *thin* if these aspects are not built in? Stubbs has pointed out some interesting work by Erving Goffman, which by looking more keenly at some speech act situations increases this worry. A sensitive examination of a real conversation can show that there are typically more than two participants and more than two roles - the bystander is a third and different role in a conversation, and participants often switch roles. Moreover an individual may put several versions of herself into play in a conversation as in saying "*To the best of my recollection I think that I said that I once lived that kind of life.*" To label all the 'I's here by a simple S is crude in the extreme. One would expect Searle to respond in the same way to this point: yes, these things happen, but such speakers have to learn to do simple things first. But is this really so? Goffman also points out that play and pretence are endemic in the *earliest* conversations between parent and child. Can we really be sure that first children learn to speak literally, then later graduate to metaphor? Could it not be *the other way around*?

The important limitation of speech act theory for my present purposes is that it ignores the possibility that writing may change completely the nature and force of an utterance. The most profound meditation on the issues raised by writing is in the works of Jacques Derrida, and one of the more accessible entry points into that Gallic imbroglio is his rather friendly criticism of Austin³⁸ for just these systematic exclusions of the written, the nonserious, the pretended, the theatrical and most other uses of language. Searle, as the "leading expositor" of Austin wrote a response suggesting that Derrida did not understand Austin, and this led to a very long reply from Derrida showing that indeed he did. There is no hope of doing justice to the richness of Derrida's "critique of logocentrism" here, I merely indicate

³⁸ In "Signature Event Context", Derrida [1971]. Searle's response is Searle [1977]. Derrida's long, virtuoso rejoinder is Derrida [1977].

that it lies behind my own thoughts. I believe that it cuts to the heart of what is rotten in the philosophy of language in English. It does so from a direction quite different from Wittgenstein's, but to an effect quite similar to his critique of language³⁹. In any case my present purpose is not to pursue the general notion of signification which Derrida proposes as a result of his reflections on writing; I am interested in some specific possibilities created by writing, in particular mathematical text. The general point is this: if the concentration on simple spoken utterances is justified by a spurious notion of constructing a "scientific" account of communication, assuming that serious statements in "normal" speech conditions are basic and so on, then if we don't accept this kind of scientism we are obliged by the prevalence of other forms of communication in our lives to consider a less dogmatic approach. Statements are not the most common kind of utterance, seriousness is unusual and in any case a metaphor, writing is heavily involved in a large proportion of the communications which distinguish our lives from those of our remote ancestors. Mathematics and writing and a nexus of other crucial social formations emerged together only a few thousand years ago, yesterday from the point of view of the age of speech. It seems very shortsighted to ignore this concomitance on the basis of our present quite primitive understandings of language and science.

To sum up this section and the last, I have given examples of how written utterances pose some problems for the speech act framework, and suggested that its systematic exclusions may not really be well-motivated, and may indeed be seriously misleading.

§55 Text acts

The examples of §53, and lots of others, show that some utterances must be written, that utterances which are written can seriously breach the speaker/hearer

³⁹ Staten has written a comparison of these two philosophers, Staten [1985], but it seems to me not to get to the heart of their real affinity - and difference.

nexus, reducing the role of intentions and sincerity, and introducing the availability of the text as a new resource. Not every utterance is any kind of speech act at all. Austin's and Searle's conditions involve clearly and crucially the *intentions* of the speaker and of the hearer; S must intend a speech act to have the force it does, or else it won't have it; S must intend to keep her promise, H must want X if S is to promise it, dislike Y if S is to apologise for it. I am suggesting that this emphasis is not a good idea, because for some utterances, particularly those in writing, S and her intentions are not available and cannot be relevant when the H reads the text. The examples show clearly that there are many kinds of utterance which involve our intentions in much more indirect ways. I propose to use the slightly barbarous neologism *text act* for communicative transactions which rely on writing for their felicitous occurrence. I adopt it in analogy to *speech act*. By a text act I mean a signifying act, with specific conditions of success and typical consequences, which is effected by an utterance in writing⁴⁰. Many, but not all, text acts are special kinds of "speech acts" modified by their change from speech to writing. Many other text acts correspond to *nonlinguistic* semiotic acts; a third class are peculiar to writing and have no parallel out of writing. Thus asserting, questioning and ordering as speech acts, have text act versions which share many of their locutionary features, but drawing a diagram is not derived from speech at all, and citing an earlier point in the text is an essentially written semiotic act. Most speech acts have a text act counterpart, but not all do; and there are some text acts with no genuine speech act counterpart. Genuine questions make poor text acts, while bibliographic reference makes no sense in speech. I shall argue in what follows that the characteristic communicative acts of mathematics, and thereby of much science, are of this kind: they *rely* on writing for their constitution.

⁴⁰ There are of course script acts and print acts, and the differences between them are very interesting and important, though the worthwhile literature is not large, but for present purposes I am looking at what they have in common. There are also "magnetic" acts which I understand even less than text acts. The owl of minerva ensures of course that we are seriously trying to understand text acts, only as they are being transformed by the microchip into something else, just as the nature of speech became an issue only for the incipiently literate cultures of the classical age (as Havelock shows us).

A text act involves a writer, W, who utters the text, usually as part of a larger, integral text. The writing context includes a corpus of relevant existing texts as well as the rest of the human world involved as context in speech acts. The writer has specific intentions which must be to some extent recognised for the act to be consummated. Generally these include the occurrence sometime of a complementary act of uptake by a reader R (or more often, readers). Where the text act differs from the speech act is that these intentions are not usually recognised at the time of utterance, and they are not simply essential to its success. It must be possible for the writing act to be completed successfully without its consummation by any particular act of reading being presupposed. It is matter of considerable debate in literary theory whether a text act can be consummated regardless of R's recognition of W's intentions. It is indisputable that many successful text acts have been completed with important consequences in great ignorance of such intentions. Archaeology is a witness to that. The plethora of interpretations of ancient philosophers and poets is another. We should beware, though, of rejecting writers' intentions as entirely irrelevant. One dimension of the success with which a text act is complemented by a reading act is the extent of R's recognition of W's intentions. It is characteristic of text acts that this fit is loose, and that there are other dimensions. It would be stupid to ignore the way texts are mostly used, on account of some kind of pure criterion of their successful complementation which rules out most readings as defective. The text is much more independent of W and of any particular R than is a speech. What I am arguing here is that it will not do to think of text act communication as a kind of slow conversation with aids. The gross effects of text acts are different and more important than that: they make new kinds of communications possible.

§56 Text act, text and discourse

Texts acts occur in texts, or perhaps we should say texts acts make texts, typically several acts together. The natural unit of written communication is the

text rather than the sentence. By *text* I shall mean a written sign which is circulated in complete physical tokens - books, articles, handbills, currency notes are common examples. I do *not* use 'text' as has become unfortunately the practice with some linguists keen to get away from the dominance of sentences, to mean a sample of actual linguistic usage, and a generally spoken one at that. For me a text is always from this point on, written. The salience of the text is a main reason for rejecting the logico-formalist disregard for the vehicles of signs. It is *important* to the significance of Euclid's Elements that it consists of thirteen "books" always tokened together.

Texts form *discourses*, aggregations of related texts with vague boundaries but clear distinctions from one another. Mathematical discourse is the example I am interested in here. English literature, quantum physics, Roman law, Hebrew religion are others. A discourse is a complex praxis with a collection of texts as its organising focus. A discourse is constituted by a common subject-matter, by cross-references, by its persistence and distribution as physical collections of texts. Discourses are most unlike sets, they can have common members, have vague boundaries, can contain sub-discourses (Eighteenth century English literature) and so on. It will not be necessary to be more specific for now.

Texts being typically composed of more than a single complete sign like a sentence, and often many more, they can easily include signs from various different codes⁴¹. We have seen this already in our analysis of a mathematical text. Many signs in texts play roles which organise the text and relate it to its discourse. Text acts, like speech acts, may have different purposes, and since purposes are the differentia of speech acts, they will distinguish text acts from speech acts, and text acts from one another. But the classification of text acts must also take account of the discourses and the texts in which they are found as well as the particular roles they play within the text.

⁴¹ Some speech act literature considers sequences of speech acts, for example Ferrara [1980], but not much.

§57 The points of text acts

So what purposes, what illocutionary points or perhaps better, what **semiotic points**, may text acts have? I have no intention of proposing like Searle some partition of all possible text acts. I don't believe that all speech acts can be classified as he wishes, indeed I think that they are all different. The purpose of his five basic types and of the two strength components in his account of illocutionary force is actually to make a recursive formalisation for a "logic" of text acts seem plausible. This in my view is to travesty the whole speech act framework. I prefer Austin's more botanical method of comparison and contrast, happily accepting that there ^{are} over a thousand different kinds of performative which he can only group loosely and with overlap and redundancy. I am sure that interesting "genetic" accounts of the various presently existing acts can be constructed. But the various speech acts we now have, and a fortiori the text acts, have different functions, so they are different. None is basic and none derivative. Instead I shall simply suggest some ways to group text acts which have similar functions.

One way to classify text acts is by the discourse they contribute to. Thus we might identify legislatives, educatives and so on. Consider *legislatives*. These look very like speech acts written down, and of course that is how they began. But it is essential to legislation that it be written; the kind of dispute about the facts which is hammered out in our legal system must have some kind of indisputable anvil, and that is the law of the matter in question. It is removed from dispute by being archived. In a purely verbal society, there will of course be a form of law, but it must rely on personal authority. Consequently it is not in essence different from my settling the kids' squabbles. In the end, what I say goes. The facts are a matter of personal recall, and the rules for dealing with them are too. Now I know that our form of law is subject to processes of interpretation, and this certainly reintroduces doubt and personal opinion. But the interpretations themselves become part of the archive, so we are not back to square one, on the contrary we are sitting on top of a

pile of case law! Indeed, laws are really *standing possibilities of interpretation*. In any case the point I wish to make is that our form of law could not be made and used by speech acts alone. When the ruler's word is law, there is *no law* in our sense of the term.

Another kind of text act is the *instructive*. This is found in discourses which are aimed at assisting readers to do something they could not do alone or already. Very frequently these are *exhibitives*: "look at the diagram D; notice that lug A is larger than in the primitive Coleman textpump..." Such text acts generally rely on nonverbal semiotic systems deployed in the text along with written words, such as diagrams, photographs or notation systems like that of music. A music score is a paradigm instructive. So is a textbook in the ordinary sense of text, or a computer manual. Now any particular use of such a text could possibly be replaced by a verbal instruction from someone knowledgeable; but that is not the same as the *permanent possibility of instruction* which a text can be.

Another means of classification of text acts is provided by considering the kind of text which is being made: is it a resource, or a display, or a communiqué or a formality? I have already illustrated this by mentioning exhibitives, *texts which display*, for these are kinds of text rather than discourse. Again, some text acts are directives just as some speech acts are, but a thing can have several functions at once, and a directive may also be a *documentive*. By this I mean an utterance whose purpose is to create a document, a written resource. Some examples are records, certificates, cheques and birthday cards. There cannot be speech acts of which these are merely derivatives because their essential condition is that by their occurrence they leave an enduring trace. Speech is essentially transient. One kind of documentive is the *referentive*: a text act whose point is to create a document which can be easily accessed at will for reference purposes or is saliently present for reminding, like logarithm tables and roadsigns. Most legislatives also create reference texts too (though not perhaps all). The physical nature of a text is important for its being a standing possibility of referral. Another kind of

documentive is the *archive*, whose point is that a document shall be kept and be known to be kept - though perhaps never looked at (Hansard, most of the periodical literature for some sciences). A third is the *certificative*, the point here being to create a document which can be used as a warrant for some attainment or past fact, such as being born or married or graduated PhD. On the other hand the point of a classified ad is not to create a document at all but to be as widely read as possible; it is what I call a *communiqué*.

In practice these various possible points or roles for text acts are not clearly distinct, since many text acts function in several ways at once. They can, because they are continuously available for repeated uptake, for different purposes if necessary by different readers. The consequence of these multiple and separate uptakes is to variegate the point of a text act, in most cases. Insofar as these sign functions are new possibilities, they derive from the specific properties of writing, its visibility, durability and replicability. I will return to these in §59.

§58 Felicity conditions for text acts

Text acts, particularly those deriving from speech acts, have felicity conditions. A text act of a particular kind, say an assertion, has certain conditions on its successful performance, and certain other conditions on its consummation by a reading act. The content and sincerity conditions transfer largely unchanged to the written mode. The intended audience for a text act is usually plural and vague, not individual and specific like the typical H of a speech act. Nevertheless there must be an identifiable intended audience. The preparatory conditions become determinants of the audience - or rather, the congruence of their implications for the audience and the writer's intention is a condition of success. Moreover, the preparatory conditions nearly always include reference to previous discourse (this is called 'intertextuality' by some writers like Kristeva and Eco). Writing is typically conscious of its writtenness, and imposes such a recognition on its successful reading. For example: statutes or scientific papers refer to other statutes

and papers, making demands on a reader which presume that R knows how to interpolate reading parts of such other texts where it is relevant in reading this one. But are there not homelier written texts which lack this self-conscious feature? Certainly. Any spoken words can be put down on paper. But when this is done we may notice a curious effect: if the text is of any length it becomes more and more obvious that it is a written down speech, because it is not sufficiently conscious in the right way of its own unavoidable minimal written context - namely, its own earlier parts. Hence the characteristic patterns in originally spoken epics like the Iliad, which have many devices created to aid memory and recall, which are simply obtrusive in written versions.

In contrast to the content, sincerity and preparatory conditions, the essential condition always undergoes an extraordinary change. Any written assertion with deixis carries the reader into a new kind of situation. There is a presumption that it describes how things are, but R is not necessarily thrown if this proves wrong, because R knows that context changes. Consequently, the assertion is acceptable if R judges that it represents how things were, or are elsewhere, or for other people. That is, R judges how congruent what is said is with what W could reasonably have believed. There is no *literal* assertion in writing; R is always conscious of the gap between utterance and reading. Apprentice readers take time to grasp this fact.

In each of the examples of §53, there are easily identifiable conditions for its felicity. In each case we can identify conditions on the uttering, the utterance and its uptake. The will can go wrong if it is not signed or witnessed, or if the testator was forced to sign or was insane at the time; writing a cheque fails if the signature is omitted; the properly constituted authorities must authorise the form and location of the road sign. But for a proof like Euclid I.47, the conditions on its being uttered are all textual: what is previously accepted in the discourse, or earlier in the text, and how does the proof fit in? Again, a will may be void if it crumbles to dust or is never found or attempts impossible bequests, or departs too far from common

practice, for it may be set aside by legal process; a cheque must have stamp duty paid on it, have the right things written in the right places, survive in the right hands - if any of this goes wrong the act cannot be completed; the roadsign must be made of the right kind of materials and colour and size and located appropriately. Similarly, for me to understand Euclid I.47 it must be in the right language. Finally, a will may go awry if it is read by burglars or only to a philosophy research seminar or after 500 years; a cheque must be presented, at the right place (another bank won't do!), before W's funds are all gone, and W must have an account; drivers must be aware of the meaning of roadsigns (this is a condition of being on the receiving end of another text act, getting your licence) and the drivers must exercise due care, so that they notice the sign. But what conditions are there on the uptake of Euclid I.47? Can't anyone successfully take up this text? Certainly not, most people *can't*, and the reasons why not are very important.

This last point introduces something which I will discuss at length in Chapter 9, the fact that text acts only succeed in specific *contexts*, much more specific than those of speech acts. There are *general* conditions for the felicity of text acts which we can extract from these examples. The first point is that it is not enough to be a fluent speaker of English to use a text in English. One must also be *suitably* literate. This requires that the *reader* have a certain kind of understanding about the text itself and what role it plays in possible transactions: if you try to read Euclid like a novel you will get nowhere; if you look in Proust for philosophical arguments about time you will look in vain. So text acts cannot succeed without readers. The second point of course is that the *institution of writing* must be in place, assuming that this can be considered separately from the existence of writers and readers. Since its institution certainly predates most of the text anyone is concerned with, perhaps this is uncontentious. The third point is that for all real text acts it is necessary for the *text to survive* beyond the moment of production in a form suitable for reading. Fourthly, there must be a *need* for the production of texts of whatever kind is in question. This raises very general questions about the kind of

society in which text acts flourish. I will say more about these points in the next section.

§59 Text effects

I will now argue that it is important for us to understand text acts and how they work, as distinct from speech acts, because text acts, even more than speech acts, have consequent effects. I shall call the analogues of perlocutionary effects *text effects*. Let me recall three main facts about written utterances. As written, they are *durable* and they are *visible*, and because they are signs they are *replicable*. The consequences of these simple facts are vast. Because written utterances are replicable, in the sense that multiple copies of the text can be produced which all endure and circulate in society, it is possible for the effects of a given text act to disseminate widely, to a vast and varied audience. Nothing would turn on the possibility that no two speech acts were ever exactly the same, but many of our institutions rely on the identity of the texts used on different occasions. The *mass audience* is now routinely possible.

Most speech is recycled. The examples given in works on the philosophy of language tend to be single, isolated, often strange sentences; but in real life, most utterances in speech are repeats or mild variations of things that one has recently said or heard. The Chomsky wunderkind idea, that the main thing to be explained about language is the infinite capacity of users to produce and understand new sentences, has sadly obscured this simple fact. There is novelty in what we say and understand, but it is partial, tentative and spasmodic, building on past use. This is even more true of written communication, as anyone who has written a program or an essay will agree. But there is a significant difference. The context of text production and consumption is a text-filled one. Text acts occur in a textual environment: the context of text essentially includes old text. Because written text is *enduring*, the utterance and uptake can be, and generally are, separated in space and time. The first effect of a text act is to add text to the world. The uptake can be

deferred, the utterer can be absent at uptake. Consequently the commitment and the authority which are mobilised by written text take on quite different forms to those at work in speech situations. Authority for, or commitment by, a text can be dispersed or disguised. How many "authoritative" documents are actually signed? The signature is a *relic* of the personal responsibility for utterances, not its parallel in writing. The possibility is created with written text of *impersonal authority*. Texts carry their authority with them, we generally have to judge them on their own merits without access to their utterers, their authors.

In such ways, all our concepts, our material environment and our praxis, are heavily informed by the specific semiotic environment which textual discourses provide. Concepts which had a use before writing, such as authority, are simply not the same as they were. Concepts which *simply did not exist* before writing such as *criticism* and *theory* are now central to our cognitive framework. The pen really is mightier than the sword: text effects are the reason why. In particular the concepts of knowledge and truth have been greatly affected by the textualisation of culture. Much philosophical argument about concepts like knowledge is in my opinion rather pointless, because it does not recognise the evolution of concepts and the material and social forces working for their change. In particular, the cardinal problem of epistemology, the question how in mathematics we can have certain knowledge useful in an uncertain world, cannot be answered with preliterate concepts of truth, certainty and knowledge. The only hope is to approach the mathematical discourse with some kind of readiness to see it and its effects as specific to its written form. To see mathematical discourse as a simple transcription of a lot of speech acts, and all assertions at that, is to completely miss its social reality. These far-reaching claims will be supported by the sequel.

Because written text is *visible*, it can be used in quite different ways to speech. A written road sign is useful because it is visible from afar and is repeatedly, even continuously, available for the driver to act upon. A spoken version could not

be - it has to be gone through before you can repeat it; as you approach a bend, when do you need the warning? Even though reading is much faster than listening, we make road signs brief so that their uptake fits well into the temporal structure of driving; spoken "equivalents" would not. Mathematical text is powerful partly because of the enormous information compression which is achieved by written notation compared with speech. The mind's grasp is expanded by redesigning its objects to fit more economically into its senses! The eye is a channel of much greater capacity than the ear. In particular, thoughts become **surveyable**, and subject to scrutiny, comparison and doubt, when they are sitting there visibly frozen for you. The very term 'theory' derives from the Greek word meaning 'a looking at'. Thus the possibility is created by text acts of new concepts including those of *objectivity* and *criticism*.

Being both visible and durable, a cheque is useful because it can be kept and furnished as justification for the payment into my account; once a verbal order is given, it is gone and could be later denied. An aural record such as a tape-recording would obviate the latter possibility, but would be intolerably clumsy for accounting purposes. A will obviously has to be kept for however long is necessary - and must then be readable since there is no longer the possibility of its being spoken! (The "video will" I am sure will come, but this does not refute my case since there is no essential difference here: it too is durable and visible. The audio will is possible too, perhaps. It will still not be a speech act!) All kinds of records are possible. Text acts leave traces which become part of the enduring necessary context of later text acts, being simultaneously available for a given agent. The possibility is created by text acts then of *cumulative traditions*, involving new institutions. Mathematics and other characteristically textual discourses are constitutive of our world. Nonverbal text is also possible once the general system of written utterance is instituted, indeed it was always present: graphing, calibrating, scoring (e.g. cricket match, quality control). This is of extreme importance in the institutions of science and engineering which rely on *measurement*, and I return to it in Chapter 11.

Thus various new and important cognitive possibilities are created by written utterances. It is not just that individual texts are utterances of a kind different from spoken ones, but that because they endure, and because artefacts made using the knowledge which they support also endure, the context in which texts acts are made is itself a very textual one. The context of text use is not merely old text, but an old text-guided material infrastructure. The world we live in is full of items and practices formed by praxis with text - roads and bridges, laws and lawsns, computers and commuters, you name it there are texts behind it! *The context of text also essentially includes the text-structured world.* Moreover, texts have to have locations. They are mostly retained and stored in some kind of archive. A whole subculture of text production, distribution, curation and consumption has evolved which simply has no parallel for speech. The material vehicle for text has entrained an enormous material infrastructure in the text-based economy. Further, in a particular text act there may be many participants, not just two as in the archetypal speech act situation. There may be an utterer and a chain of copiers, distributors, readers and uptakers. There may in fact be a group of utterers, and there is usually a dispersed set of uptakers. *Text acts disseminate meaning*, rather than convey it like a relay baton. Finally, not everyone uses texts: but who are those who do not? The old, the young, the mad, the poor, the remote aboriginal, the illiterate - text is massively implicated in social structuration. The actors in the text act are typically, in virtue of the social distribution of text, role-differentiated and of unequal power; in the typical speech act they are both full adults, and though they may have roles, they could easily be reversed. This is not true for producers and users of documentives, of legislatives, of instructives. Most texts are produced for the prosecution of social ends which perpetuate and reproduce and intensify the differentiation of society. *Text mediates power and structures society.* This preliminary discussion of context will be enlarged on in Chapter 9.

To sum up what I have said in the last few sections, first, there are communication acts which simply cannot be effected in speech. The signature is what is common to simple examples of this fact. But more importantly, the existence of text has led to a situation in which the standard, conventional, characteristic way to do certain things is in writing, because the context of their being done is now a massively textualised one. That one can plausibly imagine any given text act replaced by a speech act does not imply that one can plausibly suppose that they all could be. Our lives are finite and our capacities limited. Nor can all text acts be replaced by speech acts; mathematics requires to be written because it employs many text acts and text effects with no spoken parallel.

I recognise in the concept of text act a certain instability. Does it really make sense to couple together the actions of some utterance in Alexandria 300 BC and an uptake in Adelaide in 1987 and call the result a text act? Perhaps not. But that just emphasises that we won't understand my reading Euclid on the model of my promising to remember your birthday next year. In the next section I will illustrate some of the preceding discussion by giving an account of definition as a text act; later, in Chapter 12, this will be extended to an account of definition in mathematics which explains some logicians' difficulties with the idea of definition.

§60 Definition as a speech act and as a text act

I can certainly give a definition in speech. Let us consider how successful such an act can be, by trying to apply the theory of speech acts. (Curiously, although Austin lists 'define' as an expositive, he does not discuss it, and Searle never mentions it.) Suppose S says to H "I define a bike as a two-wheeled self-propelled conveyance". The conventional set-up and roles are easily satisfied for this speech act, and assuming H hears S, then all of Austin's first four conditions seem readily satisfied. The others are satisfied if H wants to know and S intends to tell him what a bike is, and if the definition is correct and they both accept and use it subsequently.

But now what is the point of this locution? It seems clearly to be an example of a *declarative*. By saying it, S has made it so: if she is sincere, she (now) uses the term 'bike' to mean 'two-wheeled self-propelled conveyance'. On the individualistic theory of Searle, that is all that is involved. (I am assuming he would treat 'define' as more or less like 'abbreviate', 'name' and 'call'⁴².) To learn that S is going to use the term 'bike' in a certain way may or may not be of interest to H. The preparatory condition concerns the correct use, surely. It is not a definition if it does no more than inform H about S's linguistic habits. It must be a contribution to *common* linguistic practice. Otherwise we are in Tweedledum's territory⁴³. That implies some common purpose in the use of these terms for S and H. But that implies that S cannot really act unilaterally. The speech act analysis only works if it is not a genuine speech act situation! To the extent to which S cannot just stipulate how *she* will use 'bike', the point is not declarative but commissive (let's use 'bike' this way..) and assertive (people who know, use 'bike' this way...) and directive (you should use 'bike' this way too). The point of a definition in speech seems not to be just *one* of the alleged five points at all. This is reinforced by a consideration of some related locutions which could equally well be used for the same purpose: "'Bike' is defined as two-wheeled self-propelled conveyance" or "A bike is a two-wheeled self-propelled conveyance." These also contain in varying degrees all of the four illocutionary points which I have identified in S's definition. And yet S *can* define bike however she likes! Are there perhaps many kinds of speech act called definition?

A spoken definition like S's is presumably a response to a question of the form "What is an X?" (This is part of Austin's first condition.) It is used to explain the meaning of 'bike'. In speech the explanation of meanings is carried out in a great variety of ways, as Wittgenstein argues at length in the *Investigations*. I gave a few other possible formulations; lots of other sentences could be used in response to "What's a bike?". Moreover, ostension is used to a very great degree in those

⁴² Cf Searle and Vanderveken [1985:210]

⁴³ "When I use a word, it means what I want it to mean."

explanations ("that's a bike"; "no, that's an exercise machine" etc). So giving a definition in speech can't usually be more than one contribution to a more general process of conveying the meaning of a term to H. Let us now consider the text act of definition, which differs from its speech act counterpart in point, felicity conditions and effect. The point has changed from the assertion of a discovery or observation, to the laying down of a rule; the main felicity condition has changed from the concordance of the definition with prior speech, to the concordance of posterior text to the definition; and the effect, from the refinement of speech, to the creation of linguistic authorities.

The point of a text act definition is to create a resource, an authoritative guide to the meaning of a word which can be referred to at need. What a term is used to mean is what it means, and dictionaries don't themselves mark any distinctions between common usage and real meaning. They may record dialect or regional variations. But they don't carry out surveys; what the people in my street mean by 'bike' is not the point. The dictionary definition of a bike acts as a rule for all of us; if we don't follow it, then that dictionary has no authority and becomes otiose.

Now as to the felicity conditions, the difference is this: for a spoken definition to work requires that normal speakers agree to its determinations on doubtful cases. A verbal definition must be consensual, we can't have Socrates with one and Thrasymachus with another. However the rightness of a saying tends to be granted at most by those who hear it first-hand, falling off with its retelling; so spoken definitions are unlikely to gain very general currency; their authority is not evident enough. That is why it is crucial that textual authority be established for dictionary definitions. Johnson's written definitions attained an authority which far exceeded any of his verbal pronouncements by being *permanently* available and by being *widely* available in a form which readers could trust to be his and not some tenth-hand garbling. Of course they also had to reflect tolerably well common usage among his readers - but they put the boot on the other foot: later readers had to reflect tolerably well his definitions! (Early mathematical definitions such as

Euclid's are like this. Within a mathematical text, a writer can of course easily adhere to self-imposed definitions, but this only applies to terms invented by the writer. On the other hand with such terms the writer has complete freedom in what definition is made.)

The effect of a verbal definition, even if it is successful, is transient or perhaps to become a proverb: "it's long road that has no turning" perhaps evolved from oft-repeated remarks about such things. Or "justice is the interest of the stronger" might be used to decide whether a sample is a case of justice or not. You just see what is the interest of the stronger, or whatever it is. So the effect is the concurrence of talk, speakers continue to use a word with the same meaning as one another, even if that meaning changes over time and place. But the effect of a text act of definition is to create a textual resource, which will acquire some measure of authority and therefore will *constrain* subsequent usage. Dictionaries are not only permanent resources, they are *authoritative*. Dictionaries have been a great conservative force in language because they are so much more effective than spoken attempts at definition in this regulation of the usage of others⁴⁴. One can now distinguish correct usage from general usage - impossible in an aliterate society.

A complication which is important when we turn to mathematical definitions is that there is no clear distinction in speech between the tacit and the explicit introduction of new terms ("I define a buke as ..."). This is because there is a spectrum to explicitness, and because conversation can backtrack to, and add to the explicitness of what has gone before, if the participants so desire. But in text we can mark a definition explicitly. It is possible for such a mark to be entirely explicit, because the text cannot be revised during its reading; consequently there is a fact of the matter as to where a term first appears in a text, and whether that point is marked as a definition, and if so, whether it has the appropriate form. Spoken

⁴⁴ But in mathematics, the definitions which have been important are not like this at all! The effect of a mathematical definition in modern text act form is to structure text in a peculiar way. Some would put this rather more grandly: defining things consistently in mathematics is the same as creating them. I will discuss this in Chapter 12.

discourse cannot be so definitively divided up as can text, nor is it open to retrospective inspection, so that whether a certain usage is the *first* use of a term is vague. There are ways to signal that one is now defining a term in speech but they may be ignored or overlooked and cannot later be noticed; even an explicit declaration "I define X as Y" can be taken as all sorts of speech act - confession, joke, or observation for example - and its place in the discourse is quickly covered over by subsequent speech acts which may easily subvert and conceal its intentions. The written definition does not rely on intentions for its effect. If one wishes to make use of one of Euclid's mathematical definitions, one simply consults the relevant text. Typically, I am free to repudiate what I find, but there is complex social practice which assigns to such texts greater or lesser authority which I can only flout if I have sufficient cognitive authority myself - which must itself be established textually. Texts have authority which individuals do not. (I go on in Chapter 12 to consider further why definitions in mathematics, though clearly text acts, are unlike dictionary definitions.)

It might seem obvious that the text act of definition has evolved out of the speech act; but actually I suspect that the reverse may be the truth. Definitions, mathematics, Greek philosophy and alphabetic writing seem to have had a common birth. Definition may well be an invention of literate culture, originally a text act taken over into a speech act. Informal meaning explanations rarely take the form of explicit definitions because ostension and the use of example are so useful that they are all but universally used in explaining a meaning - except by bookish persons! We shall come across other examples of this rather confusing possibility of *code inversion* as we proceed.

Chapter 6

Diagrams

§61 Kinds of diagram

Diagrams from some sample mathematical texts are collected in the exhibits. Their contexts may be partly seen in the pages they appear in; they are described further in the front page of the exhibit booklet. These examples include each of the kinds of diagram¹ which are generally identified in the literature² on graphics. The classifications which can be found in Bertin, Holmes etc rely on the form of the diagram, sometimes on the content. Thus we have geometrical and statistical diagrams, a superficial distinction by content. But in the usual classification, as best exemplified in Bertin's magistral work, these are jointly contrasted (under the term 'diagrams' - he uses 'graphic' for the more general idea) with maps and plans, and these all together with networks. Within each kind there are of course subkinds, for example statistical diagrams include line charts and pie diagrams.

The diagram in exhibit #4, CS174, is *analytical geometrical*. The figure from Euclid I.47 in exhibit #5 is a *synthetic geometrical* diagram. Exhibit #6A, from Schroeder's book³ *Number theory in science and communication*, and #6B are examples of *statistical graphics*; the first is a *histogram*, the second a *scatterplot*. Exhibit #6C is part of the *plan* of a staircase, while #6D is part of a *map* of part of Adelaide. I shall argue below that tables are diagrams, and that lists and forms must be considered as part of the same sign-system, so exhibits #6E and 6F are examples of a *list* and a *form* respectively. Exhibit #6G is a *chess diagram* (included as seeming to standing outside the usual classifications). Exhibits #9 and 10 also contain diagrams; both are *networks*. I do not wish to reject the loose classification of diagrams which I have illustrated here, and which is adopted with mild variations by many writers. But I point out that it is a classification by form and/or

¹ The terminology used for diagrams, charts, figures, graphics etc. is not fixed; since I use 'graph' more generally for written signs, I shall use *diagram* for any sign that might by other writers be described by one of these terms. I exclude pure pictures.

² I refer principally to Bertin [1983], Cleveland [1985], Tufte [1983], Schmid [1983] and Holmes [1984].

³ Schroeder [1986]

content, to which I propose to add a different classification by *function*. Naturally there is a certain lack of orthogonality here, certain functions being best played by diagrams of a particular form.

Diagrams are very commonly used in mathematics, and some books are positively full of them - the book by Schroeder is a good example⁴. They are particularly common in *expository* texts. However, there is a general repudiation of diagrams by purist mathematicians, which is part of the logico-formalist hegemony. This idea that diagrams are unnecessary and dangerous, serving "only" psychological functions, and should be extirpated from mathematics, is based on fallacies, is quite wrong and will be refuted later in this chapter. To do so will require a close examination of what roles diagrams *actually* play in mathematics, and how they do so. I shall therefore be continuing the investigation of text acts into this independent sign-system.

§62 The formal elements of diagrams

The formal elements of diagrams are points, lines, areas, characters, text and pix (arbitrary picture elements, such as the stick figures used in pictograms). In this chapter I shall sometimes use 'text' as a short term to refer to written word signs, as an element in diagrams. This is a common usage which I hope will not be confused with my more theorised uses. The simplest diagrams consist of only lines with labels; Euclid's in exhibit #5 is one such. (Descartes' *Géométrie* also has many such diagrams.) A diagram cannot consist of *only* lines, though a picture may. I shall explain why this is so - diagrams have to be grounded in a context - in the next section. Geometrical diagrams often use points and areas too; they do not actually use volumes, though they may represent them⁵. Statistical diagrams use these same elements, maps and plans tend to involve pix while networks are limited to points

⁴ Actually, this book is such a good example of so many things I wish to illustrate that I tried to think of a way to incorporate it as an exhibit! Failing that, I will mention it several times again later.

⁵ Some writers, even Bertin, seem a little confused on this point.

and lines. But they all use either characters or text because *they must have labelling*. The chess diagram apparently eludes this generality, but I shall show that this is not so.

I distinguish characters and text because single letters (and other signs) are often used as labels - as in Euclid. But one also finds significant written word used in diagrams, for example the label 'Figure 11.3' of the diagram in exhibit #4, and the labels on it of 'Loss' and 'Profit', '12500' etc. The distinction is important because meaningless labels are evidently playing only a single role, that of intratextual anaphora: they enable the diagram and its text to be mutually referential. The text in diagrams may do this too but plays other roles *at the same time*. The label is a highly significant element of diagrams which does not occur in speech or even written word and I shall be explaining why this is so and why it is important in the next two sections.

"Points" are often really pixels used as atomic signs in the diagram with some mnemonic or persuasive purpose, particularly in statistical graphics - as when a point is made to look different if it represents several data values, or the squares of the chessboard are represented by real squares, and colored. The pictogram makes a principle of the idea of pictorial points. The literature, for example Bertin, goes into this in some detail. Diagrams can be anatomised in great detail by the formal elements of which they are constructed. But it is not a very deep classification because it is too easy to make up or find counterexamples. Scales can be left implicit: Tufte has an example of a histogram made out of only the data values, but it's still a data *graphic*. Geometry diagrams can be drawn with a paintbrush: areas are now the elements, not lines. And so on. The famous London Underground diagram is a "network" with some metric and colour overcoding, but it remains a map because of the use for which it was made and is used. One of Euclid's diagrams was used in a different context as the Pons Asinorum, Gauss's construction of a regular 17-gon he wanted for his epitaph! A more illuminating classification of diagrams will need to involve the uses to which they are put.

§63 Diagrams are formally articulated by juxtaposition, alignment, spacing incidence and measure.

A diagram is constructed: it consists of subsigns put together in a suitable way. By this I mean that the diagram in exhibit #5 contains the subsign BC, which plays a specific role in the diagram; on the other hand the left hand half of this line is *not* a subsign, though it is *part* of the sign BC. It is important then that the subsigns from which a diagram is constructed may be themselves any kind of sign: written word, paragraphy, notation, or even a simpler diagram. But some have no significance alone and are elementary. The "zero degree" of diagram is a picture; pictograms use pictures as elementary subsigns.

A diagram is a sign which is constructed in the two-dimensional plane of the paper by juxtaposition, alignment, incidence, spacing and measure. In some kinds of diagram metric relations are important, in others not. The crucial point is that there is no binary relation of sequence as there is for the signs of speech. Each subsign of a diagram has *simultaneously* many relations to all the others which cannot be reduced⁶ to repeated concatenation on the model of speech. In the Euclidean diagram the actual order of construction is referred to in the text and plays a role in the proof. In other cases the fact that the diagram has been constructed from identifiable parts is important for their assignment to their referents and for the use of the diagram in articulated action. In the diagram from Euclid the roles of incidence (the intersections) and juxtaposition (labels to points) predominate, but metric relations are used too: BAC must be presented as a right angle, the squares as having uniform sides etc. In the CS diagram, the alignment of the two axes is crucial, the label for the whole ('figure 11.3') must be juxtaposed to it but not in it, but measure is not important beyond the point where -1.25 is shown to be on the curve. In the chess diagram the figurines must lie on the squares, but if the files aren't the same width, it's no great matter - incidence and alignment are the main construction principles here. In all these cases it is essential to the diagram's

⁶ Descartes "reduction" of geometry to algebra is an illusion. I return to this in §69.

being of any use that it be *well enough* proportioned. Versions distorted like Holbein's anamorphic ambassadors would not be *useful*. The relatively vague terms I have used to describe the articulation of the diagram are aspects of the fact that a diagram is a two-dimensional, that is planar, sign and not a linear one like speech. Any speech sign has at most two neighbours, but no such simple contiguity attends the diagram sign. Consequently any "grammar" of diagrams must be vastly more complex than the kind of model that formal systems can offer. Bertin's work is the best of a bad lot here; despite his title - *Semiology of graphics* - he has nothing to say about the philosophical issues raised by diagrams (Have they linguistic equivalents? How do they signify? Are they necessary in mathematics? and so on). He restricts his efforts to description and examples of good practice. This is the common fault of the relatively scanty literature I have been able to find on diagrams; mostly it arises from geographers or designers. (The Goodman characterisation of diagrams which I considered in Chapter 4 seems quite impotent faced with the variety and richness of any decent selection of diagrams.)

Because there are elements in a diagram which play roles other than reference, the formal articulation of the diagram must not be confused with its functional articulation. The former concerns the arrangement of the formal elements of the diagram on the page which I have been describing so far; but the latter concerns the way in which these different formal elements of the diagram are combined to give the diagram whatever semiotic functions it has. Only the icon⁷ and iconic function of the diagram can really be described formally. I now turn to the various semiotic functions of diagrams, that is, the uses to which they are put, the kinds of text act which are effected by their placing in a text.

§64 Diagrams as text acts

Diagrams are used in many ways, and it is not the case that a diagram plays only one role in a given text. Many diagrams play several roles in the working of a

⁷ See the section after next.

text, because the use of text is not a simple process. Some important functions of diagrams in mathematics are these: the instance, the illustration, the specification, the display, the resource. None of these should be seen as a kind of second-rate description. They are all ways of making sense: they may all be seen as filling, showing, constituting, or preserving the sense of their context. Thus the triangle in the Euclid text is an *instance* (or case, or example) of what the text is about, not a nonverbal kind of description of it. The figure in the CS example is an *illustration* (a depiction, a presentation) of the idea being put forward in the text, not a different version of it. The architectural drawing does not (mysteriously) tell us how things are or will be when the building is done, but gives us a *specification* (model, template) for making it. The network in exhibit #9 *displays* (exhibits, shows) the spatial configuration (topological though not metric) of the electrical components, it does not "state" them in a graphic way. The table of $N(0,1)$ is a *resource* (reference, archive) to which we refer, not a statement to which we assent.

A diagram in synthetic geometry is used to *present* a problem or answer it, to *demonstrate* a technique or make evident the truth of a theorem, or all of these. In analytical geometry these same uses persist but a characteristic new one is added: to permit the substitution of an algebraic problem for a geometric one. This is an example of the *specificatory* use of diagrams. A statistical diagram is used to *present* data, summarise it, illustrate it, show the pattern in it, to let the user judge large scale features of the dataset - is there a trend or cyclical pattern? are there obvious outliers? is there some "glaring" fact that can be made visually plain? The specificity of these uses is the reason for going to the effort of making the diagram in cases where there really is justification for it: it is too hard or impossible for the user to judge these things from the numbers⁸. The use of a plan may be for comprehension only, as in the floor plans of the buildings discussed in books on architectural history; but typically they are used for *construction*. So a plan is used as a guide or algorithm or pattern or template. These different terms all refer to

⁸ Sometimes this contrast is not real, which provides one criterion for good practice - some diagrams are simply unnecessary.

various modes of employment of such a diagram, differing in the kind and extent of interpretation to be supplied by the constructor - it may be a worker building a wall, or a machine following a specification, or an artist weaving a carpet. A map is kind of plan, but is not used for construction. What it is used for is *traversal*, which is still an articulated use of a map as a template, for making a journey or an exploration rather than an object. Tables are used to *display* large amounts of data in a perspicuous form, and are generally *resources* to which the reader refers at need. The layout of a table *reflects* the relations between its classifiers and thereby *illustrates* the dependence of the tabulated values on them. For example, the table of the standard normal distribution, which CS174 calls on, illustrates the dependence of the area below z on the first and second digits of z - though the relation would be *stated* in functional form.

All these uses of diagrams are text acts performed with them. Obviously they are graphic in character, much less contentiously so than are written words. It is also a good deal less plausible that they have spoken equivalents than it is that written word text acts do, which is part of the motivation in dismissing them as of "only psychological" effect. I will now provide some explanations of how they are effected.

§65 A diagram is a grounded iconic graph

In this section I will explain how a diagram can be used to effect the kinds of text act mentioned in the previous section. Because a diagram is graphic, that is a written sign, it has the possibilities of display and resource. Further, a diagram consists of icon and ground. The ground sets the icon in its context and shows the reader how to take the signs of the icon; the icon through its construction exemplifies or specifies. Explanations of these dicta follow.

Diagrams are graphic

Diagrams are resources in the same sense that all text is⁹, but also in a specific sense: a diagram is drawn on in completing the sense of the Written word which the text also contains. Of course the relation to some extent can run the other way! The way that diagrams display their content is peculiar to their two-dimensional character, in this consists one of their radical differences from speech, even written speech. Moreover the two-dimensionally distributed parts must be simultaneously available. Diagrams need to be *enduring* in order to act as resources and displays, for display depends on simultaneous spatial distribution on the page, resource on relocatability. They must be *two-dimensional* in order to use disposition and alignment, which are required for them to display and to model. To instance spatial relations they must *have* spatial relations. They must be *visual* in order to be useful, in the ways I have listed, to creatures with sight as their most powerful sense, in dealing with items conceptualised through sight and constructed using sight. These three - endurance, two-dimensionality and visibility - are the marks of the graphic. They make it possible for diagrams, like other text, to be a display and/or a resource.

Icon and ground

A picture can stand alone, but a diagram has a context of use. Diagrams must be read. Consequently they rely on paragraphy: a diagram does not stand alone and it needs to be connected to its context¹⁰. So a diagram consists of icon and ground: the icon is that part of the diagram which might be called its subject, the ground is the rest of it, which could in a sense be excluded without destroying its "content", though it would destroy its usefulness and point. The **icon** consists of signs for the topics of the main discussion in the text, or implicit topics. The **ground** consists of

⁹ The sense explained in chapter 4.

¹⁰ This is an aspect of making diagrams which is hard to learn.

the signs which are not in the subject of discussion in the text¹¹. The iconic relation is the one described by the classical correspondence theory of truth as formalised by Tarski: sign elements are paired to items under discussion and relations shown in the diagram correspond to the relations in question between the subjects of discussion. In mathematics it is called isomorphism. For example in the CS diagram, exhibit #4, the curved line and the vertical at -1.25 and the areas they determine are the icon signs; the baselines are grounding signs. (They are ground despite being discussed in the text: they are discussed in an *aside*¹² which clearly changes the subject from the problem under main discussion to the text's own methods.) Notice that the baselines are needed to complete the diagram's function of showing two areas proportional to two probabilities, but do not represent anything under discussion, whereas the curved line stands for the distribution of probability of attendances, the vertical line for the distinction between profitable and unprofitable attendances.

The ground consists of two kinds of signs, *labels* and *frame*. By frame I mean everything in the ground which is not label: scale lines are the paradigm example. In the chess diagram everything except the figurines is frame; a form is mostly frame; the map of Adelaide has a literal frame. There may be no frame beyond labels: in Euclid's diagrams, there are only lines making the icon, and labels which ground it in the text. However, one cannot always identify spots on the page as frame, for the functions of grounding are complex and subtle and include all the ways which are drawn on in aiding the reader's comprehension by the integration of the diagram into its physical space in the text. The chess diagram provides us with a good example of this. At first sight this diagram is almost all icon, it seems to be rather literally a picture of a chess position. But this is wrong. Actually it

¹¹ Or if they are, this can be seen to be so through a shift in level or topic. An example of this is given below.

¹² An interesting metaphor, that. In writing a great variety of asides are possible which are extremely difficult to mimic in speech, and clearly derive from the nature of writing, not from pre-existing speech acts. For example, this footnote, and the bracketed sentence from which it derives.

exemplifies a chess position just as well as a board setup in the same way would exemplify it. But it is a diagram because it a graphic version of the position, and because there are in fact labelling and framing elements here. First, consider how we know which side of the board is white's. Answer: there is a convention that it is the bottom (not the side with most white forces) - a framing element in such diagrams. Second, how is such a diagram used? Most likely, not to exemplify just a position but a game, or a theme or opening or something. It will play a part in a sequential discussion in some related text. How is that done? By using the set of **implicit** labels - explicit in some diagrams - for the squares a1 to h8. Third, the reader must know the signs for the pieces. Fourth, such a diagram will need to have its own label - which may be implicit due to its position in the text - because the setup itself is actually ambiguous. We can't necessarily tell from a diagram alone which white rook is which. Yet this may be important for application to the position of the draw by repetition rule¹³. Nor can you tell whose move it is. All this shows that a chess diagram has in fact labelling and framing elements and cannot stand alone - though a photograph of a chessboard can stand alone: but only as a picture!

The selection and articulation of the icon and the ground enable the diagram to display its "content".

Diagrams are iconic

Diagrams must be articulated in order that they may represent articulated reality, be instances of geometrical features of things, be iconic. They could not be used as templates if they had no principles of construction to correspond to the constructions to be made or found using them, be they objects, proof-followings, or journeys. As they have, they become exemplary, illustrative or specificatory. In the map of part of Adelaide the framing signs enable the reader to determine the directions of streets represented and distances along them and to connect

¹³ Cf Hacking [1985]

information from this map with that from those for adjoining areas; the labels allow the iconic relations to be recognised. Consequently the icon constituted by the points and lines can be used via the frame and the labels to plan a journey or measure it retrospectively.

Whatever is represented by a diagram must be itself articulated so that an iconic relation can be set up. Whatever has not articulation by virtue of its being made (like a cart or a cartwheel, a camp or a campaign) or being analytically conceptualised (like my childhood or the magellanic clouds) will not be amenable to diagram, though it may well be picturable. Language is not iconic. 'Hegel's triad' has no three elements - though analysis will "find" them for you! Because a diagram has an icon which has identifiable parts these must be identified for it to be used. Some of them are identified by labels and their reference in the text; some of them are identified by the frame and its use by the reader in comprehending the text.

Goodman pours scorn on the (originally Peircean) notion of the icon¹⁴; actually this is the height of irony because the only mode of reference which he recognises is in fact the iconic: the model-theoretic account of reference is actually a theory of iconism, and I adopt it for the mode of representation of the icons in diagrams. It does not apply to real language, though the unnatural "language" of mathematics inspired it¹⁵. And it overlooks the role of grounding. It would be a mistake to think that the icon *by itself* has the semantic role of exemplifying, illustrating or specifying, because none of these can be achieved without the icon's being grounded. This is the "imagery debate" point to which I referred in my CS174 analysis: the same icon can be used for all sorts of things. The grounding in the text and context determines which, in any given case. "What" may be determined in some unsayable sense by the icon, but "what about it" waits on grounding. Indeed, since we cannot say until we can read it which lines in a diagram represent and which frame

¹⁴ Goodman [1981: 231]

¹⁵ Because natural language disguises the logical form of the statement, there are not necessarily three signs in a statement about three objects, etc. See §14.

(or even label), it is clear that out of its context the icon plays no semiotic role at all and is merely an object, like a leaf or honeycomb surface.

Diagrams are grounded

The grounding of the diagram has two aspects then, grounding in the text by the labels and grounding in the context of reading by the frame. The labels ground the diagram in the text by providing a means of reference between the two. When Euclid refers to the triangle ABC we know what he is talking about, indeed he is talking about it, only because the diagram has the labels 'A' 'B' and 'C' at suitable points. The frame grounds the diagram in the reading context by providing a means for the reader to give sense to the plot signs used - the lines on Euclid's diagram are to be taken as lines, as examples representative of lines in general, but those of CS174 are not examples or representative. How do we tell the difference? We tell by the frame in CS174, which we know from prior experience has this implication. Scales always relate the icon to systems of measurement existing prior to, and outside, the diagram. The specific nature of this grounding by the frame in particular will determine whether we will use the text as specification, template, example or whatever. Of course this is no simple one way street: the purposes for which we have obtained the diagram may dominate the standing conventions which exist about the interpretation of the frame. If I have a revelation that the map of Adelaide contains a hidden code revealing the name of the Antichrist, the usual use of the frame may well be distorted.

I have composed a table (exhibit #6H) which shows the analysis of the example diagrams into formal and functional elements and the way they would most naturally be classified by form and function. An examination of the common features reveals that it is not unreasonable after the preceding discussion to suggest that a diagram is a **grounded iconic two-dimensional constructed enduring visible**

sign. These characteristics are entailed by the features of diagrams which I have discussed.

§66 The semantics of diagrams

We can see how the parts of diagrams effect the complementary roles of icon and ground, by looking at the way individual elementary signs play their part. How do labels and plot and frame signs relate to the subject of the text? The single most important point I wish to make about the semantics of diagrams is this: the various parts of a diagram have *various* semantic relations to the content, to all of which the naming relation is largely irrelevant. The relations of exemplification, labelling and enframing are all sui generis, and all have various different subkinds.

Exemplification

The first point of resistance to the Augustinian picture of language as applied to diagrams is that the icon and its parts *do not denote*. They exemplify or illustrate or specify, display and archive, but do not denote. This is the key to one puzzle about abstraction: how can a particular scalene triangle "stand for" any triangle, since any particular drawn triangle has a different set of properties to another, for example the particular relative lengths of its sides? The answer was sketched by Aristotle: we take the triangle "qua" scalene triangle¹⁶, not as triangle with sides of relative length a and b. How do we know to do this? Aristotle does not say, but it is by the relevant text which enables us to use the figure. There will always be something of this kind, as in E 1.47: "let ABC be a right-angled triangle". The fact that no more is specified about the triangle means that nothing more can be used in the proof. On the other hand we must be clear that the proof will work with nothing less. Without this reference from the text, we can take the diagram any way we like.

A line or a point does not function as a name because it endures: all the signs in a diagram endure, that is how it can be a resource. We can come back and look at

¹⁶ The logic of the operator 'qua' has been investigated recently by Lear [1982].

it again, perhaps attending differently. It also permits display: we can shift our attention from place to place in the figure knowing that the part to which we are attending will not vanish once our attention shifts. A name functions in speech as an ever renewable *transient* tool; written names play quite different roles in our use of language, there are no "proper" names in writing, or at least they don't remain so. A second difference is that names do not need to be grounded. Their arbitrariness is a once and for all thing; I could be called Nameloc instead of Coleman but this possibility is not normally a live one. The exemplars to be used to illustrate Pythagoras theorem get chosen each time, the arbitrariness is in the neglected features, the overall actual size of the figure or the color of the lines.

Icons are not sentences, nor analogous to them, because they are all subject, yet none of their parts are nouns. They also contain no predicates, and this is the source of the "imagery debate point". To suppose that a link in a network represents the *fact* that **a** is related to **b** by the relation we are interested in, is a category mistake. Rather, the icon represents the situation because *the signs* have a visible relation in virtue of the link. The icon signs are all on a level as individual signs just as the words of a sentence are. Unless asserted or otherwise employed in a speech act, a sentence is simply a string of sounds just as a diagram is a placing of marks. To be a sign each must be used. The speaker makes the sounds into a sign, but the ground must do it for the marks of the icon. Perhaps this was the realisation to which Wittgenstein was eventually driven as he grappled with the failings of his *Tractatus semiotic*. The picture theory of *language* is no good at all, and it tells us nothing about pictures either. But it has its place in the semiotics of diagrams!

The *sense* of Pythagoras' theorem is given via our intuition of triangles; a purely algebraic version of the proof, which of course can be had from an axiomatised geometry such as Hilbert/Tarski, is not an *equivalent* of Euclid's theorem, because the terms have only the content given them by the axioms. That is the point of an axiomatisation. The content we value in E1.47 is the *geometric* content however, and for that we need to clothe the bare logical bones with spatial

flesh. The diagram plays this role. If a proof is given in an "advanced" text lacking diagrams, the sense is being assumed to be provided by the reader, drawing on past experience with diagrams or by supplementing actual reading of the text with pencilled figures. To understand Pythagoras' theorem, we must know it to be about triangles and that means that we must connect any sign stating it to exemplars of triangles, if not on the page then through the experience we bring to the reading.

Labelling

Do labels denote? Yes and no. When A is mentioned in the text, 'A' is being used rather like a name for the point so labelled. But the 'A' in the diagram is not so used. On the contrary, it is there precisely for the text 'A' to refer the reader to! It is a **tag**. We can see that this situation is different to the relations of the tokens of a name if we think of these as forming a tree ramifying from the baptismal namings (on a causal theory); whereas all uses of the label for the point A refer directly to the point itself through its tag. The tag is used consciously for this purpose, and it can be because it is *attached* to the point itself. If we change the tag, we have to change all the tokens referring to it. On the contrary with verbal names, previous tokens are of course not available to present users and the role they play is indirect and unconscious. Moreover, even on the causal theory, verbal names typically have multiple baptismal uses to establish them. If one of these were changed, *per impossibile*, that would have no effect on the genuineness of later uses of the name. But of course it is impossible, and that too is an important difference.

So one source of semiotically important marks for mathematics are the physical brand and tag, early forms of the written label. One reason they are important for the semiotics of mathematics is their role in the evolution of counting; a second is their close involvement with the origin of writing, which I have mentioned in Chapter 5. They can perhaps be traced evolutionarily to self-decoration, which will not be considered here, but those which are of concern here are probably very late in origin compared with speech. It is true that a brand or

tag functions as a classifier, but it is not primarily a descriptive classifier so much as an economic one: what it declares is, this is mine! Goodman¹⁷ treats labels as sometimes names, sometimes predicates, but apart from the sheer inconsistency of this line it is anachronistic: labels are a more primitive semiotic device, in the form of brands and tags, than are either names or predicates. Goodman's basic function for labels is as classifier; this is correct, but his concomitant claim that labels always come in sets of alternatives seems to me untrue. Moreover he overlooks the fact that the simple label must be attached to its object. He is led to ignore this fact by his wish to treat of verbal as well as nonverbal labels. But this is a mistake, I think, because it leads us to overlook the priority of the tag or brand in the evolution of the label. Verbal labels are so only at one remove - literally. The sequence is: first, nonverbal, "physical" label or tag ; second physical, attached label with inscribed mark; third, verbal mark corresponding to that on a label used as a metaphorical label. (This is the evolution of writing according to Schmandt-Besserat, viewed oddly.)

Labelling elements such as title, legend, key, scale label and tick mark labels fall into two main classes. Apart from the simple sign like 'A' there are more verbal examples, like the title of the chess diagram. The title is often even a sentence. Labels for plotting symbols may be genuine written names, and words used in scale labels may "refer to" a system of measurement, thereby involving the data diagram with that whole semantic field. The semantic relations set up in this way are among the most complex, going beyond even those entrained by the use of such terms in plain text. Of course, these more complex labels are also serving framing functions. One of the most characteristic features of diagrams, indeed of written text generally, is the tendency for several functions to be served at once.

¹⁷ Goodman [1981: 52-68]

Enframing

The fact that some texts include decorative borders, printer's flowers, gold blocking or whatever should not delude us that all framing elements are decorative. Noone would think this of labels, but a similar idea is to be found in the dismissal as merely psychological of "aids to comprehension" whether part of a diagram or elsewhere in a text such as the title page or marginal headings. It is true that in one sense framing signs do not convey the content of the text, and "only" make it easier for the reader to apprehend it. This does not mean that they are otiose. If the reader cannot apprehend the sense without it, but does so with it, then they are in the strictest sense constitutive of the sense apprehended. *All readers are limited cognisers.*

The borderline is the earliest form of framing I can think of. It seems to be on a par with the brand in respect of antiquity, indeed the two are clearly related developments in property relations. Here again we see the economic element in the constitution of the semiotic. The territorial boundary is a well-known zoosemiotic feature which certainly has very early examples in human culture. One source of the line as a geometrical element is certainly the need for the creation of boundaries; another may be the representation of the sun as the origin of the circle. These may be of great antiquity; only in connection with techniques for producing straight line and circle do they become the origins of geometry. The tools involved in the construction of built objects with social significance and shape desiderata, such as altars¹⁸, are found only recently in postlinguistic cultures; this fact must be remembered when trying to characterise their separate semiotic nature. Diagrams clearly derive largely from these sources.

Framing is a very general semiotic feature. The linguistic polarisation of the sentence into topic and comment is just one kind of case. In the visual mode, the contrast between figure and background is the form it takes, which I have adapted into my terminology of icon and ground, which is intended to be more

¹⁸ Cf Seidenberg [1962a,b]

specific and confined to diagrams, not pictures in general. The root phenomenon is the focussing of conscious attention on some elements of the environment while holding others not ignored but "bracketed". One is not prevented from attending to the frame, one often does so in the course of examining a diagram, but these expeditions to the fringes are carried out in order to make firmer our grip on the centre, a kind of epistemic boundary riding. The frame does not name what is being presented in the diagram, though it may be used to name the *diagram*. Its relation to what is being presented is entirely indirect, it allows the icon to make its presentation. The relation of framing and framed signs is originally that of container and contained; one might expect my claim that enframing is not derivative from more primitive semantic relations to be welcome to Goodman, given his partiality for mereology.

A more traditional way to look at the enframing relation, the function of grounding signs, though perhaps surprising here, is to say that they are part of the text's meta-language. I have rejected any simple distinction between object language and metalanguage in mathematical text, but this does not mean that I reject the distinction itself, or the idea of semantic ascent. On the contrary, I shall argue in Chapter 8 that text has many strata; but they are like geological ones, not those of the set-theoretic hierarchy. They are only partially distinguished, playfully confused, unfounded and each of indefinite level. But they are there. The sign 'Figure 11.3' in CS is an example which will indicate the complexities here. It works to show the reader how to connect one piece of the text with others - clearly metatextual in a simple sense. There might be a list of figures in the back of the book, or a mention of it in the index. That gives us two levels of ascent. The text uses it to refer to the figure of course, so we really had one level already. There are many such figures through the book and the numbering systems ensures that this sign refers us to all the others. And so on. (The text refers to itself in the aside about the two scales, showing how little Tarskian paradoxes matter to real discourse.)

§67 Tables, lists and forms are diagrams

I can now justify my inclusion of tables as diagrams, explaining the fact that they are often treated together in the literature on diagrams¹⁹, though generally with no explanation as to why. The salient features of diagrams which emerge from my discussion are that they are graphic, a displayed resource, showing iconically, grounded by label and frame. This admirably describes a table like exhibits #1 and #6H, or the table of the standard normal distribution called on by exhibit #4. A table consists of values labelled in two ways organised to show the labels in the two directions of the page. (Of course there are exceptions and extensions which make this definition "wrong".) Thus the elements of a table are the systems of labels and the data values, which are used as plotting signs, with their position in the table significant. Paragraphy is crucial too, alignment at least but often various rules, spacing and type variation. A table is *graphic*, that is two-dimensionally organised, visible and enduring; these are the features which permit it to be used as a resource whose use rests on its displaying its content. The point of table design is to make the display facilitate such recourse to it. A table is *iconic*, each column represents a value of one independent variable, each row a value of the other, each intersection the value of the dependent variable. The table pictures a function of two (or more - more complex tables are common) variables. A table is grounded, the stub (row labels) and the headings (column labels), and perhaps a title, providing a framing system which shows the reader how to interpret the values entered at each intersection. Usually there is some accompanying text which provides further information on what these labels mean, making them labels more strictly speaking. Where not, as in exhibit #1, the reader must already know how to use them to ground the table in prior text and action.

A table displays a two (or more) dimensional function, so naturally a one dimensional function can also be displayed. The result is a list, which is also a diagram, but one which is less clearly so. 'Diagram' like most terms, shades off into

¹⁹ E.g by Tufte [1984], Bertin [1983], Holmes [1984].

related terms, and in this direction is such a transition. I distinguish a list from a sequence²⁰. A *list* is a sequence of labelled values. I do not count as a list a mere transcription of a sequence of related words; this is what Geach²¹ understands by the term in his discussion of the "logic of lists". The basic form of the list is this: first, edwin; second, ronald; third, coleman. A list uses pairing as well as sequence. We must beware of being misled by the fact that this list, like some others, can be read out loud, into thinking that a list is a speech sign. The values may be any signs at all, but must in the same list all come from a *range of values* of a *variable* (in some cases, one given by the list itself, in others by a rule). The word 'value' is used because the originally important examples of lists are lists of amounts - of money or beer or whatever. The original function of the list was in accounting: organised records of financial or at least valuational items; other early list topics, of kings and names - discussed by Goody²² - are perhaps explicable as diversions and practise for scribes. The great cognitive consequence of the introduction of lists, as Goody makes clear, is the sharpening of the practice of classification which it makes possible. The sequence of the list is shown in the way it is shown on the page: the basic system is one item per line (or column if the writing goes in columns). I count lists as diagrams because they are the simpler cousin of tables, and because the label as such is not found in written speech. The fact that a list is easy to linearise, as in my example, shows that it is less clearly distinguished from written word than is a table, and provides some explanation for the temptation to try to reduce tables in a similar way "in principle".

However, a table is a quintessentially written sign. Its contents might be presented in other ways, but it cannot be thus "translated" because the display and resource functions of two-dimensional tabular presentation are its semiotic point. A table has several virtues over the "equivalent" linear list. It can be read two ways;

²⁰ There is some etymological justification for my possibly slightly unorthodox distinction here: the root meaning of list is *border*, related to the division of the hair or the ground after a kind of hoeing!

²¹ Geach [1962; ch7]

²² Goody [1977; ch 5]

it can be grasped as a whole; it is economical of space and access time. A table cannot be reduced to a linear list cognitively equivalent to it. The idea that this is possible is a mistake. In order to present the information in the table in its linearisation it is necessary to create lexicographic labels; their mutual relations must either be ignored or comprehended. If they are ignored, the pattern given to the values by the table is lost; if they are comprehended we must ask how this is effected, and whether one can achieve it if one has never used tables, and whether it can be done at all for large tables. Once again I reject any response which depends upon the phrase "in principle". My interest is in how the signs of mathematics actually function, *not how they might for gods*.

Finally let me briefly indicate why I count forms (questionnaires) as diagrams too. The form is an interesting item semiotically, and somewhat neglected. It has two states: blank and filled. Other texts do not show this particular ambivalence. As blank, a form might be classified with diagrams because the problem of reading it, and the problem of designing it to be read, are similar to those for other kinds of diagram. But when filled, an individual form becomes a list, with the fill-ins as values labelled by the parts of the blank form. (I do not deny that it is sometimes little more than a list of answered questions.) But forms are never found as one-offs: there are always many different copies of the blank form filled in differently. They are filled in differently so they are not all copies of the same sign anymore. On the contrary, the actual filled in values are destined to become the body of a table or tables with the common part of the form contributing the labels for the table's stub and headings. Thus a form is a diagram which aims to become the basis of a table! A form is well named, indeed filled-in forms are the only kind of sign for which the claim of a common "logical form" is unexceptionable. But it is a diagrammatic, not a linguistic commonality!

§68 The logicist attempt to extirpate diagrams from mathematics

In the rest of this chapter I will show that the logicist position on diagrams and constructions is wrong: *diagrams and constructions are essential in mathematics*. They are essential for us to come to know mathematics and for us to use it. And, as I shall show in Part III, these in turn are essential for mathematics to have sense. In this section I will explain why logicists wish to exclude diagrams from mathematics, in the next I will show why this is wrong, and in the last I will add a further argument which will lead us on to the next chapter.

Many mathematical texts contain diagrams and constructions. I have illustrated this, and discussed the various kinds of use made of them, earlier in this chapter. For many of these examples, it could be said that they are not essential to the mathematical reasoning of their text; I will return to this claim and rebut it in the next section. For the present, consider the example from Euclid, where this claim is not sustainable. As in many propositions in Euclid, according to the standard arrangements of the text explained by Heath²³, after the enunciation of the proposition, its ecthesis and specification, comes the construction. He quotes Proclus: "*The construction or machinery adds what is wanting to the datum for the purpose of finding out what is sought.*" In the example, exhibit #5, the step is quite explicit of introducing the points H, K etc and the lines AD etc. The main reasoning is about these items. One shows the desired property to hold of the diagram using these auxiliary signs. The step taken in the construction is what Peirce calls "theorematic"; such steps are the source of novelty in mathematics. Others Peirce calls corollarial. In Peirce's philosophy of mathematics there is²⁴ an insistence on the necessity of diagrams in mathematical reasoning, what he called the iconicity of the mathematical sign, which is related to this distinction between theorematic and corollarial reasoning. Theorematic proofs, which need constructions, are the informative, surprising ones. This is a development of the Kantian distinction between analytic and synthetic. The theorematic steps are *not*

²³ Heath [1956: 129ff]

²⁴ Cf Peirce [1976; passim, e.g iv, 38, 49]

analytic and make the arguments in which they occur nonanalytic. They depend on the use of constructions and essentially involve diagrams in mathematics. Central to Kant's philosophy of mathematics is the doctrine that mathematical claims are synthetic, not analytic. This relies on the concept of constructions in intuition, discussed already in Chapter 3. A main plank of logico-formalist hegemony is that no such use of intuition as is involved in the employment of these items is needed in mathematics. Russell is quite explicit about this²⁵: "...the supposed peculiarity of mathematical reasoning has disappeared...No such intuition is relevant to any strictly mathematical proposition." Logicism has been generally abandoned as a complete philosophy of mathematics, but it is still commonly accepted as part of the logico-formalist hegemony, what was taken as established by Russell and his heirs, that the partial successes of logicism in "founding" mathematics on set theory and providing sound axiomatizations of various geometries, at least prove Kant (and so Peirce) wrong: all mathematical inferences are analytic, not synthetic. The demonstration of this claim is the construction of geometry within *Principia Mathematica*. I will discuss the "reduction" of mathematics to logic in the next chapter. If one accepts that *Principia Mathematica* is an acceptable formulation of mathematics, then diagrams and constructions are **not** needed in mathematics.

They are therefore *best left out*. They are best left out because they are a notorious source, or so it is claimed, of fallacious reasoning. For example, there is a famous "paradox", proving that any triangle is isosceles²⁶. The inference rests on the quite unjustified assumption from a poorly drawn diagram that a certain point is inside the triangle. Actually it cannot be. The source of such errors is said to be the unreliability of the intuition which is employed in synthesizing our perception of such a diagram. The logicist argument for the eliminability of diagrams and constructions is that all the desired propositions can be proved rigorously in a purely algebraic fashion, as *Principia Mathematica* allegedly shows. The

²⁵ Russell [1937: 458] see also Russell [1919:145]

²⁶ E.g Northrop [1960:100ff]

observation of Hintikka²⁷, that "theorematic" steps can be recognised formally as instantiations, would seem to reinforce this conclusion.

The change from the clearly diagrammatic text to the present algebraic linear text (in higher mathematics, anyway) seems to have been little investigated or remarked; for example Hatcher passes over his repetition of it with the remark that his presentation of Frege's diagrammatic system in modern notation is merely a "notational simplification" and does not "falsify the original"²⁸ This is a contestable view and I shall contest it in the next chapter. In any case, this change foreshadowed the burying, depreciation and suppression of diagrams which reached a high tide in mid-20th-century mathematics, so that now we have the ludicrous situation that many standard books about geometry contain *no diagrams at all*. (One interesting exception to this is in the rise of category theory - which cannot be reduced to set theory, oddly enough.)

§69 Why mathematics needs diagrams

I shall argue in this section that *Principia Mathematica* does not show that mathematics can dispense with diagrams and constructions; rather, its lack of diagrams shows that it is not an acceptable formulation of mathematics. For Diagrams are needed to pose interesting problems, to give sense to the propositions which we elect to demonstrate to solve these problems, and to motivate and support the particular proofs of these propositions which we actually employ for these demonstrations. Moreover, diagrams are no more dangerous than other signs.

I do not say that the logicist derivations are lacking something, I say that they are not adequate versions of the mathematics they purport to represent: what "reduction" omits is necessary for the mathematics, if not the deductions. But why identify the mathematics with the deductions alone? The essential point is that real mathematics arises from problems and consists in their solution; but problems

²⁷ Hintikka [1968]

²⁸ Hatcher [1982:73,76].

have no place in the logicist discourse, because questions about the choice of propositions to prove and means to prove them are "merely" pragmatic. This "merely" is an unjustified stipulative redefinition of mathematics.

The functions of diagrams identified above indicate simple answers to some puzzles created by the logicist attempt to narrow mathematics to deductions. Diagrams mediate the application of mathematics through posing problems and specifying answers; they convey mathematical knowledge through display and as resources; they form part of its construction through their use as illustration and instance. Mathematical thought contains synthetic and analytic strands; to identify the analytic ones and declare that they exhaust mathematics is a distortion which creates pseudoproblems.

Logicists would be happy to grant a role to diagrams in *applied* mathematics, though they would offer no account of their "logical" role there; so I shall not argue for it. My point against the logicist conception is that the idea of "pure" mathematics as primary and its applications as secondary is quite the opposite of the true situation. Pure mathematics gains its sense and rationale from its applications. The launchpad of mathematics is the problem, not the solution. The logicist account pretends that we are only concerned with the solutions to largely unstated problems²⁹. Logicist philosophy has nothing to say about the choice of the few predicates we investigate from the infinity we might. The reason is simple: they arise in trying to solve problems, which logicism does not describe. There is a venerable tradition in mathematics, to which Euclid's Elements in fact

²⁹ This is so far from the truth that we really need to consider why such a misrepresentation is possible. My suggestion is that we look at the content of *Principia Mathematica*. *Principia Mathematica* solves no mathematical problems at all, it is entirely well-known mathematics. So we should beware of taking it as any kind of account of anything more than the completed product of mathematical thought. If we wish to see something of the process of mathematical thought in order to understand it rather than its dead product, we may examine the marginal elements of PM, the introductions and prefaces. There we find grappling with ways to make definitions, and *the use of diagrams*: in the introduction, pieces of notation are displayed as objects and discussed; whereas in the "text proper", we find only a single stream of signs of the formal language. I will discuss this further in the next chapter.

belongs, the tradition of geometric problems³⁰. This is familiar through the most famous examples, the problems of cube doubling, angle trisecting and circle squaring with ruler and compass. Another very famous one is the one Gauss solved: the problem how to construct regular polygons with ruler and compass. Gauss showed that this is possible for only a small class of polygons, those for which the number of sides, n , has a special form. So 13-gons cannot be constructed, but 17-gons can. Gauss was so pleased by this piece of work that a text of it was intended for his tombstone. But notice this: that text is a diagram illustrating how to make the construction. Although one can now demonstrate that no such diagram with the right properties can exist for bad n but can for good n , the real point of the problem and of Gauss' solution is to actually *make* objects which have 17 equal straight sides! Logicism completely distorts this geometrical tradition.

Another aspect of choice on which logicism is entirely silent is the means of proof of theorems. It is common in mathematical talk to distinguish between explanatory and nonexplanatory proofs³¹. Indeed, Lord Rayleigh went so far as to coin an epigram on the subject which I have used as an epigraph to a later chapter.³² The point here is quite simple: from the logicist point of view, there is nothing to choose between any two proofs of a theorem which are valid. But from the point of view of real mathematics, long proofs are not liked, for the simple reason that we find it hard to grasp them. For logicism a proof with a billion steps is fine; but it can have no role in our mathematics³³. Even among short proofs we can find the distinction in question. It clearly plays a significant role in the creation and use of mathematics - Gauss invented four different proofs of the law of quadratic reciprocity in an effort to understand it, and six of the fundamental theorem of

³⁰ Cf Knorr [1986]. Logicism is conspicuously silent about the kinds of text act involved in propositions like Euclid I.46. There is still proof involved, but the proposition is not easily thought of as stating a fact; it is basically a technique for constructing a square on a given line, a technique which is needed in constructing the Pythagorean diagram.

³¹ Steiner [1978]

³² "Some proofs merely command assent; others woo and charm the intellect."

³³ This issue has been discussed in connection with the four colour theorem. See Tymoczko [1979], and Detlefsen [1980] and Teller [1980] in reply.

algebra. So philosophy of mathematics must have room for it. Occasionally in off the record discussions one comes across mathematicians admitting to the role of criteria of "depth" "interest" "generality" and so on in the comparative evaluation of theorems and theories.

A third choice on which the logico-formalist hegemony is silent is that of axioms. Why study *this* set of axioms rather than one of the infinitely many others? Hilbert, one of the prime movers in the rigorous founding of geometry, showed³⁴ the unavoidability of an axiom "of space intuition". (In later writings³⁵ he explicitly states that he sees **all** signs in mathematics, diagrams and numerals alike, as "expressions of our space intuition" in their role in mathematics. This point of view is not compatible with the logicist view on diagrams. I return to it in the next chapter.) Hilbert showed that a system of geometry, that is a system we would call geometry, must have such an axiom. This creates a difficulty entirely parallel to Russell's embarrassment about the axiom of infinity. If you are giving an account of mathematics and not merely of possible formal inferences, these axioms are needed; but why just these? Russell flagged use of the axiom of infinity in *Principia Mathematica* as an *empirical* assumption, but this is plain wrong³⁶.

These deficiencies of logicism show that excluding diagrams from mathematics is not acceptable; but what of their danger? Recall that logicism was created when everyone was obsessed with the security problem for mathematics. Actually, diagrams are no more dangerous than any other sign. Of course, a poorly drawn diagram can lead you to make wrong inferences. But then so can a poorly arranged addition, if you don't line up the columns for example. Littlewood, Hardy's long-time collaborator and the best analyst in the world for a time, rejects the bogey of "dangerous" diagrams thus³⁷:

³⁴ In Hilbert [1971]

³⁵ Hilbert [1902:443]

³⁶ Less clearly for Russell, whose system is interpreted from the start; more so for modern logico-formalists whose systems are uninterpreted.

³⁷ Littlewood [1986: 54]. Of course, by "pictures" Littlewood means what I call diagrams.

"My pupils will not use pictures, even unofficially and when there is no question of expense. This practice is increasing; I have lately discovered that it has existed for 30 years or more, and also why. A heavy warning used to be given that pictures are not rigorous; this has never had its bluff called and has permanently frightened its victims into playing for safety. Some pictures, of course, are not rigorous but I should say most are (and I use them whenever possible myself). An obviously legitimate case is to use a graph to define an awkward function....pictorial arguments, while not so purely conventional, can be quite legitimate..."

Littlewood goes on to give two examples of such arguments. If the best mathematicians accept this as rigour, there is no more to be said.

§70 Tu quoque

Why do logicians think that intuition is not involved in reading and grasping their own texts, which are presented as material objects to the senses? Is there no possibility of error in the following of derivations in *Principia Mathematica*? Didn't Russell and Whitehead have some difficulties with creating some of the proofs? Frege's system eventually turned out inconsistent, after all (as did one of Quine's). The unimpeachable security of the reasonings of logicist mathematics is their great advantage over the earlier form of mathematics, according to Russell. A strange claim in a book (Russell [1919]) which confesses to being unable to deal with the Russell paradox! Mathematicians still have to write and read proofs - can we be sure that intuition is not involved? Hatcher³⁸ suggests that the arithmetisation of analysis, when completed by the Peano axiomatisation of the natural numbers, "had substituted appeal to set-theoretic intuition for appeal to geometric intuition". One might fill this out by saying that our ability to visualise the divisibility of the continuum was supplanted by our intuitions about inclusions expressed by the rules for operating with the quantifiers. This was itself originally a geometric intuition, as is evidenced by the extensive use of diagrams by 19th century logicians such as Boole, Venn, Peirce, DeMorgan, Carroll and Frege in particular in his *Begriffsschrift*. But this was rather suddenly replaced by a different intuition

³⁸ Hatcher [1982:70]

when the Peano/Russell notation for logic was adopted and the set-theoretic mode of mathematical discourse developed.

When do we grasp a geometrical content? Can we grasp one even though we follow a proof of it, without the right kind of geometrical grounding? It is common to distinguish between being able to follow a proof and understanding it. The fact that geometry can be presented in texts without graphics does not prove that geometry does not need diagrams. **Texts need readers.** The logicist view ignores the contributions of the reader to the comprehension of the text. There are several aspects to this. First, the nature of the reader's readiness - readers of geometry books typically have concepts prepared by the use of diagrams. Second, the process of reading is a visual one, involving geometry in its origins and in its details. Third, the supplementation of the text which the reader supplies; mathematical readers "figure" as they go, and they structure the text in building a grasp of it as a whole. The logicist view pretends that a mathematical text has no context, whereas all such texts presuppose other mathematical texts. (I enlarge on this in Chapter 9.) The real question is whether you can learn geometry *and understand it* without any experience of diagrams, anywhere, anytime. Moreover, the process of step by step construction which the mathematician goes through in proving a theorem to herself or another is a very diagrammatic process even for the proofs of abstract algebraic facts, consisting of the the perspicuous display of the transformational possibilities of certain crucial formulas, which can be made evident from their form. What's more, a diagram of the proof is very commonly employed in this process. Example: in proving Lagrange's theorem about subgroups of a group, which is entirely abstract and algebraic in character, the concept of a partition in the geometric sense may be used to get the idea of the proof across. Continuity does not enter into such a procedure; what does, is the spatial display of parts of the object and parts of the discourse.

There is another argument of the "tu quoque " kind which can be made against the logicist view. Recently, Hintikka and Stenlund independently³⁹ have given formal distinctions within the class of valid proofs which can be taken to revive the analytic/synthetic distinction. Indeed Hintikka explicitly argues that the distinction he proposes is the one for which Peirce was groping in his homely 19th century way. If such a logical distinction can be established to play the role in question, then the logicist view on constructions is refuted, in an ironic way. The point is that algebraic formulations are iconic too. This Peircean point turns the tables on the logicist. Far from showing that even theorematic steps can be reduced to logic, the Hintikka/Stenlund point actually shows that logical notation is really diagrammatic too! Peirce claimed that all mathematical reasoning is iconic; in the next chapter I shall argue that iconicity is a key feature which notation shares with diagrams.

³⁹ Hintikka [1968], Stenlund [1974]

Chapter 7

Notation

§71 What notation comprises

By *notation* I mean written signs introduced specifically for special purposes of mathematics, music or whatever. 'Notation in mathematics' means the special signs of mathematics. I distinguish diagrams from notation (and both from the writing system), although I shall be arguing in this chapter that actually notation has diagrammatic properties. Notation consists of signs which must be written, being jointly derived from written forms of spoken signs, and from diagrams. I will call them N-signs.

The exhibits contain a great variety of examples of notation some of which I have highlighted in yellow on each of them. There are several kinds of N-sign. First, peculiar use is made by mathematics of single letters like x f Z . Second, and most obvious, are the specially introduced, specifically mathematical signs, which do not occur in written speech, such as $-$ $=$ $+$ $-$ Σ \sin \log . Of course they are related to words; I will come to what that relation is, shortly. Equally apparent to the eye as characteristic of mathematics are paragraphic signs, such as bracketing and underlining, which do occur in written speech of course, and layout features such as superscript and subscripting, which do not, for example $()$ \underline{y} x^y . A sign like $\sum_{i=1}^{\infty} a_i x^i$ illustrates all three kinds combined. I will discuss the role of paragraphy and the relation of the paragraphic sign-system to the others in the next chapter, so I put it aside for now; I shall make no points here about notation which leans heavily on such features, like a matrix of partial derivatives for example, and will restrict my attention to the example in exhibit #7 which is firmly of the kind which the logico-formalist characterisation of mathematics takes as typical. I have chosen it because it is a typical example of what is presently regarded as rigorously presented mathematics, for it contains lots of notation but no diagrams; and it is short enough to be considered as a whole. (Its mathematical content is of no interest here.)

Mathematics contains various notational subsystems which can sometimes be in conflict. Some signs have many different uses, such as '+' and '0'; and many

different signs are used for one and the same thing in different texts. The derivative of a function has several quite different notations. Cajori deplores this lack of a universal univocal notation system and argues in his concluding chapter that the study of the history of notation shows the need to do something about it¹. But really it is an important fact that there is no universal system, and in fact there *cannot be*. Notation is *necessarily local*, at least in part. There are several reasons for this.

First, in many texts it is necessary to introduce signs for items specific to the cases being considered, such as an unknown or a coefficient of an equation. It would be intolerable if the unknown quantity in every quadratic equation problem had to have its own sign because was a different thing; intolerably burdensome on the memory and on the invention of new signs, and intolerably clumsy since it would lose us the advantage of the uniform procedure for dealing with such problems which the quadratic formula records by using 'a' indifferently for the quadratic coefficient in any application. Second, the processes of making and using mathematics require the introduction of new and possibly temporary notation for new ideas. It is inevitable that with thousands of people producing new mathematics all the time, that this will result in different notations being introduced for what is essentially the same new idea, and the same notation for different ones. Third, such rival systems to the same ends must be tried out, as in the example of the Newtonian and Leibnizian notations for calculus, or those of Frege and Peano/Russell for mathematical logic. Such rivals tend to live on because the superiority of one over the rest is never absolute or clearcut or psychologically acceptable to everyone. Moreover the texts in which systems are put forward remain in use so that there is pressure to continue to use both of two rival systems in order to extend existing text². Fourth, different uses are made of the same sign from text to text and even in different parts of the same text, because there are really not that many different signs available (particularly in print until recently). It is important

¹ Cajori [1929: §§712-50]

² The Newton/Leibniz notational struggle comes to mind here!

not to overload the reader with the need to memorise too many new signs, and old ones' prior uses are often irrelevant to the matter in hand and not likely to be called to the mind of the reader. So they are often redefined *pro tem*. Disclaimers are rife in mathematical text, that the danger of confusion caused by this practice will be avoided by context. Of course the idea of context on which this relies is not one which philosophy of mathematics usually explains, much less practising mathematicians, but I shall do so in Chapter 9. It is the nature of text that it is read by a cognitive being of limited attention.

Some notation is near universal in mathematics, because it is used in or relevant to almost all mathematical texts, particularly the signs for the integers and arithmetic operations, but even this agreement is not quite absolute, for example 3.000 means different things in Paris and Adelaide, and the agreement tails off from there. Considering the limited extent of this universality, it is amazing that the idea has such a popular hold that mathematics (meaning notation) is a universal language³, quite apart from the mistaken idea, which I criticise below, that it is a language at all.

§72 An example described in logico-formalist manner

Let us examine in detail a brief recent text⁴, exhibit #7A. Scrutiny of this text shows that it contains many paragraphic signs (indicated in red), no diagram, much written word (left unmarked) and a great deal of notation. I have highlighted the notation in yellow. The first thing that strikes one about the result of this highlighting is how the N-signs and W-signs are entirely intermingled. I will come to the significance of that shortly. Let us begin by trying to look at his text from a logico-formalist point of view.

³ Any number of expository books ring changes on the title " Mathematics, the language of science".

⁴ Hansen [1984]

First, we consider the *syntax* of the notation. This means we give a list of the primitive signs, and rules for their combination which delimit the classes of terms, wffs and proofs.

The primitive N-signs in this text are as follows

| | |
|-----------------------------------|---------------------------------------|
| S1 (paragraphy): | $() \{ \} [] ,$ |
| S2 (individual variables): | $S P x y d g i n$ |
| S3 (individual constants): | $0 1 2 \infty$ |
| S4 (individual constants): | $+ . < \geq \in \leq = \neq - \Sigma$ |

The modes of composition actually employed with these signs are juxtaposition, subscription, and superscription, but let us pretend with logico-formalism that the latter two are not essentially different from juxtaposition (in the next chapter we shall refute this idea). The formation rules are rather complex; those for further individual variables (constants) would take in part the recursive form

R1 If X and Y are variables, so are $X+Y$ and XY , X_Y and X^Y , the result being a constant if both X and Y are.

The main rule for forming wffs would be something like⁵

R2 If X and Y are terms and Z is a constant term from the class S4, then XZY is a wff.

R3 Logical compounds and quantifications of wffs are wffs.

On the basis of these rules and some others, the complex signs which actually occur in the text would be recognised as terms or wffs.

C1a(individual terms): $x_1 y_1 a_1 x^2 y^2 x_1^2 xyx d(x_1)$

C1b(individual terms): $(S, +, .) (S, + \dots \leq) \sum_{i=1}^{\infty} a_i x^i$

C2a(wffs): $y \in S \quad x \in S \quad xy \in S \quad y_1 \in S \quad x_1 \in S$

C2b(wffs): $x \geq 0 \quad x^2 \geq 0 \quad x^2 \geq 0 \quad x_1^2 \geq 0 \quad a_i \geq 0 \quad 0 \leq i < \infty \quad [g(x)]^2 \geq 0$

$d(x) \leq 0 \quad 0 \leq xy^2$

C2c(wffs): $P \neq \{0\} \quad d(x) \neq 0 \quad d(x) = 0 \quad \dots\dots\dots$

⁵ I am being illustative here, not completely precise.

..... $0 \leq x d(y)x = -d(x)$

' $0 \leq x d(y)x = -d(x)$ ' is an abbreviation for ' $0 \leq x d(y)x$ and $x d(y)x = -d(x)$ '. Abbreviations are endemic in mathematical text, and I shall suggest below why the logico-formalist account incorrectly conflates them with defined usages. Actually they are rarely explicitly defined. There are 35 different signs in the class C2c, mostly of the form $A = B$, some of the form $A \neq B$ and a few combining signs of the forms $A \leq B$ and $B = C$. I have not copied them all. (The rules which would ensure that all these signs were indeed wffs and terms would be more complex than those I gave, there would have to be provision for terms of the form $d(x)$ and $\langle S, +, \cdot \rangle$, but these details do not interest us here.)

The logico-formalist syntax is completed by declaring certain sequences of wffs to be proofs, namely the closure of a list of them called 'axioms' (which would include identities of the form $X = X$) under the operation of derivation. One sequence of wffs, A , is derived from another, B , if A consists of B with one appended wff, as long as the appended wff is the conclusion of a "valid inference" from premises which are elements of B . "Valid inference" would be clarified by giving formal specifications of the patterns of premises and conclusion which are accepted in the system. This would be necessary to give rule R3 properly anyway; some natural deduction system for the lower predicate calculus would do for the present text. Final elements of proofs are called (derivable) theses of the system.

So far we have only a description of a kind of a meaningless calculus in which signs can be generated which will include those used in the text and will not produce nonsense signs of the kind which clearly would not occur in any extension of it, such as $+))x^S =$. The second stage in the logico-formalist account is the semantics for this uninterpreted system. One defines a *model* for the theory of which the text is a fragment, as a set equipped with certain suitable relations, and a map from the individual constants to the set and its relations which satisfies a natural list of conditions to ensure that terms correspond to elements of the set,

and wffs to relations which do or do not hold among the set elements corresponding to the constants in the wff. Then a valuation of the wffs consists in an assignment of either "truth" or "falsity" to each "atomic" wff according as the relation to which it corresponds holds or not. Wffs involving logical operators are assigned T or F by familiar (or in some cases, not so familiar) rules. A wff is valid if it comes out T in every valuation. A valid wff is a theorem. In nice systems, theoremhood and thesishood coincide. Gödel showed that in reasonably strong ones like that of *Principia Mathematica* they don't: theorems outrun the means of deriving them as theses. But for the text we are considering no such possibilities are relevant, since it will be formulable along the lines which I have sketched in a system which is of a kind which has been proved to be consistent, which means that anything derivable is valid. This is what we need to explain that the theorem proved (i.e. derived) in Hansen's paper is necessarily true, namely that its interpretation in any model of the theory is a true relation.

A much longer, more detailed and precise, version of this sketch can be found in books on the foundations of mathematics, like *Principia Mathematica* or Church [1956] or Hatcher [1982]. Many nice points arise in connection with the preliminary explanations required in setting-up formal systems, but I shall not go into them here. In any case, the logico-formalist account of the example text has a number of shortcomings; in the next section I will point some of them out.

§73 Some omissions from this account

If we return to the text itself (exhibit #7A) from the logico-formalist idealisation we can see that there are various ways in which it is inadequate. I will argue for three things. First, there are many different text acts effected in it, where the logico-formalist account would recognise only assertions. Second, it does not make clear what has been learned by someone who reads it with understanding. Third, contrary to the logico-formalist account, the terms do not refer nor do the

wffs state anything. These difficulties all stem from ignoring the presence in the text of non-notational signs - Word and Paragraphy.

There *are* assertions in this paper, for example the main theorem (W25/26) is one. Another is W8 - "Before proving the main theorem, we need the following lemma.", but this is not a mathematical claim. It is an example of the text discussing itself. There are several other remarks of this kind, and such remarks are common in mathematics, but not modelled in the logico-formalist account. Other text acts which are effected in the paper include definition, proof, inference, calculation, grounding and illustration. Logico-formalism ignores the differences among them, but if we consider some of them we can see how this is unacceptable. Consider the text act of illustration. The whole of §3 of the exhibit is quite apart from the main deduction, not needed at all. Actually it gives an example of a ring in which the conditions of the main theorem are not quite satisfied, and in which consequently the conclusion of that theorem does not hold. What the logico-formalist account omits is any explanation of why in this text this deduction follows the main one. I am not saying that its explanation for this conjunction is wrong, but that it sees no need for one. But just such conjunctions are constitutive of knowledge: we wish to know the limits of applicability of our theorems, and whether their conditions are really necessary for their truth, and so on. This kind of concern is natural to the *mathematically educated* reader, and she realises the point of §3 without further thought. But its inclusion is not accounted for by the logico-formalist account of mathematics. This is a further illustration of the *choice* problems which were listed against the logico-formalist hegemony in the previous chapter. Why do we communicate just the sequences of theses that we do, out of the infinite variety of acceptable derivations? Logico-formalism has no answer to such questions.

Secondly, I suggest that the logico-formalist account does not make clear just what is known by someone who reads this paper with understanding. Putting off to Chapter 9 the question what are the conditions for reading to be "with

understanding", the present point is this. While the theorem of this paper is stated in the abstract, P7, almost in W-signs alone, and could clearly be purged of the two tokens of 'S' which are its only non W-signs, it is nowhere stated in notation alone or any where near it. In fact if we examine W25/26, once again it seems to be the N-signs which could more easily be replaced by W-signs than the reverse. Of course, *Principia Mathematica* shows that this is wrong does it not? That is to say, Logicism has shown, surely, that it is possible to replace all the W-signs in mathematics using the logical N-signs pioneered by Frege and perfected by Russell and others, so that Hansen's theorem could (and for rigour's sake therefore should) be given entirely in notation. I will discuss this putative "other text", and the idea of re-expression, at length in §78ff below. For now it suffices to point out that Hansen's paper is an example of rigour as presently conceived by professional mathematicians, and that in it, wffs are not used as statements - rather, statements are made about them - literally! They all have the form of W12: "Using the fact that N it follows that M". The W-signs make the assertions using the N-signs. They are not "about" the N-signs in the sense that the subject of the paper is signs. That is not true and a formalism which implies that it is, must be wrong. The paper is about rings. But these W signs are "about" the N-signs in the sense that they physically frame them. Indeed that is exactly what the relation is, **framing**. The W-signs are used to put forward the facts *exemplified, but not stated* by the N-signs. The W-signs do the stating. It is a bit like indirect speech - except that N-signs are never spoken. Each wff and term is actually embedded in a W-sign. This is strikingly shown by the highlighting of the exhibit. In the main deductions of the paper Written word frames the Notation. Thus the relation of framing recurs at several levels in the text. Parts of the Written word component frame the actual inferences in the paper, for example W1 and W2 and W41. Indeed it is also true that sections 1 and 3 of the paper frame its main section 2. The grounding which is achieved by other paragraphy in the paper (such as P1-9) is the subject of the next chapter, but clearly the framing relation which the the paragraphy has to the rest

of the text, while quite apparent as highlighted in the exhibit, is not limited to one simple instance but pervades the whole paper at several levels. What is known by the knowing reader of this paper is not something expressed in any N-sign, it is expressed in N-signs grounded in W-signs themselves grounded in P-signs. The negative part of this is no more than common sense really - knowledge cannot be expressed by uninterpreted signs of the form 'p-->p'. I am pointing out how the interpretation is actually part of the text.

My third criticism of the logico-formalist account of the text is the mistaken idea that N-signs are statements. The illusion that uninterpreted N-signs codify knowledge is the idea that a mixed sign like 'We see that $d(x) \leq 0$ ' is a sentence in which another sentence is used. But in fact the N-sign here is not a sentence at all but a diagram and an analysis like this is much more accurate: 'We see that: $d(x) \leq 0$ '. The whole idea that ' $d(x) \leq 0$ ' is a sentence, a complete thought, is nonsense. What sort of a thought is the axiom 'p->p' of PM? The N-signs of class C2 are **not** sentences. We can read the signs in the class C2 as sentences only as long as we read out the signs like x and y without flinching. They seem alright as long as we read through them to their presumed referents: ' $d(x_1) = 0$ ' tells us that the element referred to by ' $d(x_1)$ ' is the same as the zero element referred to by '0'. This is the Fregean interpretation, if we add that "of course", ' $d(x_1)$ ' and '0' have different senses. Well, they do too. But do they have referents? Putting '0' aside for now, to what does ' $d(x_1)$ ' refer? To find out we must see what ' x_1 ' refers to. It is first mentioned in W27 and we find that it is actually an "arbitrary element" of S for which $d(x_1) \neq 0$. It does not have a referent at all! Nor do any of the other individual variables. In the LF orthodoxy this is disguised by talking about not this piece of mathematics but another which is "equivalent" to it, in which there is no talk of "letting" but in which the "implicit" quantifiers in the text are made explicit, and we see that all these equations are really general truths about elements of the ring in question. Once again we find that the logico-formalist account does not apply to the real text but invokes another text to which it could "in principle" be transformed.

I have pointed out some ways in which the logico-formalist account oversimplifies the complexity of the text. In the next section, I will consider the response that N-signs only abbreviate W-signs, so there is no difference of principle between knowledge expressed in a mixture of signs and the same knowledge expressed purely in W. In the following sections, I turn to the complementary idea that the W signs are eliminable, so that an equivalent pure N-sign expression is possible. Neither of these responses addresses the first of my points, the variety of text acts to be found in such a text. This will be discussed in subsequent chapters.

§74 Notation abbreviates

I listed in Chapter 2 a number of effects of notation which have been observed in the literature. In this and the next three sections I will discuss how these are related to one another and to the written nature of notation. I begin with the fact that many notations are abbreviations. Histories of mathematics make this clear, about signs for unknowns and arithmetic operations in particular⁶. In logico-formalist orthodoxy, all definitions are abbreviations. There is a strand of thought in logicism which might be called the *mother tongue fallacy*, the idea that the language of mathematics is "really" one's mother tongue, and mathematics could really be expressed in ordinary *spoken* English by expanding all these abbreviations. I have argued in Chapter 5 that it would have to be *written* English. But even making this adjustment, there are several reasons why this is still a fallacy.

First, not all notation is introduced in definitions; a great deal of it is defined "in use". The setting-up of a formal system is a kind of definition perhaps (Frege and Hilbert argued about that), but it's not one that can be unpacked (that's why). It is the historically earlier notations which are most purely abbreviations. Later ones also involve more and more various other functions which I shall come

⁶ Cajori, Menninger, op cit.

to in the next three sections. But notation is generally abbreviatory, and it is worth considering in detail the idea of dispensing with it "in principle" and doing mathematics in plain written English. Let us try with the beginning of Hansen's main proof.

Suppose an element of the ring is not mapped to zero by the derivation. From the preceding lemma, the derivation takes the square of this element to zero since its square is greater than zero. Appealing again to strong regularity there exists another element of the ring such that the first one squared multiplied on the right by the second is the same as the first. This (the derivation of the first) is the same as the derivation of the first squared multiplied on the left by the second plus the square of the first multiplied on the right by the derivation of the second, which is the square of the first multiplied by the derivation of the second. Multiplying the preceding equation by the second.....

I'm getting lost already; mathematics expressed like this is hard to follow because its prolixity makes the cross-references to different elements impossible to keep straight in the mind. So the first idea, that notation makes things easier by abbreviation, is certainly true. If the abbreviations are removed it is harder to follow the text. Still, it only makes it harder, it doesn't seem to be necessary, merely desirable. "In principle" we (at least, some of us) could do mathematics using only words. The abbreviatory point deserves more serious consideration, however. It may be possible to globally "English" mathematics, but if it were done, would the result be understandable? I suggest not. The reason is that it would become so enormously long and so circuitously expressed that the difficulty we have adumbrated for a single text would become a sheer impossibility for the whole discourse - none of us could learn significant amounts of a purely English mathematics. Just try solving a simple quadratic problem in words, or squaring a seven digit number (expressed in words of course). For any finite creature there are cognitive tasks which though finite are impossible to complete in its lifetime, or in time for the solution to be useful. No-one can count the raindrops in a shower - but it can be done "in principle". This appeal to principle is not acceptable however; **if mathematics could be done in principle but not in practice then it would not exist.** It is *existing* mathematics which poses the problems of philosophy of mathematics. Calculations cannot become part of our social institutions if they cannot be

completed in time for the results to be used in those social institutions. If counting the citizens of a town to tax them took so long that they would be greatly changed in numbers through death or natural increase before their dues could be determined, then their dues simply could not be determined in this way. If solving inheritance problems in words was so difficult that most of them were insoluble reliably within the lifetime of the legatees, then they would be solved another way (violence). Mathematics as a social reality is only important because it is conducted in ways which do produce results when they are needed. It could never become an institution of cognitive high status if proofs for even simple theorems were always so long and difficult to follow that agreement on them remained less well-established than in theology. Mathematics would not be a paradigm of knowledge; the problems of philosophy of mathematics, which derive from its being one, would not arise.

But in any case, abbreviation is not the *only* effect of notation and the corresponding improvement in comprehensibility of mathematical text not its only consequence. We must guard against the tendency to think of '+' as *simply* a brief alternative to the word 'plus' or even 'and', as in 'two and two is four'. The latter sign 'and' has an ambiguous status, because the definitely notational sign ' $2+2=4$ ' has specific properties which 'two and two is four' lacks; but it is often assumed that "in principle" they are interchangeable. This is not so. The virtue of the logico-formalist analysis of notation is that it enables us to make the difference between these signs clear. Unfortunately it also makes the relationship between them obscure! The cardinal point is that the notational sign obeys explicit rules we lay down deliberately; the difficulty of the logico-formalist account is the further claim that these are the only rules to which such signs are subject. This makes their applicability mysterious and inexplicable. In fact, notational signs are always embedded in text which includes signs from the three other signsystems, and these ensure that such signs are actually under the sway of rules which ground them in possible application. To understand my English version of the text, we must

understand terms like 'element', 'ring' etc. Now our actual knowledge of these has been gained using notation. So a fair test of notation's necessity would also assume that any prior texts had also been entirely Englished. To create globally English mathematics would require *all* notational devices to be removed, not merely hidden by the use of words which replace them. To use 'ring' for something which was originally defined using notation *after the definition is made* is one thing; to create the concept without any notation at all is quite another. In my Englishing of Hansen above, I have replaced variables by verbal forms which can play the same role - but can they play that role for someone who has no experience with variables? Can such a reader follow their uses? Frege tried to mimic his notational development of the Begriffsschrift in his *Foundations of arithmetic*; despite the clarity of his writing, he admitted himself that for precision of statement and rigour of proof the notation would be needed. This is because a calculus of proof requires tokens whose transformations are governed entirely by known rules: that is, explicitly introduced signs. Of course we could use nonsense words which were not already governed by prior linguistic rules, like this:

A1. Orez is a number.

A2. If rebmun is number, so is txenrebmun.

A3... and so on.

But this device is tantamount to the usual notation in an inconveniently unabbreviated form.

It does not follow then, that because we can mimic mathematics in W-signs, that all mathematical text could be simply W-signs; what it shows is that semiotic effects created in one code can be transferred to some extent to others - but only with difficulty, and in a context of reliance on the original media and discourses as sources for the mental structures necessary for the transferred significations. In this section I have pointed out that the abbreviatory effect of notation creates possibilities not available in Written word alone. I go on on the next section to consider further the significance of the explicit rules we have for notation.

§75 Notation is formal

If we make an entirely verbal counterpart for the text we find that it is very hard to follow the inferences, unless we can treat whole phrases *emptily*. A phrase like *'the number you first thought of'* must be sugarred⁷ so that it can be manipulated in phrases like *'the square of the number you first thought of, 'seven times the number you thought of plus twice the square of the number you thought of all divided by two'*, and so on. That is, we must ignore or bracket the actual meaning of such a phrase while going through inferences involving it, in order to attend to their validity while prescind from their truth or import. It is necessary to do this to gain certainty of inference. The difficulty of doing it is one motivation for introducing notation which has no prior interpretation but whose manipulations are subject to explicit rules. Insofar as notation is chosen with this end in view it becomes less abbreviatory. One wants an abbreviation to remind one of what is abbreviated (this is a criterion for good notation) but one wants inferential counters to remind one of as little as possible (which is why the unknown is x , whatever it may be).

The main deficiency in the logico-formalist account is the assumption that all the semiotic roles in mathematical text can be identified in speech. These are thought to be reducible to reference and predication, so all mathematical signs have to be classified accordingly. But really, mathematical signs bring with them new kinds of semiosis. A variable like ' x ' is not a variable name, nor a name for a variable object⁸, nor a temporary name, all ways of speaking scorned by the early logicians. But nor is it a kind of pronoun, the image favoured by logico-formalism.

⁷ For me, the word 'sugar' epitomises the phenomenology here; when I was young I used to think or say this word until it lost all meaning and became a mere soundshape whose semantic connection with any real item seemed unlikely in the extreme. This emptying of the signifier is practiced systematically and routinely in mathematics, originally via variables. It is the root idea in Husserl's [1936] and is the real formalist trope - but must be thought of psychologically. That is the mistake of formalism, to anathematise the mind!

⁸ Fine has recently revived this old idea, from an entirely logico-formalist perspective. I shall not pursue the matter here.

It is like a pronoun in allowing anaphora. But like the general phrase from which it really derives ("a number less than 6") its reference is indefinite, that is it does not really refer at all. It is not a pronoun because there is no noun for which it is pro⁹. Indeed its emptiness is its real novelty. Unlike the general phrase it is apt for constructive manipulation because it can be treated *formally*.

This is the aspect of notation on which formalism has seized, this role of the "empty" sign. Husserl¹⁰ discusses it in his *Origin of Geometry* under the name "unfulfilled intentions" and it would seem that Hilbert must have been influenced by Husserl in his thinking about meta-mathematics. The idea of an empty sign is that it can be "manipulated" (how these metaphors insist!) without regard for the acceptability of the signs into which it is transformed. For example, consider the use of a letter to stand for the unknown quantity in an algebra problem, stated verbally. There are two difficulties in trying to solve such a problem verbally, the length of the sentences makes it difficult to attend to all the features of the situation at once, and their content makes it hard to reorganise them with confidence that truth is being preserved. Letters solve both aspects of the cognitive problem: they enable short statements to be used for the data, the abbreviatory function; and they allow a relative disengagement from their "referents" during algebraic manipulation, the emptying function.

The formality which arises from emptying is created by the need for explicit rules for the manipulation of N-signs. One knows about notation that one must know the rules of use before one can hope to use it, in a way which does not apply to most words: you can use words how you like, as long as you have some notion of how others use them then your deviations may be creative. The question of the "correct rules" is largely irrelevant, though some social groups think otherwise. Any mathematical sign, no matter how familiar, can be explicitly defined in a

⁹ People first learning algebra commonly think that the teacher knows ("always already") what x stands for, and is just being difficult. Later, some of them acquire the appropriate instrumental attitude to variables. They are the ones we say are learning mathematics.

¹⁰ Husserl [1936]

given local context and if used there according to those rules, no mathematician will object much, unless it is implied that the new definition should supersede some established one. For example '+' is a very commonly used sign for wide range of different operations. This rarely leads to any actual ambiguity however, because the explicitness of the N system prevents it. Readers can get confused - but the means for sorting it out are in the text. With words on the other hand, because their usual use is established and communicated informally and inexplicitly and vaguely, no such freedom is available.

The explicit rules of combination of N-signs is what constitutes a calculus, and thereby leads to ease of calculation and similar "facilitatory" aspects of notation. For example, the usual algorithm for decimal arithmetic addition is possible because the decimal notation has clearcut and explicit rules for the formation of numerals and for the operation of adding a pair of digits. The formality of much notation, that is, its deliberate lack of a specific and fixed interpretation, and its control by explicit rules of use, allows the features of locality and polysemy to be used positively.

So notation is not only abbreviatory but formal; these are distinct features and provide complementary aspects of notation's facilitatory role. In the next section I consider a third distinct feature, its iconicity.

§76 Notation is iconic

Notation organised in a formal system is iconic: the mode of representation of notation, as system, is iconic in the strict sense which logico-formalism has explicated and which I have transferred to diagrams. Of course there are plenty of notations, for example names of specific numbers like π , to which this concept is quite irrelevant. But the formal system, as perfected in half-century of logico-formalist effort, is purely iconic, and insofar as mathematics uses systems of notation which the logico-formalist account describes, so far mathematical text is iconic. The elements of the sign correspond to the elements of the model. The

relations between them mirror those in the model. The "truth - conditions" of the "sentence" which the sign mimics are recursively built up from partial truths about the model. Not all notation in mathematics is constructed in iconic fashion (for example, the signs for the circular functions), but it is in virtue of its iconicity that a notation is useful for computations or proofs, so that all signs take part in iconic relations.

The source of the iconicity of the N system is **not language**. Rather, it is the physical counters used for computations prior to writing, which on Schmandt-Besserat's theory actually played a role in writing's origins. N-signs function as counters in a calculus because they are replacements in the graphic sign-system of mathematics for three-dimensional ones. This alone is enough reason to give weight to the role of nonlinguistic sign-systems such as Diagram in the mathematical sign-system. For as I shall argue in the latter part of this chapter, the illusion that notation derives entirely from language is a potent mistake. It underlies the common but false idea that a notation can be a language.

That an icon needs a ground is inherent in the way I drew the contrast between them in Chapter 6, so that if it is accepted that notation is iconic, then it follows that notation must be grounded. I have already pointed out in §73 that notation in text is grounded wff by wff: words frame notation. In mathematical text, demonstratives are literal - "we see that: $d(x) \geq 0$ " is the form of the relation of W and N. Words frame notation in mathematical text in the same sense as grounding signs frame the icon of a diagram: they surround it spatially, and they direct the readers apprehension as to what is being represented. The *general* mode of presentation in mathematical text is the form 'W: N' as in 'we see that: $d(x) \geq 0$ '. Donald Davidson once suggested¹¹ such an analysis of opaque contexts in general, but with nothing more than aspects of the grammar of 'that' to support his suggestion. In the case of the relation of N and W there is considerable effort in most mathematical text to make the contrast between the W and the N-signs *obvious*, by

¹¹ Davidson [1969]

using different fonts, spacing and so on. The formality of the N-signs prevents their being taken for strange words.

I pointed out in §73 also that there is a good deal of Paragrophy in our sample text which also contributes to grounding it, though most of it acts to ground the whole main text, composed as it is of W- and N- signs. I will discuss such global grounding functions of Paragrophy in the next chapter. However the grounding of individual wffs by sentences and the global grounding of the main text do not exhaust the ways in which N-signs are grounded in mathematical text. Returning to my English version of the Hansen main theorem, notice that the author explicitly makes a semantic ascent and discusses his signs ('Multiplying the preceding equation...'). This is not an aberration, but entirely typical of mathematics. This text, like most others, introduces part of its notation at the beginning, thereby grounding the subsequent use of the N-signs in question. This "metatext" is not be part of the mathematics according to the logico-formalist account, since it must precede the mathematics, which is what is done *using* the notation. Yet mathematics as it is actually done is rarely found without it! (This is the difficulty about definition which both Russell and Frege noticed, but could not account for; I will come back to it in Chapter 12.) But this example in the main theorem illustrates how the grounding of the N-signs goes on throughout the main text, not just at the start.

In the next section I consider further the grounding of N-signs, and argue that it *must* involve W-signs.

§77 Notation must be grounded in Words

The idea that a notation could be a language is quite misguided, for insofar as it is iconic notation needs to be grounded in language just as do diagrams. Consequently the *replacement thesis* about notation, that is could take over all the functions of language, is also mistaken. I have already argued that being iconic, Notation needs grounding. An icon needs a frame. Mathematics actually proceeds

through sentences of words framing notation. Now I shall argue that this grounding of the iconicity of Notation *requires* words. First, I will argue that N-signs must be given interpretations through words. Second, I will argue that texts employing notation need to be launched: a notation cannot be initiated as both iconic and communicative.

The first reason why grounding needs to be in words is simply that a lot of it already is grounded in words, because the activities in which it is used are activities which are carried on within the great stream of words which composes human culture. Any new notation or new uses of old notation have to be tied into the ongoing accumulation of mathematised action, and words are what can make the connections. Even on the logico-formalist account, theses of formal systems express truths when their interpretations are true. It is these interpretations which constitute the application of mathematics, and it is only because they exist that the question of truth arises at all. The N-sign ' $d(x) = 0$ ' cannot be used to make any assertion, expresses no knowledge unless we know that d stands for a derivation in a ring, etc. This need is not avoided but merely shifted by providing a heavy formalisation of ring theory in which 'derivation' is a defined term in set theory. This wff only says something about derivations in rings to someone who has the concepts of derivation and ring. This requires familiarity with examples of both. The words are used to refer not merely to the formal definition but also to the examples. Logico-formalism sometimes tries to deny this, but one finds that even in the most formal presentations of mathematics there are notions which are not defined and whose interpretation is given through words. In *Principia Mathematica*, 'proposition' is such a notion. Considering how basic to *Principia Mathematica* this notion is, and the little that is done to actually define it, it would be quite amazing if mathematics *really* reduced to consequences of a few axioms about propositions, like ' $p \rightarrow p$ '. But it doesn't really, because in real mathematics most terms, even those which are defined, are given meaning at least in part ostensively. Even if a Fregean definition of number were formally satisfactory it

would *not* exhaust the concept we have of number. Particular numbers give part of it, numbers like 1,2,3... π , 666, the number of the planets and so on. These concrete instances of the concept are given to us through words. Notation must be grounded in written word. The relation is one of dependence, though this does not in the least imply that N could be reducible to W, an idea which I have already refuted in §74.

Examination of texts shows that notation is always *launched*. Indeed I shall suggest that it must be. By saying that notation must be launched, I mean that the use of notation in a text must be prepared by the prior setting-up of the notation. Now in a developing tradition, often some of this can be and is taken for granted. But rarely all of it, most texts establish some of their notation at the start. The Hansen text is quite typical, in that it defines some of its notation but assumes that the reader will be familiar with the great majority of the signs to be used, and with the systems of notation to which they belong. Here it is necessary to bear in mind the point I made above in Chapter 4, about local and global eliminability of signs. Certainly, any individual text can trade on the reader's prior experience and assume that the notation which will be used in it is already understood. But that is to assume the existence and use of other texts in which it is established. A minor point which is interesting in this connection is the extremely *long* launches to be found in some of the canonical texts of the logico-formalist tradition - *Principia Mathematica* has 80 pages explaining its symbolism, Church [1956] and Curry [1963] similar expanses.

A radicalisation of Quine's radical translation problem can provide us with a reason for thinking that launches are a *necessary* basis for notation. The situation in which that problem is supposed to arise is both vague and overspecific¹². It is vague in that the circumstances in which the problematic text is presented are merely gestured at, yet specific in that we are clearly allowed to take its producers to be creatures superficially like ourselves, perhaps even human, to the extent that we can identify sentences and the "native speaker's" assent and

¹² Cf Quine [1962: ch2]

dissent to them. These are large but highly unspecified concessions to the interpreter. In the CETI problem, by contrast, of the possibility of communication with extra-terrestrial intelligences, we simply have a trace from radio or other receiving device, which may or may not be a text from extra-terrestrial signallers. It is usual to assume that, if it is a text at all, it will be a message, that is, will be an attempted information transmission. Rather a large assumption if one considers a random sample of the texts transmitted over human electronic media. (This is an example of the identification of the "serious" and "literal" with the "real" or "essential" texts, which Derrida rightly contests as I mentioned in Chapter 4). The situation has been dramatised by various science fiction tales, such as Hoyle's *A for Andromeda* and Niven's and Pournelle's *The Mote in God's Eye*. The more active program, of trying to send a message to such presumed aliens, has been taken seriously enough to attract considerable funding and scholarly activity ¹³.

The question for us is: how can you make the initial bridgehead of communication with no pre-existing shared language at all? One extended attempt to work out a way to do this is LINCOS¹⁴. In constructions such as LINCOS it is generally assumed that the bridge which must be established with the aliens will be made of mathematics, on the grounds that the laws of the universe are by definition the same everywhere, they have essentially mathematical expression, and intelligences capable of detecting our transmissions will have an understanding of the universe which presumes, like ours, considerable mathematics. A very interesting and debatable argument, but let us not be distracted by it at present. The strategy adopted in LINCOS is in essence to make the first texts **samples** of numerals and other basic signs of a formal system, and of atomic sentences in it. This is not surprising as a product of a mathematician in 1960, although perhaps for a quasi-intuitionist it is not quite unremarkable.

¹³ See Bracewell [1974]

¹⁴ Freudenthal [1960]

Now I don't want to deny that this might work - although the basic argument for thinking that it might is that if we got such a message, we would be able to decode it. Rather, I wish to consider what is assumed in the idea that this method could possibly work, no matter what the material circumstances of the receivers. Because that is the assumption which is made by the the philosophy of mathematics which allows that notation might be a language itself. The decoding problem is at bottom a statistical one: to decide that we have a text at all, we must decide that there is a pattern in the data, that is, that there are repeated elements in it which we can take as the "alphabet" of the incoming language. This is not sufficient, however, for we are convinced of the repetition of atomic structures in samples of lead, without taking this as evidence of attempts at communication. What more must we find in the data? I suggest that we would have to find evidence of *arbitrary* and *hierarchical* structuring of these presumed semiotic atoms. By 'arbitrary', I mean that the larger units into which they are assembled in the presumed text must not be explicable by the laws of physics; by 'hierarchical' I mean that we would expect to find that there are discourse units at a range of extents: as we have letters, words, phrases, sentences, texts, traditions, languages so we would be looking for identifiable levels of construction in the text, because we must presume that the senders are finite and forced like us to use recursive techniques for text variation. The point I am trying to make through this rather speculative discussion is that in making such a decision as that a text has been received, and trying to decode it, we would in fact be driven to make a lot of assumptions about the senders.

Now this does not yet make my point about notation, because these assumptions on our part, even if true, do not constitute any kind of shared pre-message language. My point about notation depends on this: that having decoded a text, we would want to reply. To be sure that it was a text, and to establish communication, we would transmit a text in the same alphabet in the relevant direction. What would we say? I think we would quote the incoming text, and add some kind of simple transforms of it to evidence our grasp of its principles of

construction. But we would also "sign" our text. That is, we would try to set up some kind of indexicality in the texts which we hope to see flying back and forth across the galaxy in order to know which ones are meant for us. Until there are texts directed to specific communicators, there is not yet communication, merely some kind of very public archiving of information. So I think that the actual conveyed by the initial messages would be things of the nature of "We are here - where are you?" "We are homosap - who are you?"¹⁵. Looking at it from this point of view we can see that the real content of the first communication was not $2+3 = 5$ and all that "mathematical" stuff, but the fundamental indication of signification "This is a sign". After all, we are *assuming* that they know mathematics, so we can hardly be *telling* them that $2+3=5$. So the idea that mathematics in notation can be the language itself is a gross misconception of the role that mathematical notation plays in such a contact: it is actually the vehicle for a quite simple and direct message, and is not being deployed as notation in order to convey or even state mathematics at all! Signs that we normally use as N-signs are actually being used as W-signs.

In this section I have argued that N-signs must be grounded in W-signs. In the next two sections I will discuss a common misinterpretation of Frege's Begriffsschrift as a perfectly logical language, an idea which ignores the need for grounding. In the three after that, I argue that a scheme of Field's for avoiding "ontological commitment" to numbers depends on this same mistaken idea that N-signs can by themselves be a "language" (for science at least); it also depends for motivation on the idea which I will refute, that re-expressions in canonical notation show ontological commitments. These are powerful and widely accepted errors, based on misunderstandings of notation. They show the logico-formalist hegemony at work.

¹⁵ The message actually put on Voyager was clearly of this kind.

§78 Begriffsschrift as perfect logical language

Motivated by a different effect of notation, its lack of ambiguity, vagueness and uncertainty of inference, many people interpret Frege's *Begriffsschrift*¹⁶ as an attempt to create a perfect logical language. Diamond has recently explicitly interpreted Frege this way. Most other writers on Frege adopt a similar view, though Sluga and Currie show some caution. But none of them show sufficient regard for Frege's own statements about *Begriffsschrift*, which contain several arguments for what is an extension rather than replacement conception of notation. In this section I will expound these views, in the next argue that they are untrue to Frege and misinterpret what notation is.

There is a long tradition of longing for "logically perfect language" in which various puzzles would disappear. Motives vary from the illusion that calculation is the solution to all disagreements (Leibniz, AI today), to the desirability of a language everyone could speak. An enduring motive is the idea that ordinary languages are vague, illogical etc. Frege's achievements in his *Begriffsschrift* are generally recognised to include the inventions of the modern concept of the quantifier, or quantification theory, the modern concept of formal system, distinctions between variables and constants, or between kinds of variables, between axioms and rules etc. Bynum¹⁷ lists eleven such achievements, none of them inconsiderable, but those I have listed are the ones most relevant to my present argument. Most modern writers would essentially agree with Bynum's list, but many of them also suggest that these are local achievements to be set against Frege's failure in his global aim of creating a perfect logical language to substitute for ordinary language.

Diamond has recently made explicit¹⁸ an interpretation of Frege's *Begriffsschrift* which takes it as an attempted perfect logical language. Diamond is

¹⁶ Frege [1979]. I shall italicise his book, but not his system, in the following discussion.

¹⁷ In Bynum [1972]

¹⁸ Diamond [1984]

the most explicit example of the modern interpretation of a 'concept script' as a substitute perfect 'logical' language. Her overall strategy in her essay is to contrast the philosophical consequences which Frege on the one hand and Wittgenstein on the other envisaged flowing from the introduction of such an "adequate notation". For either writer, according to Diamond, a good concept script is a "tool of intellectual liberation" which would allow us to see clearly the "character of thought" in general. She devotes a good deal of space to the impact of a perfect language on semantic puzzles such as the one about concepts and objects - the concept 'horse' imbroglio - and to the contrast between saying and showing. Much of Diamond's long discussion may be summed up as the argument that if concept and object are characterised by different roles-in-a-thought and an adequate notation shows all and only such differences of role, then in it one cannot even state a question presupposing the same roles for concept and object. So a "perfect" language will preclude the kind of muddle we get into with ordinary language. This may or may not be a valuable insight into semantics - I incline to think it is not - but I don't think it is the answer to her question "What does a concept script do?", because it presupposes that the "adequate notation" must substitute for ordinary language rather than supplement it. None of this has much to do with the central aim of the *Begriffsschrift*, which is the perfection of *mathematical* discourse.

Diamond's is a view consistent with the majority of writing on Frege. Bynum agrees that Frege sought and failed to find a perfect logical substitute for ordinary language. He cites Frege's admiration for Leibniz' project to create just such a replacement. Thiel¹⁹ is another writer who takes the *Begriffsschrift* as an attempt to implement Leibniz' project and make an artificial language for the clear expression of all thoughts; his first chapter expounds this idea explicitly. These writers seem less cautious than Sluga²⁰. He, while identifying Frege's intellectual lineage from Leibniz through Kant and Lotze, rightly points out that Frege attempts

¹⁹ Thiel [1968]

²⁰ Sluga [1980: 95]

something much less ambitious than to implement Leibniz' program. It is not even quite true that Frege was seeking a replacement for the existing "language" of mathematics, since part of his explicit purpose was to "supplement" existing notations such as that of arithmetic.

"When we step back from the particular issues that have so far concerned us and ask why Frege was interested in language and the construction of a logical language we discover that his interest was never an end in itself...Thinking about language was essential for any real improvement in logic. But even the interest in logic was not philosophically ultimate for him.." and "Is not such narrowness not due to the fact that he was first and foremost a mathematician?" 21

Currie's avowed aim²² is to present Frege's work as a contribution to epistemology, to the long-standing problems about certainty and objectivity, rather than to late twentieth century controversies about meaning or ontology²³. I agree with his negative thesis here, but would want to sharpen the point: Frege's work is in fact all directed solely to the philosophy of mathematics, not to the problem of knowledge in general, not even scientific knowledge. Frege was no more an epistemologist than he was a philosopher of language. Currie claims²⁴ that Frege was only interested in mathematics as an example of an epistemological ideal. But the rest of Currie's book, which treats Frege's works in detail, refutes this view, since it treats no examples of knowledge other than mathematics. Frege was not interested in other knowledge. Insofar as Currie goes on to consider Frege's views about general issues in epistemology in his last chapter, I believe he distorts the point of the *Begriffsschrift* in presenting it as a tool for such general purposes. For example, the 'objectivity of knowledge' as a *general* problem for epistemology is not the same as the *special* problem of the objectivity of mathematical knowledge which concerned Frege. Actually the general problem only has its life in a cultural context within which mathematics exists as a demonstration of the possibility of objective knowledge. It is by no means obvious that such objective knowledge is possible

21 Sluga [1980: 61]

22 Currie [1982]

23 As Dummett's prolix interpretations of Frege have suggested. Cf Dummett [1973].

24 [Currie 1982:13]

unmediated by mathematics. To assume that this is so, and transfer Frege's tool for a special problem to an alleged general problem, is to put two carts before the horse.

So one must question whether Frege intended his Begriffsschrift as a perfect logical language. If we examine what he actually says we find that he thought of it in terms much more like those I have attributed to notation earlier in this chapter.

§79 Begriffsschrift as notation

There are three main aspects of the correct view of the Begriffsschrift. First, Frege puts the Begriffsschrift forward as an extension of language, explaining this relation through analogies with microscope and "artificial hands". Secondly, he explicitly puts forward as significant a number of its graphic features: Begriffsschrift is a *written extension* of ordinary language. Thirdly, in the way that it is modelled on algebra, as Frege says, Begriffsschrift is notational: it uses formality, iconicity and abbreviation. Reasons for the general neglect of Frege's actual Begriffsschrift in favour of Russell/Peano notation reinforce these considerations. I will expand on each of these claims.

There seem to be two main elements in Frege's declared aims for the Begriffsschrift. His first aim is to remove the imprecision of ordinary language; and this is so that he may pursue his second, main aim, which is "**to keep the chain of reasoning (in the foundations of arithmetic) free of gaps**". It is **not** intended as "*a presentation of abstract logic*" which, as for example in Boole's symbolism, is "*not suited for the rendering of a content...But this is exactly my intention*"²⁵

Frege intended the Begriffsschrift as an extension of ordinary language. He explains his idea of the relation of the Begriffsschrift to ordinary language (Sprache des Lebens) using two analogies: the microscope, and the idea of 'artificial hands'. In each case the point is that just as other tools extend our ordinary capacities through an increased but inflexible precision for specific purposes, so too does his own technological innovation. His notation extends our capacities, in this case our

²⁵ Frege [1972: 104, 97, 93]

linguistic capacities, through an increased but inflexible precision, and the tool is therefore *useless for other purposes*. He specifically refers to his innovation as technological. He says²⁶

I believe I can make the relation of my "conceptual notation" to ordinary language clearest if I compare it to the relation of the microscope to the eye. The latter, because of the range of its applicability and because of the ease with which it can adapt itself to the most varied circumstances, has a great superiority over the microscope. Of course, viewed as an optical instrument it reveals many imperfections, which usually remain unnoticed only because of its intimate connection with mental life. But as soon as scientific purposes place strong requirements upon the sharpness of resolution, the eye proves to be inadequate. On the other hand, the microscope is perfectly suited for just such purposes; but for this very reason, it is useless for all others."

This is entirely incompatible with describing Begriffsschrift as a language proper²⁷, which Bynum does, rather vaguely, but which Diamond does quite deliberately as a necessary feature of her argument. It is a fundamental error to think of formal languages as capable of independent use, as alternatives to ordinary language rather than as essentially written extensions of it. Frege gives no arguments for regarding formal "languages" as languages proper, or for the general applicability of the Begriffsschrift. On the contrary his stated views seem more consistent with the view which I am putting forward. He offers²⁸ another simile indicating his view of notation as extension in his *On the scientific justification of a conceptual notation*:

"...[ordinary] language can be compared with the hand, which despite its adaptability to the most diverse tasks is still inadequate. We build for ourselves artificial hands, tools for particular purposes, which work with more accuracy than the hand can provide. And how is this accuracy possible? Through the very stiffness and inflexibility of parts the lack of which makes the hand so dextrous."

The idea common to these two similes is that the Begriffsschrift supplements ordinary language by giving up some of its virtues to improve on some of its inadequacies for specific purposes.

²⁶ Frege [1972: 104-5]

²⁷ Recall that I mean by language a signsystem adequate or the general communicational needs of people. Many logico-formalists vacillate between meaning this and just any kind of signsystem.

²⁸ Frege [1882: 86]

In the second place, Frege was very conscious that the Begriffsschrift was designed to exploit graphic possibilities. Frege has powerful arguments³⁰ for the potential advantages of visible over audible symbols which seem to have been ignored by his commentators. Some advantages he cites for visible symbols are

[1] they are "*sharply defined and clearly distinguished*"

[2] have "*permanence and immutability*"

[3] give "*the possibility of keeping many things in mind at the same time*"

[4] and "*symbols on a two-dimensional writing surface can be employed in far more diverse ways to express inner relationships than the mere following and preceding in one-dimensional time*" This latter advantage is not exploited by the written word, Frege thinks, but can be in conceptual notation such as that of arithmetic formulas "*since it directly expresses the facts without the intervention of speech*". Thus he is attempting to "*supplement the formula language of arithmetic*". For Frege, Begriffsschrift is a graphic extension of speech.

Further reasons for denying that Frege conceived Begriffsschrift as capable of replacing ordinary language can be found in his remarks on the generality of its applicability. Frege did not suggest or apparently believe that the Begriffsschrift would be generally applicable. It is important to note that Frege, in conformity with his declared purposes, only *applies* the Begriffsschrift to the analysis of mathematical and purely logical reasonings, though some of his *illustrations* concern scientific reasoning. He does not attempt at all to use the Begriffsschrift to deal with philosophical puzzles stated in ordinary language, puzzles such as that concerning the semantics of 'the concept horse'. This puzzle is in fact created by his *later* distinction between concept and object, which Diamond attributes to his purposes for the Begriffsschrift, but it seems to me that this is anachronistic. If he uses the Begriffsschrift as a "tool of intellectual liberation" as Diamond asserts, it is only with regard to certain particular enslavements concerning mathematics. This kind of philosophical generality is, I believe, entirely alien to Frege's much more

³⁰ The quotations in the rest of this paragraph are from Frege [1884: 85-6]

modest aims in the Begriffsschrift. He says merely: *"I am sure that my 'conceptual notation' can be successfully applied wherever a special value must be placed upon the validity of proofs"*. In other words, precisely *not* generally; rather, only in mathematics and perhaps the more mathematical parts of science. Yet this is on the very page which Diamond cites as evidence that Frege *"allows us to think about language, any language in which thoughts can be expressed.."* Frege was quite modest in the claims he made for the applicability of the Begriffsschrift; but in any case, even if he thought, or came later to think, that a Begriffsschrift could be developed that could replace ordinary language, this possibility cannot be his great actual achievement of 1879. The achievement is the technological innovation *as he described it himself and as he actually deployed it*. What is shown there, though it was unappreciated for 30 years, is that the new notation can make better proofs of arithmetical theses. The proofs are better *for Frege's purposes* because they are more perspicuous and therefore more certain. We can "see" and "grasp" the basic laws more clearly, more securely - recall Frege's two chosen metaphors for the Begriffsschrift, the microscope and the artificial hand.

The two-dimensional aspect of Frege's system of symbols is in fact an essential part of his contribution, not a quirk which can be safely neglected by a wiser posterity. Frege's notation was, and largely still is, seen as queer and cumbrous, hard to read and hard to print. There are several reasons for this. He elects to make do with only a single rule of inference, which makes proofs longer than they might otherwise be. And he deliberately created a novel two-dimensional notation which makes proofs both very extended spatially, and organized quite differently to simple written speech. Frege chose cumbrousness in the interest of *perspicuity* which he rightly held was necessary for certainty. Frege clearly underestimated the reluctance of even the best scholars to adopt new methods, and most commentators, contemporary or modern, simply dismiss Frege's system as too new and too different. They rarely spend much time on the real advantages it offers. Only a few scattered asides are to be found: Sluga admits that "when one gets

used to the notation, one discovers in it a clarity which contemporary notations lack" but counters this with the point that it is not very useful for modern meta-theoretical studies. But that was not Frege's intended application! (Recall his emphasis on the specific uses for which tools like Begriffsschrift are created.) Thiel admits the considerable advantages of Frege's system, and that its bypassing in favour of the Peano-Russell system was certainly due to a combination of scholarly inertia and the typesetting problem which it posed, rather than any symbolizing inferiority.

The advantages which Frege correctly saw as flowing from a two-dimensional notation are essentially two. First, the possibility of using the two dimensions to mark significant features of two different kinds. Second, the possibility of increasing the assimilable information on a page. Two particular advantages are gained in Frege's system. First, the separation of the modes of representation of the form and the content of formulas, which in particular permits the clear exhibition of the reference and generality in them (this is the contrast between saying and showing). And second, an extreme simplicity in the representation of the deductive relations between formulas, and therefore security in the apprehension of them. As a matter of fact, various more recent systems of logic *do* attempt to gain cognitive advantages by exploiting both vertical and horizontal articulation of the signs used for formulas. Subscripts and superscripts, including exponents, are of course time-hallowed devices which make some use of the vertical dimension in the articulation of basically horizontally linear symbols. They are essentially vertically displaced from the line in which the plaintext is organised. An indication of this is the difficulty of unambiguously but perspicuously reading out loud notation using subscripts and superscripts, and the even greater difficulty of grasping the sense of the linearizations of them one finds now in computing contexts where the subscripts and superscripts are replaced by array parameters. Natural deduction systems making use of tree structures, and line labelling systems for checking the validity of arguments, are both compromises

between the cognitive superiority of Frege's method and the capabilities of mid twentieth century printing technology.

There is one good reason for replacing Frege's system with the linearized notation of Peano and Russell: one thereby obtains a conspectus of more of a proof at a time³¹. There is no doubt that Frege's system is a great consumer of paper, his arguments do extend vertically rather a lot! Yet to point this out is simply to repeat one of Frege's own arguments for visible display, since one of its stated advantages in exploiting the two-dimensionality of the writing surface is "keeping many things in mind at once". The Russell-Peano system allows us to keep *more* formulas in mind at once by allowing more of them to be present together in the visual field. The price is the lessened perspicuity of each formula in that its internal structure is harder to keep clearly in view. Thus the apprehension of proofs is an *economic* problem, in that more detail in view costs greater effort in assimilation, *ceteris paribus*. Notational innovation might be seen as the attempt to change one of these holding the other fixed, under pressure for a greater cognitive return on effort expended.

Frege is quite explicit that he has modelled his Begriffsschrift on algebraic notation, and it therefore shares with it the features of formality (he says³² that the main point of overlap is the use of letters), iconicity (it expresses things directly without the intervention of speech) and abbreviation (obviously). Thus the Begriffsschrift was intended as a notation, that is as an abbreviatory, formal, iconic written **extension** of speech, and **not** as a logically perfect replacement for it.

§80 Analysis, replacement and avoiding ontological commitments

The replacement conception of the role of notation also derives from an alleged effect of logical notation which I do not accept, that of revealing logical form. Russell's analysis of non-denoting expressions takes notation to re-express

³¹ Cf exhibit #7B.

³² Frege [1879: 104]

perspicuously what one really means in ordinary language³³. Quine's regimentation of ordinary language for science is explicit about replacing ordinary forms of expression with better ones, justified by the alleged ontological clarity of canonical notation³⁴. Field's recently proposed program to avoid the problems about mathematical objects and truth presupposes this train of ideas³⁵. But it also requires a global replacement of ordinary language by notation. In this section I will sketch Field's program and its reliance on this tradition, in the following two I will criticise it for the two presuppositions on which it relies, those of the global replaceability of language by notation and that logical re-expression reveals logical form.

There is a tradition in philosophy in this century of advocating that certain expressions, be they names or definite descriptions or whole sentences, or even whole theories, would in some way be improved if they were reexpressed in a different notation. What is achieved by the suggested replacement may be that confusion is averted, or that the correct semantics of the original are exposed, or that its ontological commitments are clarified, or only that the substitute permits progress with whatever the user of the original expression was about. How the offending expressions need improvement is not always precisely stated, either. For example, the phrase "for Christ's sake" is held to encourage belief in sakes and thereby confusion - if you start asking questions about sakes based on what seem reasonable questions about, say, stakes. Or "The present king of France is bald" is held to be misleading, suggesting that to account for its falsity we must posit reference to him or her and thereby apparently in some way recognise her or his existence. Notice that in each of these cases there is a theoretical context supporting the confusion: in the first, we think we know the kinds of criteria for identity and identification that apply to stakes, and presume that we can use them on sakes. In the second, we have a theory, again more or less explicit, about how to account for

³³ Russell [1905]

³⁴ Quine [1960]

³⁵ Field [1980]

the truth of a sentence, in terms of reference and we apply it to the questionable sentence. The paradigm example of using notation to avoid such muddles is Russell's theory of descriptions. He re-expressd

[A] The present king of France is bald.

as

[B] There is a unique thing which is a present king of France and bald.

This sentence is a *simply* false claim that something exists. To any residual worry about "speaking of nothing", or any protest that [B] does not say exactly what was meant by [A], it is suggested that ordinary language is vague and that what is really meant by both of them is best shown by

[C] $(\exists!x)(PKFx \& Bx)$

which reveals the logical form of [A]. Notation displays logical form.

Quine systematised this line of thought. First, philosophical puzzles are deemed to arise from talk about mysterious objects. Some expressions enter discourse in such a way that theories using them create difficulties because those expressions involve the theory in assuming to exist things whose properties seem mysterious or impossible, or harder to explain than the phenomena which the theory is intended to explain. Their use in a theory ensures that the theory will be explaining the obscure through the more obscure, clearly a poor strategy. *Ontological commitment* of theory to Xs means that for it to be true, Xs must exist. Now in many cases it is difficult to establish exactly what are the ontological commitments of a theory, that is to say, what must exist for its claims to be true. But secondly, says Quine, a theory expressed in the classical first order logic has no such unclarity. It is committed ontologically to precisely those things which must be reckoned as values of its bound variables for it to be true, because just those things are what must exist for it to be true. Quine's Rule - to be is to be the value of a bound variable - is a criterion proposed to support a general claim, namely the claim that progress will be facilitated whenever there is any problem with the ontological commitments of a piece of discourse, if a reexpression into classical logical

notation is made. Such a re-expression, it is held, makes clear just what *its* ontological commitments are, however it is with the original formulation. So it will help scientific progress to re-express theories in *canonical* notation, or *regiment* them, as he also puts it. (Interesting metaphors.)

The difficulty for nominalism which Field's work is intended to overcome is an argument whose locus classicus is Quine's "On what there is"³⁶. The argument is this:

P1 Our surest knowledge is scientific knowledge.

P2 Scientific knowledge requires mathematical theories such as R, the theory of real numbers .

P3 Mathematical theories such as R quantify over, that is have as values of bound variables, abstract objects including at least sets.

P4 To be is to be the value of a bound variable. Therefore,

C we are ontologically committed to abstract objects, specifically at least sets.

This argument for a realist interpretation of mathematics has been given its most extended statement by Putnam³⁷. Field accepts premises P1, P3 and P4, but he seeks to avoid the conclusion C so he aims to refute premise P2. Field's program follows a strategy inspired by Hilbert, using a technique extended from one of Hilbert's. The strategy (of Hilbert's metamathematics) is to show that an apparently large class of inferences is really available with smaller resources, a conservativeness result; and the technique (from Hilbert's work on geometry) is the representation theorem.

Field's argument is given in detail for only one theory of physics, Newton's theory of gravitation; but he thinks that his argument generalises to all scientific theories. There are two basic steps in Field's strategy for which he makes a detailed argument. First we have to get the numbers and suchlike out of the axioms of our theory, then we have to get them out of the inferences. The overall structure of his argument is as follows. First, Field constructs a nominalistically acceptable theory

³⁶ Quine [1948]. Recently challenged by Routley [1982] - with a very paragraphic title!

³⁷ Putnam [1972]

N, and proves that N can be represented using standard mathematics S as the usual Newtonian theory of gravitation NP. This means that any model of N is also a model of NP (in a particular way). So for any nominalistically acceptable assertion n, if NP \implies n then (N + S) \implies n. In many cases, n would be a theorem of N in fact, so that N \implies n. Examples show, however, that no-one would want to do physics without mathematics, because inferences readily made in NP would be impractically long if made purely in N. So in practice N+S would have to be used to obtain n even if physicists could be persuaded to go from NP to N in expressing the theory. The answer to this honesty problem is the concept of conservativeness. The conservativeness of mathematics over physics shows that one need not eschew the briefer inferences in order to be a good nominalist, for it is the fact that any useful brief inference in N+S must have a counterpart in N, possibly long, so that the truth of theorems of N does not depend on the truth of any thesis of S. That is, if (N+S) \implies n then N \implies n. These two implications together show that the P is dispensable from NP; if NP \implies n then N \implies n. In other words any thing which is a consequence of the usual Newtonian theory is actually a consequence of Field's nominalistic theory alone.

Field asserts, with some not terribly persuasive argument, that the same technique can be used on any theory, so that mathematics is dispensable from science. Since this, if true, disproves the premise P2 of Quine's argument, Field can deny its conclusion, that we are ontologically committed to numbers and suchlike. The novelties in Field's nominalist program are basically three. First, his reexpression, unlike those produced by earlier nominalistic devices such as Craig's transcriptionism, is an attractive theory in itself. He has an account of what makes it so which is interesting and worth pursuing elsewhere. Second, his argument that his reexpression shows the dispensibility of mathematics is much more thoroughgoing than others, in showing just how the original and the reexpressed theories are related through a *representation function*. Third, he admits that no nominalist could avoid mathematics in practice; but he has an argument to show

that this admission does not amount to taking back with the left hand what was given with the right. As he says, most attempts at nominalism or fictionalism about mathematical objects can be accused of intellectual dishonesty in simply admitting without any excuse or explanation that *practically* mathematics, in its usual unnominalized form, is essential for the practice of science. How the actual use of mathematical theorems can be combined with denial of their truth is what Field tries to justify through his adoption of the Hilbertian device of representation. The details of his argument are not really relevant to my purpose here, which is to highlight the important role of current conceptions of notation. In particular, Field's whole argument would lose all point if Quine's rule were false; and it would not succeed anyway unless the acceptable notation which does not quantify over numbers etc can be extended to replace *all* such present uses. So there are two questions: does notational re-expression reveal ontological commitment? And could all (scientific) discourse be put in a Fieldian notation? In the next two sections I argue for negative answers to both.

§81 Throwing away the ladder ?

In this section I attack Field's reliance on the idea that notation could by itself be a sufficient language for science. The Russellian conception of analysis using notation as mere re-expression or translation cannot survive the *paradox of analysis* objection made by Alston, by Jackson and by Wright, that no *synonymous* re-expression could possibly alter our "ontological commitments". But Quinean replacement-analysis avoids this objection by giving up any claim of synonymy. Yet Quinean regimentations are not really replacements, but **additions** to our modes of expression. The complete replacement of language by notation which Field's program requires is impossible. (Ironically, because it would undermine its own foundations, Field's demonstration that his reformulated physics could really be used without mathematics reinstates a kind of paradox of analysis. I will come to this in the next section.)

Some writers discussing ontological commitment in general, and Quine's rule in particular, suggest that the original sentence and its reexpression must be equivalent in meaning: the word 'translation' or 'paraphrase' is common. On this view it is hard to see why the reexpression should not have precisely the same ontological commitments as the original if it is an adequate reexpression. This seems to be an idea rediscovered periodically, it has been put forward by Alston, by Jackson and by Wright³⁸, each time as if a new idea. So that, for example, all that Russell's analysis shows, is that if the paraphrase [B] or [C] has no ontological commitment to an existing nonexistent monarch then the original [A] had no such ontological commitment either. In which case we may happily continue to use it. The analog of this argument for the Field reexpression of NP to N would be that if it is successful, and the reexpression N is not ontologically committed to numbers, it must show that NP is not really ontologically committed to numbers either. Certainly it seems obvious that if synonymy were the relation, it would be symmetrical - if [A] is the same in meaning as [B], then [B] is the same in meaning as [A]. So that if [B] is a regimentation with ontological commitments thus and so, so must [A] have those ontological commitments too: if [B] says or implies or entails or involves certain ontological commitments, then if [A] means the same as [B], it would seem clear that [A] must say or imply or involve the very same ontological commitments. If not, just what does 'same meaning' mean? So in this case whatever the advantages of one over the other, though they might be aesthetic or affective or even cognitive, they could hardly be ontological. Indeed, it seems that whatever the relation between [A] and [B] is, if it is semantically symmetric, then the reexpression can have no ontological advantages.

However, Quine explicitly denies³⁹ that the reexpression need be synonymous with what it reexpresses, at least sometimes he does. He seems to allow that it could almost (but perhaps not quite) be effected in ordinary language, but

³⁸ Alston [1958], Jackson [1980] Wright [1983]

³⁹ Quine [1960:159]

then of course the ontological commitments would not be so transparent (on his view). But certainly, the more careful claims of re-expressionists do not include synonymy. Rather than claim that the reexpression will serve equally as well as the original in all contexts, Quine's claim would be only that it will serve in all the contexts we care about, or in the one where the problem arose which provoked the reexpression. The re-expression is offered as as a "surrogate" or replacement which relieves problems. Now this claim is a trifle disingenuous in some cases in that it is often only the *nominalists* who are having problems with sentences like $2+2=4$, at least, problems which could be helped by proposing that numbers don't exist. Realists tend to have problems about how we know this kind of thing, or whether it is analytic, not whether it is meaningful or true. (Scientists don't have problems with it at all.) And anyway, neither of these kinds of perplexity can very readily be considered to be impediments to the progress of science. Yet ultimately Quine's justification for offering re-expressions is that they will help to facilitate scientific progress. Allaying the puzzlements of philosophers is not necessarily or obviously going to do that. At least, not on the austere picture of what science is with which he and Field operate. And not if it promotes doubts about the meaningfulness of mathematical ways of doing science. Has the enormous foundations-of-mathematics industry (which was created to try to settle the uneasinesses at the turn of the century) *helped* science by refining its logical weapons, or *hindered* it by distraction from empirical questions?

In any case it is just not true that new formulations *replace* the old ones. People still make jokes about meeting nobody in the road, and since no-one speaks canonical notation, nobody uses the formalese version [C] when talking about the king of France. People are still embarrassed if asked in court a question like "Have you stopped beating your husband?", because the knowledge that regimentation into canonical notation would defuse the presuppositions of this question (on some current theories) does not in fact help one in a courtroom. It is true that in new theorising the re-expressions may in some cases be used where previously old ones

would have been. This certainly happens in science insofar as ordinary language terms like 'hot' are not used in theories but more theoretical ones such as 'temperature' instead. But the old uses of 'hot' continue; and not only in unscientific everyday life, but in all extra-theoretical contexts, particularly in practical applications of the theory of heat, it must. It is after all the theory of *heat*. I can't myself observe temperature; but in any case I need to turn on the heater if it doesn't *feel* hot enough in my study, not at a particular temperature.

The general point I am making is that the real effect of notational or other semiotic innovations is to make *additions* to the semiotic devices in use and *complications* in the discourse which we must understand, not replacement of one device by another or simplification of discourse. This fact can be disguised or overlooked by attending, as re-expressionists tend to do, only to the new added discourse rather than to the whole, supplemented discourse into which it is introduced.

Quine remarks that our canonical notation has to be explained in ordinary language when he is emphasising its familiarity, as being like paraphrase. These explanations must remain part of the discourse of science, for it is only thereby that its explanations can be connected to the world we experience and the facts about it which lead us to seek scientific explanations in the first place. I have already pointed out the considerable launches which formalisations of mathematics require to precede their expression. To suggest (as Quine does) that having used ordinary language to get into canonical notation, we can then abandon it, ignores the way notation actually figures in text (as I have illustrated). It ignores how that process of "getting into" has formed our minds (which I shall discuss in Chapter 9) and how theorising must remain grounded in ordinary language, for its suggestions of problems, observations about experiments and statement of results, as well as applications to human life.

Thus the semiotic moves of re-expressionists like Quine and Field are **not** contributions to a global *replacement* of ordinary language by notation, only

extensions of it. That presupposition of Field's program is false. Another, his motivation by Quine's rule, is the subject of the next section.

§82 Does regimentation reveal ontological commitment?

There seems to be a fundamental paradox about such an attempt as Field's which doesn't even give up the use of the dangerous locutions. Surely whether we are "committed" to numbers doesn't depend on whether Field has produced his argument. But if we can happily use mathematics in practice and be committed, but without worry because we could avoid it, why should "commitment" matter? It seems to mean only that we *seem* to imply that there are numbers but that nominalistic physics, which we could "in principle" use to do physics instead, does not. So what? How can this affect whether there *are* numbers? Or perhaps the Field argument shows that there are no numbers, the apparent commitment was misleading; in that case it would seem that the Quine argument is not so useful - our commitments are not so clear as canonical expression seems to suggest - for Field began with the argument that science as it is entails the existence of numbers. In short, if there was anything objectionable to talk of numbers before Field's re-expression which has been dispelled, it can only have been that we couldn't avoid it. Yet we don't *have* to do physics. (I suppose God might be "ontologically committed" to numbers.)

Something like the paradox of analysis can in fact be revived against Field's program, due to the particular technique which he uses. I will now sketch this point. The essence of my objection to the idea that expression in canonical notation is ontologically pellucid is captured in Joseph Sneed's phrase 'burying structure'. In discussing the possibility of expressing a theory like Newton's in different ways he points out⁴⁰ that

"A great deal of mathematical structure can appear 'implicitly' in a predicate without appearing 'explicitly' as a part of the entities which satisfy the predicate. For example, when we define a function as the solution of a certain differential

⁴⁰ Sneed [1979: 195]

equation, there may be functions referred to in the definition that are 'part' of the entity defined. In this way, some of the mathematical structure that we intuitively regard as characteristic of the theory might get 'buried' in the predicates used in a certain formulation of the theory and not actually appear in the models for the predicates."

I suggest that Field is burying the mathematical structure of NP in just this way. For example in note 48, where he proves that his nominalistic definition of the relation PRODLSS is equivalent to the usual notion, there occurs the phrase "for sufficiently finemeshed R_{st} and R_{scal} ". Thus he slyly assumes (for his proof to work) that *equally spaced regions* of any fineness of mesh exist for both st and $scal$, where these are subscripts indicating equally spaced regions in spacetime and scalar mass respectively. But this is to make *numerical* assumptions about what subsets of space-time there are! Putting the point a little more accurately, his proof, which of course is being conducted in N+S, requires that for any small mesh size m , there is an equally spaced region R_{st} of mesh less than m . Now of course this platonic statement has a longwinded nominalistic counterpart, a purified N version which conservativeness assures us exists - if it is true - but notice two things. First, all regions are acceptable to Field no matter how bizarre or needing no matter what kind of specification. And they exist if they can be so specified. So the collection, let's not say set, of regions of spacetime in Field's theory has a considerable structure, far more than what the axioms of Tarski alone provide for segments. Field has, he admits, been accused of building the structure of the real numbers into his spacetime, and his response is twofold. He reiterates that it is purely the *abstractness* of numbers he can't accept, he is quite happy about their infinity or what he calls the *structural* assumptions about them. And, he claims, his points and regions of spacetime are not abstract, and are quite different from the reals *because you can't multiply them properly* - he grants a kind of addition is available. Here it seems to me he is simply wrong: the relation of representation is stronger than he wants it to be, I believe.

Webb⁴¹ asserts that Hilbert

⁴¹ Webb [1980 p77ff]

"was also anxious to oppose Kronecker's exclusion of geometry from pure mathematics on the ground that it dealt essentially with continuous magnitudes...Hilbert wanted to show that pure geometry did not depend on any kind of number...[but his Streckenrechnung showed that] numbers were not so different from geometric objects after all; however disparate their epistemological origins may have seemed to be, the essential properties of both were expressed in the field axioms."

The last sentence is certainly something of an exaggeration, since it ignores the specifically geometric predicates the field axioms do not directly represent, but the passage captures the ontological ambivalence of such a result. One can as well interpret the theorem to mean that geometrical objects are essentially numerical (since all models have a distance available), as that they are essentially not numerical (since distance is not in the axioms). Here is a parallel. Does the fact that the propositional calculus may be formulated using only \sim and $\&$ show that this part of logic is essentially free of disjunction, since the whole theory is available without mention of disjunction? Or should we say that that these operators are essentially disjunctive, contain disjunction within them or some such? Neither kind of response seems warranted if it is supposed to carry any ontological significance. This argument seems to me to be sufficient in itself to defuse any implication there may be in Field's demonstration which goes beyond the claim that talk of distance *may* be avoided; equally, it *may* be introduced, is what the representation theorem says. That is the way such a theorem is usually viewed - positively! Here is how Tarski states his representation theorem, which Field extends to prove his own:

*For M to be a model of E^2 it is necessary and sufficient that M be **isomorphic** with the Cartesian space $C_2(F)$ over some real closed ordered field F .*

So the representation theorem actually shows something much more damaging to Field's case: it shows directly that the axioms of R are true if those of N are. Just as Hilbert's result shows that the axioms of R are true if those of synthetic geometry are. This is just to say that if we may apply the theory N to the physical world, then R is true of exactly the same world. To illustrate this, just consider the axiom of R which says that there is a number, namely 0, which when added to any other

number gives that same number as result. The Streckenrechnung shows that this is satisfied in any geometry E^2 , by any segment [aa]. The uniqueness of 0 is not lost by the fact that any a will do because equality is interpreted in the Streckenrechnung by congruence! To object that this is not what equality is, is to give the whole game away: it is to confess that the ontological inferences one is pretending to make from the reexpression are merely assumptions that one is bringing to it.

§83 Surveyability

Summing up then, in this chapter I have shown that the logico-formalist account of mathematical discourse ignores important features of notation, in particular the fact that it must be grounded by words. I have shown how the prevalent misinterpretation of a notation as an attempted logically perfect language, a replacement rather than an extension of language, is a mistaken interpretation of Frege's Begriffsschrift and underlies Field's typically logico-formalist program of re-expression. I have also criticised the motivation of Field's program as dependent on a related misconception about notation, Quine's rule.

If we return to the actual properties of notation from these more fanciful misconceptions, notation is a written extension of word. The three main features of notation - abbreviation, formality and iconicity - together make **survey** of mathematical text possible by allowing the whole text to be visually assimilable and objectified. We need to be able to view it all together and to view it for its form. So these sign functions are versions of visual display and iconicity, which is why I say that an N-sign in context is a kind of diagram. Hilbert had a view of this kind⁴². We can also give credibility to this claim if we look at the less successful attempts to invent formal logic which precede Frege's - Peirce, Boole, Venn, Schröder etc. All these, including the Begriffsschrift, are clearly logic diagrams, and often are explicitly called that⁴³. The task was to find a way to *diagram inferences*. Frege

⁴² Cf Webb [1980: 88ff]

⁴³ Cf Gardner [1982]

found a powerful way, by adapting the preexisting notation of algebra, the two dimensions of the page and explicit diagrammatic elements, as I have pointed out in the previous section. One glance at some *Begriffsschrift* is enough to show that it's diagrammaticity: the lines involved are functional. The Peano/Russell notation which replaced Frege's disguises but does not eliminate this diagrammatic character: the dot brackets are clearly iconic, the use of page lines and alignment is crucial to it, and the rules of substitution depend on shape. The basic role of notation is to provide surveyable tokens (counters, samples, etc) which can be visually manipulated. An N-sign is iconic because as part of a system it shows, rather than says. This is the core of Wittgenstein's theory of notation. And I certainly agree that proofs have to be surveyable - some of them anyway: the concept of surveyability has been and is being stretched - for example by the "enormous theorem" (the 100000 page proof of the classification of the finite simple groups). I will take up the topic of the surveyability of notation again in Chapter 12 when discussing Wittgenstein's philosophy of mathematics, for it is the source of the most profound consequences of notation. In the next chapter I will consider the last of the four signsystems of mathematics, Paragraphy.

Chapter 8

Paragraph

§84 Nature of the Paragraphy signsystem

This chapter will be shorter than others, not because Paragraphy is unimportant but because it is largely ignored. There are no misconceptions about Paragraphy in the literature of philosophy of mathematics, as there are about Diagram and Notation, because it is not discussed at all. So in this chapter I shall continue with the exposition of the mathematical sign-system as I have in the previous three chapters, but without any immediate application to prevailing opinions. The conception of Paragraphy which I develop will be applied in the following chapters to improving the concepts of mathematical context and practice.

One reason for the marginalisation of Paragraphy is its natural specification as a residual category - those signs in mathematical text which are neither W nor D nor N. There are no "naturally" paragraphic signs. There are no "naturally" N-, D- or W-signs either in fact, but to find '+' used as word, or 'rigmarole' as a variable would be so odd as to inevitably suggest some ulterior motive, some second-level signification. Actually, there *are* signs which are used standardly as Paragraphy, for example punctuation signs. But for our present purposes, the most important P-signs are ones which in themselves might be W- or D- or N-signs, or composed from all three, but are *put to paragraphic use*, such as titles, indexes and footnotes.

Paragraphy begins as a supplement. In speaking, there is some scope for physical arrangement of the successive signs by pacing, intonation, volume, timbre, emphasis and rhythm. In simple speech situations accompaniments such as gesture and facial expression are usually used too. These all contribute to the meaning of what is said. The common denigration of these meaning elements is a consequence of the value given to written text, in which they are not present. In many spoken communications, these elements are far more important than the actual content. This is why politics, social negotiation and life in general are so much more difficult and complex than philosophical discourse about them sometimes suggests.

One source of Paragraphy is the attempt to find substitutes for these unrepresented aspects of speech. As is usually the case in the semiotic world, these inventions soon take on a life of their own. Kings were painted bigger than citizens in ancient Egypt; headlines nowadays attract attention by the same device. Supporting Paragraphy as a reaction to this need is the nature of the medium: the writing page is a two-dimensional surface on which the written signs are displayed. They may also be disposed and embellished and so on - possibilities Frege took up in *Begriffsschrift*, as I have discussed in the preceding chapter. A second kind of need which Paragraphy meets is created by extensive texts: readers need mechanisms to assist their grasp of the "logical space" of the text, which is commonly, particularly in mathematics, not a simple linear tale. For example, in Heath's Euclid, references are given to theorems or axioms as they are used. This makes it possible for a reader to understand the whole text without having to memorise it all in order to know the location of any proposition which is used. The third important motivation for the development of the paragraphic system is the proliferation of texts and readers. In order to enable a large and unevenly dispersed community of readers to have effective access to a large and unevenly dispersed resource of stored texts, mechanisms of reference and citation, concordance and index and so on have been developed. Most texts contain paragraphic signs which locate the text with respect to this prior discourse.

It is easy to dismiss Paragraphy as making no essential contribution to the text because it is by definition not part of the "real" content. But even for users who conceive of it purely as the transmission of information and aim to use it so, communication actually fulfils various functions simultaneously and its success or failure in any of these roles, even the transmission of information, depends on the pragmatics of the situation. Paragraphy can make the difference between a communication which is understood and one which is not; there can hardly be a greater contribution. Of course, it is always possible to see how the same message could have been conveyed with less Paragraphy and more time and effort - this is

the argument for poor textbooks and bad teaching. But this kind of appeal to what is "in principle" possible is of no interest if what we want to understand is how certain kinds of knowledge and knowledge contexts are possible *given our finite limitations*.

Paragraphy facilitates communication, and by facilitating it makes possible what would not in the circumstances be possible without it. The idea that the world's present research effort could "in principle" continue if all paragraphic aids were removed is correct but absurd: it is equally true that we could all be entirely self-sufficient, make our own clothes etc. But not with the same resources we actually have. The objectificational and grounding roles of Paragraphy make possible the existence of cumulative traditions. The term 'cumulative', though traditional, is rather a misnomer in that it simply means heaping up, whereas the peculiarity of the kind of tradition for which mathematics is the paradigm is just that it consists of a vast but dispersed variety of different but related texts, not a collected mass of interchangeable copies like a heap of grains of sand. The articulation of such a discourse, which is the sine qua non for such traditions, is effected through Paragraphy. Each new contribution to such a discourse presumes the existence of the corpus of previous texts in it, and typically makes reference to parts of it. There is no metalanguage in mathematical discourse, but every text has metatextual elements. The written word has that relation to the notation and the diagrams, the Paragraphy to the written word: these are typical relations within a single text. Between texts they can be much more complex. Many later texts make into their subject matter paragraphic elements of earlier texts, reversing the 'natural' relation. The mathematical discourse is a very complex and multilayered construction from such citation, enframing and anaphora.

These text effects of Paragraphy, and the characteristic text acts effected in text through its use, are what justify its identification here as a separate system.

§85 What Paragraphy comprises

In view of the novelty of my claim to identify Paragraphy as a separate signsystem in text, it might be useful to consider a criterion for a sign's being paragraphic. Most philosophers would deny that writing does anything that speech doesn't, and the claim is often heard that mathematics can be read aloud¹. The falsity of this follows from the independent needs for mathematics to be written, to use diagrams, and to use notation, all of which I have argued already. At this point I will simply claim that when text is read out, paragraphic signs are ignored: the criterion I propose is that a sign is paragraphic if it will normally be disregarded in reading the text aloud (rather like the contrast between the speeches and the directions in the script of a play). Of course, a believer in *nasc*s² will jump on the caveat here - "normally" - to prove this criterion useless and circular. But I have already indicated that I think *nasc*s to be undesirable and unobtainable for concepts with content.

In fact, I don't believe that anyone is going to be in much doubt about the classification of a sign as paragraphic after I have discussed an example text or two. Nor would much depend on any hard and fast classification anyway. My overriding aim is to show that there are a large variety of important text acts and text effects to be found in mathematical text. My classification schemes are means to that end. All the same, I believe that there really are these four semi-independent systems at work; but the boundaries between them are not *rigid*. We have already seen that framing effects exist at various levels in text, and part of the reason is that some Paragraphy is intercalated with other kinds of sign. Let us look first at exhibit #4, the page CS174.

There are two groups of P-signs to be seen here. There are a few *framing signs* which I have labelled P1 to P4; and there are a large number of *intercalated signs* which I have not marked. The most obvious of these latter are the usual

¹ Teachers often say that a good way to test the sense of mathematics you are writing, is to try to read it out aloud. I am dubious about the value of this advice.

² Necessary and sufficient conditions.

punctuation signs such as '(', ',', '?' , and so on. But there are some more subtle signs to be found which I also classify as paragraphic. There is *capitalisation*: the first letter on the page is an upper-case 'S'; that it is upper and not lower-case is significant and the choice of case is a further choice of sign beyond that of the letter 's'. Next, in the ninth line of the text the word 'any' is printed **bold** for emphasis. Again, this choice of *style* is significant above and beyond the choice of the word 'any'. Next to the clearly paragraphic sign P1, is a sentence 'What percentage...' which shows us another form of Paragraphy - it is printed in a smaller *size* than the preceding lines. This indicates that it is slightly *aside* from the main exposition. Similarly for the whole of the subsequent "example". There are a number of further choices which the authors have made in composing their texts which are not without significance (though they are not all equally important, of course). There is interline spacing of various densities; the diagram has a specific size compared with the sentences and the page; compactness has been weighed against clarity in deciding when to start new lines in many cases. Turning to the four P-signs which I have explicitly labelled, the first three are all *labels*, the fourth is a *footer* which contains page number, chapter title and a couple of D-signs (straight lines) which clarify the relation of the whole sign P4 to the rest of the page.

This simple example will provide sufficient motivation for the classification of P-signs which I shall now introduce. Different kinds of sign can be found used paragraphically. There are special signs invented for the purpose, such as punctuation marks - e.g. ", {...} - ; as well as typographical sign-features which trade on the specifics of the written mode by marking physical features of the page appearance, such as the choice of font and its size and the amount of border, or whether footnotes are at the foot or at the end of the text. There are also the organisation of the text into parts; and signs taken from the other systems of diagrams, written words and notation and used out of the main stream of text. I mean here such signs as boxes and lines dividing up the page, and titles and page

numbers. The mixed nature of these elements derives from the fact that the paragraphic system is originally a secondary one. I classify Paragraphy into five formal categories: *paraliterals*, *rubrication*, *disposition*, *paragrams* and *paratext*. I will illustrate this classification with further examples taken from the sample texts. I have marked on the exhibits some of the paragraphic signs. Rubrication and disposition, such as differences in font and border size, are not easily highlighted; but I shall indicate some of these in what follows. I will show through these examples how paragraphic signs of all five kinds can contribute to the paragraphic functions which I shall shortly identify. In the rest of this section I will give a complete account of the framing Paragraphy of the Hansen text, exhibit #7A; this will afford examples of some more kinds of P-sign. In the next section I will explain my classification.

There is of course much intercalated Paragraphy in the exhibit #7A, but for present purposes the 27 framing P-signs which I have highlighted are more important. They are as follows. P1 is a self-referring label; P2 title of the paper; P3 author's name; P4 a kind of date label; P5 an editorial claim; P6 a label for P7; P7 a summary of the paper; P8 & 9, labels for reference to this paper; P10 title/label for section 1; P11 a copyright tag; P12 the internal page number, P13 a short title of the paper, P14 the volume page number (P12-14 constitute a header); P15 -20 are labels for short blocks of text, except for P16 which is one for the whole of section 2; P21-3 are like P12-4 except that the author's name replaces the title of the paper; P24 is a label for part 3 of the paper; P25 labels the bibliography; P26 is the bibliography; P27 is the author's address. Notice that none of these signs is part of the mathematical content of the paper. It will be apparent that many of them are *labels* and many of them are concerned with *citations* of one kind or another, mostly to or from this paper or within it. One or two, such as P4, P5, P11, will only be fully explained **much** later.

§86 Formal elements of Paragraphy

In this section I will explain my formal classification of P-signs, and indicate the usual functions of those that I call intercalatory. The functions of framing P-signs are more important for us and will be the subject of subsequent sections. The main point of the present section is to draw attention to the ubiquity and variety of P-signs; the actual classification will not be much used below.

Paraliterals

First, there are additions to the basic set of characters, as with brackets and spaces. These signs occur within the line of writing as do the actual letters or characters. I call these extra characters *paraliterals*.

Paraliterals comprise a number of kinds:

- | | | |
|-----|-----------------|-----|
| [1] | + = - / ^ | etc |
| [2] | , ; : . ? ! ... | etc |
| [3] | () {} [] - - | etc |
| [4] | " | etc |

Examples of all these may be found in all the exhibits.

Class 1 are actually elements of notation, and are mentioned here only because they also count as extra characters beyond the alphabet; but since their role is not usually paragraphic, but notational, I will exclude them from the Paragraphy system. Notice that they must be read out when text is read aloud. Class 2, *punctuation*, are signs which try to mimic the effects of timing (,:), mark sentence boundaries (!?.), indicate mood (?!) or play purely textual roles (...). These signs may have some kind of representation in reading aloud, but their correspondents are so weakly determined that they belong in the paragraphic system all the same. Early writing lacked such devices; perhaps the way to think of them is as not belonging to any particular word. Their role in text, though originally imitative, cannot be so confined. Question marks are not found only at the end of questions - for example, they are frequently used in comic books to signify a puzzlement or disbelief which

would be hard to put in words. They may also be used in the margin of a mathematics text to show where the exercises are to be found - though these need not be questions at all.

Class 3 are signs which typically indicate subordination. Formalisation makes great use of this kind of device to "improve upon" speech in point of ambiguity. In speech, subordination is conveyed by grammar and pacing and intonation and all the other paralogistic accompaniments of speech. Devices which originally may have been introduced to try to substitute for some of these have in the end become the arbiters of subordination. The great discovery of logic in recent times is that relations of subordination much more precise and rigorous than those of speech can be imposed in writing. Brackets are the prime example of the way text goes beyond speech here: the use of brackets to disambiguate scope has no true parallel in speech.

Class 4 are signs typically used for quotation. This is something on which there is some philosophical literature. Mostly it shows a remarkable aversion from noting that explicit quotation is a *written* phenomenon. The less sensitive treatments generally do no more than warn against confusing use and mention - a distinction *created by* the possibilities of Paragraphy - as though it is an obvious and simple distinction. A few writers have suggested otherwise, but in the logico-formalist hegemony they are dismissed as "illogicate"³.

Rubrication

Secondly we have originally supplementary signs which modify or qualify others: accents or font choices or type styles such as underlining. This kind of thing is often called formatting, but I shall call it rubrication, recalling the illuminated

³ A term which I have invented to parallel to 'illiterate' and 'innumerate', to describe the common "deficiency" of many philosophers, of being less than ideally conversant with the niceties of formal logic. I am surprised no such term of abuse is already in circulation considering how often people who raise objections the logico-formalist hegemony are condemned as not understanding it. For example, the early reviews of Wittgenstein [1956] in Benacerraf and Putnam [1964] take this line.

initial letters of manuscripts. We can distinguish between (at least) font, case, size and style.

Font refers to the design of the set of characters - are they tall and thin, do they have serifs, and so on. For example, the previous sentence is in a seriffed font called Bookman, but this one is in one without serifs called Monaco. The function of font can be a rather subtle one - compare the fonts used in eighteenth century French books with those of the gutter press today. In mathematical text, it mostly offers contrast - use of several different fonts in one text can provide for emphasis and make for stratification (see §89 below). In the CS174 text, the labels P4, P6, P7 and P11 are all in a sans-serif font, whereas the main text is seriffed. This helps to distinguish the main text from the Paragraphy.

Case is choice of upper or lower case letters, such as A or a. The function of *case* is not always entirely clear. In German, nouns are indicated by initial capitalisation, as well as as in English to initiate a new sentence. Whole words may be capitalised for emphasis.

The functions of *Size* and *style* (italic, bold etc) are partly obvious - relative emphasis; and partly more subtle - tying together text segments which are at the same level but physically separated. For example P3 in CS which is in *italics*, emphasises the generality of the claim being made there. Again, all the section headings in this thesis are **bold**. This contributes to its cohesion as a readable text and helps a reader to form a structured grasp of its organisation and argument.

Disposition

Third, there is the spatial deployment of the writing on the page. In this category I distinguish alignment and segmentation, the white space usage on a page - page layout generally, and location. *Alignment* includes use of subscript and superscript, underlining and so on, choices about the signs which depend on the direction of their juxtaposition (such choices do not arise in a linear medium like speech). The possible importance of such sign devices is well shown by the power of

iterated exponential notation. By *segmentation* I mean the division of text into pages, paragraphs, sections etc, which depends on the use of choices about the line regime (left-to-right, boustrophedon or whatever); number and width of columns; whether diagrams are in the stream of text or separate; the closeness of the characters on the line and between lines (leading, kerning), the use of margins. These choices about the amount and use of the white space on a page are of some consequence, because a page can be surveyed, but separated pages cannot be compared with the same ease as different parts of the same page. All these devices can be and usually are used for stratifying the text, that is for imposing a hierarchy onto the uniform stream of letters which comprise the main written text. I will discuss this further in §89.

Paragrams and Paratext

I call *Paragrams* signs whose origin is in the diagram system such as lines and boxes, which are being used for paragraphic purposes. An example is the line segment used in the footer on CS174 to make clear that it is not part of the main text. There are also lots of literal signs - letters or written words - which play paragraphic roles. I will call them *paratext*. Paratext and paragrams partly play the role of *text organisation*⁴ in the case of what is on a page of the main text. But there is also a lot of material outside the main text in title and contents pages, indexes and so on. In this case the paratext will generally be citational in its function. Paratext is mostly pieces of extra text which assist the reader to navigate the main text: title, contents, index, pagenumbers and so on. These devices do not

⁴ The literature of instructional media calls some of the paragraphic signs "text organisers", or sometimes "advance organisers" because the reader is thought to use them to prepare for detailed consumption of the text. But what is included under such a label, and how it relates to general semiotic issues are not very deeply discussed there. Cf Hartley [1985] for examples and further references.

usually use special signs, but the signs used are not part of the main text. They are part of the Paragraphy system because of their role⁵.

This brief sampling of Paragraphy may not convince the reader that it is important, but I think it is at least obvious that there are a lot of signficatory choices which are generally neglected in the literature. To show that this neglect is a major mistake is the purpose of the rest of this chapter. The following sections introduce the main *functional* aspects of Paragraphy. I distinguish in §87 a variety of semiotic elements underlying paragraphic possibilities. In §88, I point out the great variety of individual text acts which can be made with Paragraphy, and briefly discuss affective uses of Paragraphy and true asides. In §89, I turn to the nature and significance of a text's articulation, its internal hierarchical organisation. In §90, the grounding of a text in its discourse and the Archive is shown to be the crucial basis for cumulative traditions. The overall effect of Paragraphy, text objectification, is connected in §91 to certain cognitive values which have developed with literacy and which are necessary for the existence of such traditions.

§87 Functional elements of Paragraphy

Having indicated what Paragraphy comprises formally, I turn to the more important question of what can be done with it. As for the other sign-systems, I will propose a tentative functional classification. The great variety of P-signs have a few main elementary functions, of which *spacers*, *contrastors*, *labels* and *cttors* seem particularly important.

Spacers are P-signs which organise the location of the other signs on the page. P-signs which contribute to the linear spacing of alphabetic text, which might be called *punctors*, seem to be originally an attempt to fill the gaps which early transliteration techniques created between spoken and written forms of the "same"

⁵ Occasionally one finds obscure signs at the foot of a page of a printed book which can only be interpreted as signs from the printers to themselves, labelling pages for assembly. Such signs contrast with paragraphy, since they *really* are extraneous to the text!

utterance. They clarify the identity of the speech signs which have been transposed into written form. Commas and full stops correspond to pauses, and so on. But some signs, like brackets, seem to be immediately open to novel signficatory uses. Thus punctors are a species of a more general genus.

The signs which I have called rubrication are generally used as **contrastors**, which are signs whose role is to differentiate one part of the text from another. One example in the Hansen text is the use of italics for the actual statements of the lemma and theorem, W9/10 and W25/26. A second, more complex, is the double contrast between on the one hand the size of the main text, the smaller pointsize used for most of the framing signs such as P4-P9, and the larger one for the title of the paper, and on the other hand the use of bold for the section titles.

Labels are very common among the P-signs we have seen in the exhibits, which are by no means unrepresentative. The earliest mathematical text in our exhibits, Euclid's, already uses them quite essentially. Proposition 47 is so labelled, so that it can be referred to later on. We noticed in examining the diagram system how salient was the relation of labelling there: points in diagrams have to be labelled so that they can be referred to in the text. The reason for paragraphic labels is the same: devices are required to effect reference back and forth between different text segments. All the various kinds of title in texts, of which we have seen some examples in the Hansen paper, are labels; but so are theorem numbers, signs like 'Lemma' etc, and page numbers.

Citors are P-signs used in making citations. *Citation* is a paragraphic text act which is common and obvious, in which one text is mentioned in another. It may seem to be effected in the main text through W-signs, for example, Hansen W1, but usually this is mediated by a bibliography, which consists of citors, and in any case to be successful there will need to be paragraphic access *to the text which is cited*, through its own labelling. There are three main kinds of citation. *Intra-citation* is the citation by part of a text of another part of the same text. Euclid's '[CN2]' is a classic example. One of the reasons why there are so many labels in text

is to make intra-citation possible. (There is no explanation in *Principia Mathematica* why the various theorems need to be numbered since this is clearly not part of the "real" mathematics.) *Extra-citation* goes in two directions, into and out of the text which is being considered. The text itself can only cite outwards of course, but it contains P-signs whose function is to mediate citation into it, such as Hansen P1 and P2.

§88 Elementary paragraphic text acts

The great variety of paragraphic devices, some of which have been exemplified in the previous sections, are naturally put to a great variety of particular uses in *elementary paragraphic text acts*. We could spend a great deal of time investigating the specifics of text acts like *titling* and *referring to the literature* as individual communicative acts, in the spirit of Austin and Searle. One brief example will have to suffice. In order to successfully *title* a paper, an utterance must appear in appropriate contiguity to the rest of the paper. It must be possible for a reader to see at a glance that it is a title and not the first sentence of the text or an unrelated remark. This is achieved for the Hansen paper through the use of a whole standard format of framing Paragraphy which precedes and follows each paper in volumes of the Journal of Australasian Mathematics Association. The title is printed in a large size, bold, centred on the page and well set off from what follows. It is the first sign on the page except for the *citor* P1. The experienced reader won't take *that* for the title, but a less experienced one presented with only this text might well be puzzled by it. The title achieves several things. It gives the reader some idea of the main point of the paper, preparing for receptive reading (a poor title fails in this); it provides another means of reference to *this* paper by the author; and it marks the start of the text. It is clear that there are many felicity conditions on a *titling* paragraphic text act, several text effect consequences of it, and that it has a specific point (*entitling*). All of these differentiate it from other

paragraphic text acts, for example citation of an authoritative source, or reference to an appended table.

The presentation of the text in a form which will be conducive to receptive reading has already been pointed out as part of the point of titling a text. There are various ways in which signs are used in order to make this general effect of "user-friendly" text. Many of them have effects which while undoubtedly real, are too diffuse, too interrelated with others or too debatable, for us to do more than indicate them here. For example, a font may be chosen having regard to legibility (research shows that serifs aid legibility), but it may also be chosen with predominantly aesthetic criteria in mind. Aesthetic value is not divorced from cognitive advantage by any means, but it is very hard to say just how an attractive font contributes to the cognitive availability of text. Many valuable choices made in the composition of a text we shall have to put in this "too hard basket" for the present. These signs contribute to the general articulatory and citational aims of text organisation, but they do so in a way which adds to the general attitude of the reader. The choice of printing colors is contrastive, but it also aims to make reading a restful experience; pink ink on black paper for Euclid would make it hard to follow the sense of the text. (Of course there is a tension here, since it would certainly give great contrast!) The horizontal line which is part of the footer on the CS pages, itself a paragram, has an apparently purely affective role in that it helps the reader to distinguish what is "text" from what is footer. It makes the text more "user-friendly" than it would be without it. It makes effective use of the text easier but not different. As such we might call it decoration. But this is not necessarily to denigrate it. Some such signs might seem worthy of denigration: recent first year calculus texts have all had two printing colors. The highly varied choice of the second (non-black) color seems to aimed at being unusual as much as restful on the eye. Yet there would seem to be by contrast valuable signs of this kind too.

Another species of text effect achieved through Paragraphy, which I mentioned already in the previous chapter, but which will be merely noted here, is

the *aside*. Footnotes, W-signs in brackets, and many other devices are available in text to create true asides. By "true", I mean that in contrast to speech, where the stream of talk must be interrupted to make a parenthetical (!) remark, one of the consequences of the endurance of text is that it is perfectly possible to say several things at once. Of course, the reader cannot *read* two things at the same time, but because the text endures, she can read one and then the other and then the first again. There are many variations on this which provide different forms of effectively simultaneous consumption of different text sequences - kinds of "parallel" processing of text. Although reading W-signs is like hearing, one at a time, text need not consist of just W-signs, so that one can attend to a diagram while reading the words. And a text need not be read in any particular way. Something that many people learn poorly is the variety of access points and routes through complex text which increase its value⁶. All of which is predicated on its not being read just once, which in turn is possible because text endures.

In this section I have suggested that the many kinds of P-signs can be used to effect many different paragraphic text acts, and given some examples. In the next three sections I will discuss the three most important functions of Paragraphy, text articulation, grounding and objectification.

§89 Text articulation

In concert the many individual paragraphic text acts articulate the text and the discourse to which it belongs. Spacers and contrastors form the text into segments, which are labelled in various ways, and connected by a web of intra-citations to one another, and to the textual context into which the text fits through extra-citation. The citations create a network of subordinations among segments, which produces text stratification. The textual structure thereby produced is important for its being understandable, and for the nature of what is understood from it.

⁶ Cf Morris and Stewart-Dore [1984] on trying to teach such skills.

The simple logico-formalist concept of articulation, the linear chain of elements jointed together in sequence, does not apply to paragraphically organised text so much as collapse under it. Firstly, paraliterals are simply inserted into the articulation of the phonetic text. Thus their inclusion destroys the simple relation between the articulation of speech and that of its image - linear juxtaposition - which is assumed in the transcription notion of writing. Second, paratext obeys within itself all the laws of articulation of the simple written word, but its relation to the main text cannot be accounted for within that scheme. It is *intended to be outside it!* Third, disposition provides many possibilities for articulation which have already been observed in connection with diagrams and which apply with equal force here. The number of different relations among signs which are simultaneously in play on a two-dimensional page is vastly in excess of the simple relation of succession. Fourth, the way that rubrication combines with the rest of the sign system defeats the whole notion of articulation, since the sign functions in question are not more signs joined in sequence with the others somehow, but modifications of the signs which are already there. In this they resemble the features of speech lost to writing, loudness and so on. The main role of Paragraphy is an articulatory one, but when we observe Paragraphy and its role in text we see that any simple notion of articulation is inadequate for the reality of text. Let us try to develop a more complex one.

Different text **segments** are created by spacers and contrastors. In CS174, the piece of text which is labelled 'Exercise', is constituted as a segment by its smaller point size, its bracketing and its increased interline separation from the text around it. Having been chunked and differentiated from the rest of the text, it can be labelled as 'Exercise'. It is usual for a number of P-signs to combine as in this example to effect this *segmentation* of the text. Segments of a great variety of extents can be found in texts, from the whole main text versus its framing Paragraphy, down to a single sentence as in this example. Perhaps one reason for the neglect of Paragraphy as a sign-system is that one of its main roles is in this

constitution of *supra-sentential* text segments: much linguistics and philosophy of language operates under the curious illusion that the single sentence is an adequate model of discourse.

Segments are labelled so that they can be referred to. Texts typically come with labelling on the title page, pages, lines chapters etc; with theorems, diagrams, equations labelled; with an index and bibliography. Labels always employ relatively simple disposition to make clear what they label. Contiguity and alignment generally play the main roles here, as in the labels on CS174. We know that 'Example' labels the rock-concert text segment, because that text is unified by its size contrast from surrounding text, and the P-sign is located next to and parallel with it. One could not plausibly take any of the other text segments on the page as labelled by it instead. Several of the segments of CS174 are labelled for us, as is the page itself. These labels permit citation of these segments from elsewhere in the text, for example in the contents where this chapter is listed, or the index where 'normal distributions' is listed. The page label allows the casual flipper to find the material on normal distributions without the index, since it is a large segment. These devices all permit a reader to locate and make use of such segments without (re)reading the whole text. But this is not their most important function. Most significant is the *stratification* of the text which the network of intra-citations creates.

Texts are segmented into lines, pages, chapters etc, which lie in a natural relation of subordination based on inclusion. But it is also segmented into main text and Paragraphy and into theme and illustration. From the page of CS in the exhibit we can immediately see that the text from which it comes is divided into chapters in which the main "story" is presented, illustrated by examples and supplemented by exercises for the active reader. This page is marked as p174 from several hundred and is part of the chapter on normal distributions. The page itself divides into four main blocks together with the diagram, and four framing paragraphic signs P1-4. Thus the page has significant segments, and is itself a

significant segment. A similar segmentation of the Hansen text is apparent. But it is also apparent that the system of labels in the Hansen text supports another and cognitively more significant relation of subordination, with the theorem at the top, and other segments progressively less important: the lemma, the two proofs, the definitions, the references to the literature, the example in section 3. The framing Paragraphy provides a third relation of subordination among the text segments which turns on the relations between this text and others. Here P1 is top segment, with P2,3,27 next, then P26 and so on. A further fine network of subordinations is created within the main text by citations of one part of the text by another, for example W38 refers to the "hypothesis on the ring" i.e. it cites W25/6. Notice that we can identify what is the hypothesis on the ring, because the theorem is stated with a label identifying it as such. So we, as experienced readers of such text, know that any conditions on the ring being discussed in the proof will be found here.

All this complex structuration of the text I call stratification. The cognitive complexity of mathematical text is vastly increased by this *text stratification*. What is understood by a successful reading of this text is highly structured by the textual structure. It is simply not true that the mathematical knowledge which a reader gains from it is structured as the logico-formalist account would have us believe. On the contrary, it is structured as the text is, because its acquisition has to be effected through a reading process whose complexity is itself great, and whose structure is suitably matched to that of the text.

§90 Text grounding, discourse and Archive

In this section I will argue for the following claims. The paragraphic devices which structure the text are complemented by others which, by relating it to other texts, create a greater structure, a *discourse*, which is the basis of a cumulative tradition. The prepared reader is aware of this discourse as part of the context for making use of the text. The Paragraphy which makes it possible *grounds* the text

into its context. Discourses are inter-related and together comprise the Archive, the textual product of past thought, and basis for its further development.

An individual text like Hansen's is enmeshed in a very large network of relations to other texts via the labels with which it is furnished for just that purpose. The most obvious grounding device is the explicit extra-citation, mediated through the bibliography in most cases. But there are others. P1 of the Hansen text is a paragraphic device which enables other texts to refer to it easily and in a standard manner. Titles and other P-signs also allow this reference *into* the text. The author's name and address, and those of the Journal, permit other kinds of reference from the text to related texts and communications (P5, Hansen, is an interesting example: "Communicated by R Lidl"). In fact every one of the P-signs in the Hansen text can easily be imagined as used in citation one way or another.

Most texts have paratext which locates them in traditions, genres and series ("Fontana Modern Masters"), introductory material which smooths the reader's path into the text's world, advance organisers which put the reader into the right frame of mind for consuming the text. We know we are in a didactic text with CS174 because of the use of the term 'exercise'. Many other such cues are available to locate that text in the mathematical tradition. The use of the terms 'proposition', 'theorem' and so on relate countless mathematical texts to the early Greek ones. The labelling of the propositions in Euclid is for purposes of intra-citation, but it has been used subsequently immeasurably more often by other texts referring to Euclid. This is a cardinal example of the *appropriation of text by later text*. These links do not contribute directly to the "content" of the text, but they inform the reader's concept of what kind of content it is, by marking the text as in the mathematical discourse, and not, say, literature.

Complementing paragraphic devices in the text is a complex network of access aids, such as indexes, "current contents" periodicals, librarians etc. A tradition like mathematics is supported as a discursive formation by an organised collection of texts which are marked as relevant to one another by their framing

Paragraphy. The supplementary aids which relate texts to one another can only be created if the texts they relate contain suitable labels through which they can be referred to. The reciprocal relations among a collection of texts and their supplementary literature constitute a discourse, of which mathematics is a prime example. Such large collections of texts require methods of access. Once scholars owned or had ready access to significant numbers of texts, comparative scholarship became possible and various kinds of aid were developed to assist it - the index, the glossary, the table of contents and so on. This was the point at which the technology of the printing press impacted on the nature of text. The general features of the way print changed the scholarly world have been magistrally shown by Eisenstein⁷. Some other historians of the book, such as Lefebvre, have also described these fascinating developments; but unfortunately Eisenstein is the only source which emphasises the cognitive consequences.

Of course, the boundaries between one discourse and another are vague - one might take mathematics and physics together, or distinguish within mathematics the discourse of geometry. Nevertheless, the distinctions are real because they inform the ways in which texts are read, and therefore the way in which their contents are understood. One of the sources of cognitive innovation is the realisation that something belonging to one discourse can be used in another, like using logical results in analysis (Robinson) or analytic methods in number theory (Hardy & Littlewood). These transfers are difficult just because textual contents are understood, bodies of knowledge constituted, by these contrasting discursive localisations. The summa of all these discourses is the Archive, the accumulation of all our textual resources, larger or smaller samples of which can be found in large libraries, but which is unevenly dispersed around the world. I will discuss this at some length in the next chapter. Texts written now for publication inevitably include a great deal of Paragraphy; some of it is probably legally required. Texts are produced in consciousness of their being additions to the archive and therefore

⁷ Eisenstein [1979]

include mechanisms to ensure that they are. Moreover they are generally expected (to varying extents, depending on the scholarly "level") to locate themselves vis -a-vis the pre-existing discourse and Archive.

§91 Textual objectification

Through stratification the reader's attention is drawn to the segments of text as objects individuated from other text segments. I emphasise the variety of the extent of segments. Mostly they are not sentences. Citation involves the reader in an interpretative process in which the use of text segments as objects of thought, reference and display is common and normal. Together these semiotic functions provide the text with its overt macrostructure⁸. Together with the framing Paragraphy, the results of these devices are to create a complex network of text segments which are in various relations of subordination, and to objectify these segments as cognisable items in their own right. They contribute to the objectivity of the present text or text segment by making absent texts and segments compresent and related in the mind of the reader, as objects to be called on, thought of, referred to while the present one is being attended to. Even decoration facilitates this complex process by making more easily achievable the perceptions which must be made for it to succeed. Texts are seminal just in so far as they permit such success, and ensure therefore that they will often be referred to in subsequent texts. The great exemplars in Western tradition are the Bible and Euclid.

As a result of these uses of Paragraphy, different segments in the text are at different levels of objectification: there are "simple" statements or questions, like the previous sentence. There are others which advert to, or comment on, or otherwise presume the presence of such simple signs, like footnotes. At the next

⁸ Phillips' [1985] attempt to identify the macrostructure of science texts by a statistical analysis of word contiguities seems to overlook the role of segmentation and labelling in providing the reader with a grasp of the text's macrostructure. It may give us some insight into the mechanics of text construction, but seems ill-suited to the more general task of clarification.

level are sentences like the previous one. A further complication is introduced by sentences like this one which allude to or refer to themselves, directly or indirectly. The use of Paragraphy makes indirect self-reference quite common in text. The creation of levels by the use of chapters, sections etc facilitates and makes common the practice of referring to other parts of the text; the use of the several different sign-systems in mathematical text also adds to this Russellian nightmare by the concomitant introduction of devices for referring between text and diagrams, diagrams and notation, notation and text.

The grounding of the text into the much larger discourse means that the content of the given text is like the tip of an iceberg: a much greater content is already present and at work in the mind of the reader trying to understand it. The structure of what is understood in the text is largely external to it, being given by relations among the texts of the discourse which precede it, in the reader's experience, or historically, or both. For the texts in some tradition to have developed such a structure, the readers must necessarily have developed concomitant skills of reading them, and matching conceptualisations of knowledge, of texts and of their relations (albeit implicitly perhaps).

I shall be arguing in Chapter 12 that the concepts of objectivity and objective knowledge, and the object problem in philosophy of mathematics receive much of their force from this process of objectification which text undergoes. Beginning from the simple fact that texts, unlike speeches, are objects, the Paragraphy system deepens and variegates this *being-there* in many subtle ways. Some of these have been studied by anthropologists interested in the socio-cognitive consequences of literacy⁹. The psychology involved here is presumably related to the idea of "chunking" often used (under various names) in discussions of cognitive information processing¹⁰. The segmentation supports a complex grasp of the content as more than a simple sequence of facts; the objectification contributes new

⁹ Goody [1986] and the related literature cited in Chapter 5 is once more relevant here.

¹⁰ Cf, for example, Miller [1967].

modes to the cognitions which result from understanding the text. In the Euclid text, the label [C.N.2] not only refers the reader to an earlier item in the text - the common notions - but also thereby gives the justification for the line so labelled. The psychology of justification depends on just such textual links. The labels for the propositions have been used countless times by later texts. These extra-citations give the stratification of the text much greater force because they relate it to the similar stratifications of the cognate texts.

Attitudes of scrutiny, minute comparison, comprehensive examination of all relevant evidence, and so on are encouraged by discourse formation. This has been important in the development of various traditions of literate culture, such as law and religion, but particularly for mathematics. I will illustrate this in Chapter 12 in my discussion of proof. Scholarly values have been created around these peculiarly textual activities to ensure the social security of the institutions necessary for discourse elaboration, and I will discuss this in the next two chapters.

Chapter 9

Context

§92 Mathematical text and context

The dominant philosophies of mathematics, those which are part of, or at least influenced by, the logico-formalist hegemony, are concerned with the product or content of mathematical science, to the almost total exclusion of the process of its creation, distribution and use. But since the logico-formalist hegemony is at bottom idealist, this concern is not with actual mathematical products, in particular, real mathematical texts, but idealised versions of them, and an ideal notion of mathematics as a structure of immaterial propositions, objects of knowledge eternal and unchanging, which we slowly discover but in no sense create. In the previous five chapters I have described the sign-systems which are used in real mathematical text, and argued that this description reveals some misconceptions in these widely held product-oriented philosophies of mathematics. It has gradually emerged in the discussion that there are specific conditions of the possibility of real mathematical knowledge of the kind we know, mediated by texts. What is emerging is that texts can only be used in a suitable context. In this chapter I shall describe this context of mathematical text. In the following chapter I shall rely on this account of context to identify some shortcomings in the pictures of mathematics drawn by Kitcher, Lakatos and Wilder. These less influential but more realistic philosophies attempt to take some account of mathematical process as well as its product, and thereby show some appreciation of context.

Literally, context refers to the written text before and after a passage under discussion; but since many writers on philosophy of language use it by analogy to refer to the time, speaker and other surrounding circumstances of spoken utterances, I propose to use it more widely still for all the relevant significationary environment of a written text. That context is important to understanding an utterance, is not always very readily admitted. In philosophy of mathematics dominated by logico-formalist hegemony, it is only rather grudgingly allowed that there is a pragmatic dimension which is found in the grubby human use of

mathematics, while in philosophy and linguistics there has been an equally grudging admission in recent times that isolated invented sentences are not necessarily the ideal data for the study of language. But the working assumption of logico-formalist hegemony is that context can be ignored for logical purposes, or grafted on after syntax and semantics are done. The assumption is that although some expression is necessary for a formal system, no particular one is. Therefore the logician can ignore anything specific to that one which is chosen. This argument underlies Goodman's theory of signsystems, as I showed in Chapter 4, and I pointed out its fallacy. If we recognise that we must attend to the specific characteristics of the signsystems actually used, as determined by the users of the system and the purposes for which it used, then we must also recognise that this entrains attending to their context of use, for sign and context are mutually determinative. I have described some specific features of the mathematical signsystem, its writtenness and the four codes involved. Now I shall describe the context of mathematical text and show how it is determinative of the processes of signification in mathematics.

Sometimes formally inclined philosophers of language recognise that context cannot be ignored completely, and add in a perfunctory way a few parameters to cover the time, place and other indexical features of an utterance. This is always done grudgingly, and assuming that such simple adjustments are all that is necessary to protect formal models for natural language from the obvious charge that most real sentences are not eternal and impersonal statements¹. The attitude of such epicyclists is that context is simply a minor complication, or a few of them. Lewis allows that there might need to be an infinity of "co-ordinates" to cover the various aspects of context - time, place, person, etc etc. But this makes no difference "in principle" (formalists love an excuse for infinite dimensional spaces). The assumption underlying such suggestions is that context is readily confined. Another point of view, and one with which I have more sympathy,

¹ Examples: Lycan [1984]; Lewis [1970], Cresswell [1973: ch8]

is the illimitability of context. For even the simplest real utterances of natural language, even 'The cat is on the mat', there is really an indefinite and ever-present possibility of further specification depending on the direction of interest of an interlocutor. Does 'the cat' refer to a particular individual? Did you mean the exact same cat we saw earlier? Did you mean cat as animal, or cat as visual shape? etc etc. (Quinean arguments can be adumbrated here.) These indefinitenesses are aspects of context, in that we expect them to settled or settleable "by the context" when we are confronted by such an utterance. Sometimes, for example when the utterance is of great emotional importance, such further determinations are made through a long process of discussion with no hope of finality (Freudian "interminable" analysis for example). The need to take context seriously in linguistics, for reasons like this, has been well argued by Brown and Yule². However, philosophers interested in mathematics and science tend to regard these problems as not really important to their conception of language, because in mathematics and science context is deliberately reduced as much as possible. Or so the positivist orthodoxy has it. Actually this is an illusion, as I think my examination of mathematical signification shows: there is a very specific context of mathematical utterances, which is made all the more important because the overt signs of indexicality have been removed from the statements of mathematics, and incorporated into other aspects of the mathematical process. These aspects are basically three.

§93 The trace of praxis

The conditions of possibility of the creation, distribution and use of mathematical knowledge are the conditions for the successful employment of mathematical communications, that is for the use of mathematical text. They are felicity conditions for effecting text acts, in particular mathematical text acts. These conditions are all aspects of the necessity for such utterances to be made in a

² Brown and Yule [1983: ch2]

particular kind of context. The mathematical context is the threefold trace of past mathematical praxis, consisting of the *Archive*, the *readership of mathematics* and the *mathematised world*. The Archive, preserved text, is presupposed by all mathematical texts; paragraphy is constitutive of the Archive, as I have pointed out in the previous chapter. The readership of mathematics is a community of people able to understand mathematical texts. Readership is constituted through didactic text and relies on social stratification. The material environment of mathematics, the mathematised world, is textually formed; diagrams are constitutive of the mathematised world. I will show by example how the context informs mathematical understanding, and how mathematical understanding contributes to the reproduction of context. Thus mathematics underlies, presupposes and reproduces record-based, mathematised, stratified social praxis.

Speech usually has as context its present human environment and some recent past *speech*. I have introduced the mathematical Archive already in my discussion of Paragraphy in the previous chapter. It is part of the global Archive created by the characteristic of written signification, that it persists beyond the occasion of utterance. In claiming that text also requires a community of 'text users', I mean that it is not possible to make use of texts without being aware that one is a member of a literate community in which many texts are to be found, and which the relevant persons are capable of producing and consuming³. It no more makes sense to think of a single text user in an illiterate culture than of a single English speaker in China⁴. Moreover a text user has a conceptual endowment which is massively changed from that of an illiterate - that's one reason why education is difficult. One can hardly suppose that readers and writers are simply speakers of a slightly odd dialect. Yet virtually all discussion of the philosophy of language treats speech and written text as interchangeable, as though it is *obvious* that nothing relevant to the possibilities of communication intervenes.

³Recall that I use 'a text' for an item of written discourse. By 'text' I mean the mass of such items in general.

⁴ An active one, I mean. Robinson Crusoe is always possible.

By mathematised artefacts I mean human constructions other than texts themselves, like clocks and checkbooks, whose construction relies on the use of mathematical text and leaves artefacts structured like mathematical text. They are not texts, because they are not made to be signs - not presented as semiotic tokens. Rather, they are made for other kinds of use. Such use conforms in some way to the text underlying it; text and context are mutually reinforcing. And very often, as with a clock, although the whole structure does not form a sign, its use may rely on its producing signs which must be read in what may well be a new signsystem - though one derived from pre-existing ones. There may be no production of signs, for example an electric torch has no such informed output; but of course that is its point - we want white light, i.e. visual noise, from a torch, not a message. The context of speech, the human environment has for most of human history been a cultural one and not merely natural, but only recently has it been dominated by material objects which are its own product. These are almost entirely what I call mathematised.

These three facets of the context of mathematical text are *traces of past praxis*. Each of the three is an enormous reservoir of items - traces - which were formed using patterns of mathematical text and endure through being inscribed like writing, or other relatively permanent formations. The processes of social action of which mathematics forms a dominating strand in our world are what I refer to as praxis; it is in praxis that context is both formed and informs. In the rest of this chapter I shall try to give a simple description and illustration of these three facets of the context in which mathematics is practised, the Archive, the mathematically literate community and the mathematised world. In the next chapter I will develop further the notion of mathematical praxis.

§94 The Archive is presupposed by all mathematical texts

In this section I enlarge on the point made in Chapter 8, that all texts are grounded in the Archive. All text presupposes other text as part of its context;

although it is equally true that all speech presupposes other speech, there is a difference here between speech and text. The difference is that the text which is part of the context of a textual utterance can remain present while they are both considered, but the speech which may be part of the context of a spoken utterance cannot. **Text remains.** Although some recent speech can form part of the context of mathematical text, as in a discussion, mostly it does not; but recent other mathematical text does. "Recent" may mean recently seen by the user, or recently put in the Archives and therefore available to the user. The former kind is dominant in learning situations, the latter in the creation of new mathematics; but both play a role in most uses.

Pre-existing (mathematical) text as context for a mathematical text constitutes the mathematical Archive, an enormous resource which no-one knows very well. But any reader of a mathematical text must have read other text and must know some things about the Archive - though most know only a little⁵. Most texts, as I have indicated in the previous chapter, make use of paragraphy to orient themselves within the Archive and connect themselves to it. Understanding the text presupposes the ability to use the paragraphic system to place and process the text being read, and to follow the indicated connections to other texts. The Archive is presupposed also in that diagrammatic conventions or special terminology may be used with which the reader must be familiar, or tables and charts may be called on as resources to supplement the text (as in CS174). In fact the Archive is a gigantic resource which a text may call on in a myriad of ways. The way the Archive is used is accordingly a very large topic, some aspects of which have been touched on in the preceding chapter, where I argued that the Archive is constituted through the paragraphic signsystem. Here I wish to simply raise three issues. Who has access to what in the Archive? How do they get it? And why? Those who know most

⁵ How can a person ever read their first mathematical text if my last claim is true? There is no problem here, because the boundary between mathematical text and non-mathematical text is a vague one, as is that between reading and other semiotic activity. I shall discuss vagueness in Chapter 11, so I do not pursue this point now.

mathematics know most about the Archive - where to find things, what they might find, how it is all arranged. They have the most extensive text context when they are writing or reading mathematical text. "To them that hath shall be given..." applies here too. The nature of the Archive and the construction of the mathematical reading public are interrelated and mutually reproducing. Because there are a variety of skills involved in making use of a text, in that it may depend on the Archive in a variety of ways, readers may be more or less well placed to deal with it. Because there are a variety of readers for mathematical text, anyone composing a mathematical text must make a myriad of subtle and usually inexplicit choices about the reader and what is being presupposed about her. Sometimes some of the more obvious such assumptions are collected together in the preface of a mathematical monograph, but the ability to read between the lines of such an introductory text is itself a skill which is very variously possessed, to say nothing of the varying skill of writers in stating their presumptions accurately.

§95 Contents of the mathematical Archive

Mathematical text is very widely dispersed in our culture; one tends to think of the great libraries of the scholarly world as the essential location of mathematical text. But this is wrong: actually most mathematical text, quantitatively, is probably in schoolchildren's books, their "texts" and exercises. This contrary exaggeration is intended to point out that where mathematical text is seen depends on what we count as mathematical text. My criterion is the presence in the item of the four signsystems of mathematical text. By this criterion, a very wide range of items are mathematical text. Ruling schoolwork out on the grounds that it is only practice, or not original, or whatever, would be to make a value judgment on a very obscure basis. Such a distinction might be drawn perhaps, but it should be formulable, and will be comprehensible, only *after* we have a view of the whole Archive. So mathematical text is to be found in home and business, laboratory and parliament, village and metropolis, schoolbook and scholarly library; in books

and journals, newspapers and magazines, textbooks and TV shows, electricians' vans and electoral offices. I don't believe that any kind of random sample of mathematical text has ever been attempted, indeed I don't think anyone has even tried to conceptualise the sampling frame. It's time we did.

Of course, it might be said, this could be done, but what is the point? We know that all this writing is out there, but surely "the mathematics itself" is best represented by its definitive formulation in the current monographs and periodicals in a University Library. Yet how do we know what is "the mathematics itself" before we have concluded our enquiry? How many mute inglorious Ramanujan's manuscripts are kicking about in third world slums⁶, or unknown Galois' prize essays lost in inefficient schooling systems⁷? And isn't there more than a hint of platonism in the idea of "the mathematics itself"? More than a hint, I'd say: this is the implicit logico-formalist hegemony view, the idea of the "natural" content of a mathematical sign, which one finds at every turn. We should resist an implicit identification of mathematics with what mathematicians know, and identification of *that* with what is "represented" by signs in the pure literature. There is a continuum of people knowing more or less mathematics from André Weil to young children, and a continuum of what they know from very "pure" formulations like the prime number theorem to the very "applied" praxis of check-out girls, from people who can barely recognise the number on the bus to the people inventing new theorems about nonlinear programming algorithms used by still other people to schedule the buses. The written items generally thought of as the mathematical explanandum are much more various than is commonly recognised - there are within pure mathematics original memoirs and school text-books, popular surveys and telegraphic reviews, conference proceedings and encyclopedic conspectuses, classics of long ago and encomia for the Fields medallists. Yet mathematics would be no more than a game like chess if it had only

⁶Cf Hardy [1967: 30ff]

⁷ Cf Boyer [1968: 638ff]

a self-contained pure literature. There is also the enormous and enormously varied applied literature, which includes not only what one naturally thinks of here, that is to say the literature of science, but also the equally longstanding and historically determinative use of mathematics in government, in finance and taxation, in war and administration, in engineering and architecture and even (even!) in the arts.

So we should not allow the contents of the Archive to be limited to pure mathematics. Indeed I think we can expect no sharp boundary around it; there are *none* outside the realms mathematics treats of. In the real world of which mathematics as a human phenomenon is part, all boundaries are vague. (I repeat that it is mathematics as a cognitive phenomenon which needs understanding, not the contents of mathematical theories.) Proofs, readyreckoners, program specifications, financial statements, survey reports in the press, chess studies, knitting patterns, the tote board, the tender for constructing a new parliament house, the operation manual of its security staff, there is no end to the Archive! A random walk through a University library shows that a very large proportion of the items held are mathematical in the sense I am interested in, that is to say they contain and deploy clearly mathematical terms and concepts in pursuing human activities clearly marked through and through by mathematical thought. Even if one ignored all these texts, which partly use mathematics and therefore form part of the Archive, even the "pure" literature is vast and variegated, as a glance at Mathematical Reviews and the shelves of a good library show. The Archive has expanded exponentially over time and takes on ever new forms and convolutions. Presently we are going through a profound modification of our archival technology thanks to the electronic microprocessor, the prediction of whose effects would be as foolish as might have been Gutenberg's guesses about the consequences of printing⁸. Just consider the textual context of all the mathematical signs which one comes across in a month. The Archive is so vast that it is hard to see it - an ironic case of

⁸ See Eisenstein [1979] for some discussion of what the effects really were. Profound, slow, multifarious, and surprising conceptual evolutions have been rooted in print.

not seeing the textual wood for the trees out of which it has been constructed. Nevertheless it is worth distinguishing the Archive within the mathematised world, for it is all *text*. The mathematised world includes this Archive of course, but in the next section I will point out the enormous extent to which it includes *other* non-text items too.

§96 The material environment of mathematics is largely mathematised

I am arguing in this thesis that a workable semiotics for mathematics, a desideratum raised by consideration of the various existing philosophies of mathematics, requires an account of mathematics as one complex system of signification among others⁹ in the larger human arena, systems which must also be taken into account. To understand mathematics we must be able to see it in use, and this demands that we include in our description of the explanandum the usings and all the items which are involved therein, not just the propositions being deployed. This is what leads us to consider the second facet of context, *the mathematised world*. I will try to point out just how all-pervading is the mathematical structuring of the artefacts that surround us. (I will deal with this before the readership, which might be thought to come second "logically", for reasons which will become clear shortly.)

Mathematics as a human phenomenon is both very generalised and variegated, and also quite specific. Mathematics is not merely certain linguistic products (which are usually analysed as though spoken, but are nearly always actually written), it is a whole cornucopia of significant human products bearing the form of mathematical construction such as computers, calculators, cash-registers, meters of enormously many kinds, motors and controllers for them, the

⁹ I don't think that these other systems of signification are all languages, though one must recognise that attempts to understand some of them on the model of language have been illuminating. My claim is the more modest one that signification forms a continuum including language but extending beyond it into these other arenas of life. My inference is that we cannot simply assume that they are irrelevant or parasitic, nor that aspects of our clearly language-based activities like mathematics can be accounted for in isolation from them.

regularised landscape and the city system, money and its economic substructure, the uniformity and repeatability of our clothes and meals, the form and disposition of our houses and vehicles, even the organisation of our cemeteries and crypts. These items are noticeably more regular than "natural" objects of the macroscopic level. We are surrounded by straight lines and circles, and by items which move in trajectories derived from them. It is an interesting exercise in looking at one's environment to note all the straight lines one can see within one's immediate visual field - and how few triangles there are! (There are triangles in some constructions, like bridges, but they are nowhere near as common as one might think.) The relations between the mathematical items and the mathematised ones are complex and various too. Connections may be in the making (templates, plans) or in the operation (clocks) of artefacts, or in their arrangement (books, houses).

Mathematics is very generally imposed on our environment, but is also a very specific element in human life. The world was largely free of it until about ten thousand years ago; most of the world was free of it until much more recently. Much of the population of the underdeveloped world still lives in a relatively unmathematised environment; in our own Western societies there are significant strata of society living largely blind to it. The old, the young, the moronic and the mad take little conscious part in it, as we all took an uncomprehending part in the mathematical natural world till recently; though these fringe-dwellers of our societies are also part of the mathematical world for the rest of us in that we count them and plot their geographic locations and foist mathematicized objects onto them. But even the most mathematical of us have times and arenas in our lives which are mainly mathematics-free. Predominantly bodily activities such as sex and rambling fall into this category - though the Kama Sutra and the package tour are distinctly textualised influences on these hangovers from a more carefree past.

All this specific content of the cultural environment has arisen, I contend, through the deployment of text. A great deal of cultural evidence for this is available in the process whereby its dominion is being extended over the "less-

developed" parts of the world¹⁰. In the next section I will list a few samples of the kind of item I am drawing attention to here; the increasing prevalence of just such items is among the visible signs of the developing integration of the world community.

§97 Contents of the mathematised world

Our mathematised world, then, is full of artefacts structured like and by the use of mathematical text - structured like diagrams and proofs and the graphs of functions and formal systems. Here are a few examples.

Clothing

Clothes are made using patterns to form the separate pieces which are joined together along lines; they have borders, use geometrical forms, are decorated with patterns often involving geometrical figures and usually produced in the cloth by repetitive processes¹¹. The actual manufacture of the materials of clothing usually involves machines of course. Though many items of clothing are unique, most are tokens at a fairly specific level; the trousers I generally wear are copied in respect of material, cut, size etc by pairs worn by many other people around the world. A uniform range of sizes is available. The general type - jeans - are worn by many millions. Most unique features of clothing play on this background uniformity.

Barthes has written a book¹² trying to treat clothing as a semiotic system, and although its pokerfaced seriousness reads at times like a parody of American academicism, the basic points are valid: that clothes convey meaning, because the clothes that may be worn at one time together have a syntagmatic structure, with different items chosen from different paradigmatic ranges (hat, coat, footwear...) Making too much of this leads to absurd pomposity and pretentiousness, but

¹⁰ Some older evidence is collected in the volume Goody [1968].

¹¹ Recently some geometers have turned their attention to the mathematics of possible woven fabrics. Grünbaum and Shepherd, *Tilings and Patterns*, forthcoming.

¹² Barthes [1983], and also [1967], of which it is an application.

making too little of it is what lies behind the culturally empty conception of mathematics which I am opposing. In any case French semiotics (as best represented by Barthes) generally overvalues language as the model for all signsystems, a view which derives from Saussure and has been criticised in the literature¹³. Many of the more outré pronouncements of semiotics could be improved if more semiotic variety were recognised and in particular if the salience of the mathematical text, its fourfold signsystem and its consequences in the formation of artefacts were noticed more.

Housing and the city

Shapes of doors and windows are a very ancient and universal example of my theme; there are also pillars and pipes and so on. Note also the great role of geometry and plans in the design, construction and use of houses and public buildings. The ubiquity of construction toys, beginning with simple blocks, is a particularly important intermediary for the reproduction of these systems. The psychological effect of this kind of geometrised second womb are hard to establish but surely of great significance all the same¹⁴.

Transport

Consider the evolution from the track to the road, the Roman road, the canal and the railways - an increasing dominance of straight lines, obviously in the "ways" along which people travel, but less obviously but even more so in the construction of the carriages and the social arrangements which are needed to support these transport systems. Timetables, rosters for staff, enormous quantities of identical spare parts, many of them wheels of one sort or another and so on. Some mathematical aspects of transport systems have been the subject of a great

¹³ Eg. Sless [1986]

¹⁴ Cf the writing of Mumford, e.g [1961]

deal of study, particularly in what has come to be called Operations Research since World War II.

Motors

Motors are very prevalent in our environment. Their design, construction, operation and form are all clearly mathematised. Notice that many mathematical theories are invented to describe the behaviour of artefacts - beginning with the stretched string perhaps; but it is particularly evident in the mathematics of the Industrial Revolution, the mathematics of steam and machine and clockwork. So the "puzzle" for which Wigner is famous and which many still find puzzling - "The unreasonable effectiveness of mathematics in science and engineering"- seems to me no puzzle at all if the context of mathematics is not ignored.

A large volume could be written expanding on these sketchy remarks and making similar ones for information systems (stratification of education and employment; IQ, grading and social location; organisation of libraries, the telephone system etc); public works (roads, armies, sewerage etc); capital equipment (machines that make machines and how they are made, etc). Perhaps the preceding is enough to point out the massive extent to which the world in which mathematics is conducted is a world which has been formed using the products of mathematics, to make material products of mathematical form, which in turn constantly influence everything we think and do, in particular prosecuting mathematics.

§98 Diagrams are constitutive of the mathematised world

It is impossible to overestimate the extent to which the use of diagrams is involved in the creation, use and reproduction of the mathematised world. Plans, maps, templates etc are not just used to "state" spatial relations, they are used for *making* things. And the things which are made bear the impress of these semiotic tools in their form: they are mathematically formed, *mathematised* as I say. A

house evinces its origin in plans by the enduring form it has, which can be found repeated elsewhere, in other buildings and plans.

Here is an example that came to hand as I was writing this section: I bought a tiny radio and with it came a circuit diagram (exhibit #9). This particular text is probably going to have just one use - to become an example in this thesis; other copies may be used in fixing faulty radios, though I doubt it. But the point is that the item which it accompanied is formed as the diagram shows - it has just that pattern of connections which the diagram says it has. This iconic relation holds because the radio has been made using machines and processes which followed plans just like the diagram in relevant respects, to guide the process of construction. Moreover the design was arrived at by an enormously textual process involving similar earlier designs, text on the performance of similar artefacts, and calculations about these performances as well as the economics of production, and so on. The radio is full of straight lines and connections and its performance is full of sine waves and amplification factors. Not only is it statically formed like and through text items, its dynamic behaviour is also described by further textual items. Similar remarks could be made about a great variety of other common objects in the present environment.

The role of diagrams in all this would be hard to overestimate, yet it could hardly be more slightly regarded in the literature of the philosophy of mathematics. Goodman is quite right when he says: "scientists and philosophers have on the whole taken diagrams for granted"¹⁵ and the one exception he cites (James Clerk Maxwell in the 11th edition of Encyclopedia Brittanica) is the more disappointingly narrow in its scope and interest when tracked down.

§99 The third facet of the context of mathematical text is a readership

Now I will turn to the third facet of the mathematical context, the readership. Obviously texts need readers just as speech needs hearers. I claim that

¹⁵ Goodman [1981: 171]

no reader can be conceptualised as a speaker of slightly unusual language; writing is not a translation of speech. In presupposing readers, texts presuppose the processes of forming readers too. And to become *mathematical* readers is a more special process still. This process is one which requires, experience indicates, a large, explicit and directed effort - education. Education must be carried out by educators and educating requires considerable material resources. To educate people to read mathematics requires (at a minimum) competent readers as teachers and suitable texts as practice materials. All our experience shows that these didactic texts need to be very carefully constructed, or selected, to lead the speaker gradually into the mathematical reader's world. In particular mathematical texts of gradually increasing abstraction and sophistication must be encountered - and therefore must be available. But the texts must also be encountered in a context which gives sense to their contents. So they must relate their mathematical content to the learner's known world, and show how the concepts which they introduce have application there. This presumes that those concepts do have such application: that the learner is learning in the mathematised world. Part of the difficulty found by culturally deprived groups, such as the indigenous Australians, in learning mathematics is the fact they they do *not* live in that world. Many "textbooks" are quite appalling in the inadequate relation of their content to the mathematised world. A lot of mathematics that is supposed to be, simply *is not* learnt.

Specific texts are directed towards specific readers. Many mathematical texts make some attempt to indicate the intended readership at the outset, though this is probably often vitiated by the bizarre assumptions current¹⁶ about what it takes to read a book. Thus it is a feature of the reading public that it is differentiated in respect of the kinds of texts which it is capable of reading. The idea of reading age glosses over many other ways in which readers differ! This is a fact which cannot be

¹⁶ Many writers have statements equivalent to the claim that all that is needed to follow their text is "mathematical maturity". That is, being a suitable reader! Many also think that by including "chapter zero", prerequisites can be reduced to nil, an absolutely stupid shortsightedness considering how *they* got to be able to write it.

ignored in trying to understand how readers understand texts. A randomly chosen reader would *not* understand a randomly chosen text. The differentiated readership is supported by the differentiation of the locations of texts as well as by a considerable infrastructure of access routes whereby readers and texts get together - librarians and so on.

The formation of a readership is a long, arduous and therefore costly business and the organisation of it on a large scale has considerable implications for the economic and social organisation of a mathematised society. The free time and the resources needed¹⁷ for the instructional process presuppose an economic base considerably above the simple societies beloved of philosophical speculation. In short there seems little prospect of the conditions for the formation of a readership unless the social structure is based in the division of labour and the employment of considerable technology, technology which must be greatly informed by mathematical knowledge, since it is so useful and, by hypothesis, available.

§100 Readership is constituted through didactic text and social stratification

In saying that education relies on suitable text I am not pretending that it is impossible to teach with unsuitable text. Much teaching succeeds to some extent in spite of its resources. Rather I am trying to emphasise the extent to which learning to learn from text¹⁸ is a difficult task and that in order to succeed the learner must not only understand the content of the signs presented but come to understand the systems to which they belong. A large portion of mathematical text is didactic. This has become particularly noticeable as education has become widespread in the community. Becoming a reader means having formed certain specific concepts. The social stratification which I have argued to underlie the formation of a readership for mathematics is of course much more complex than its mathematical facets

¹⁷ The physical production of texts presupposes a considerable technical infrastructure.

¹⁸ Cf Morris and Stewart-Dore [1984], Christiansen [1986].

alone. There is plenty of evidence that the origins of social stratification and mathematics are part of a single complex process¹⁹, but we cannot go into that now. The main point is that all we know about the existence of mathematics supports the contention that it only exists in complex, technological, stratified society.

The social location of readers for mathematical text is determined by the nature of the mathematical readership formation process: a large majority are in the education system. The connection between research and teaching functions in higher education reflects the link between the formation of text and the formation of a readership for them. The production and consumption of mathematical text outside the education system, which is of course also extensive, is prosecuted by "mathematically mature", i.e. mathematically literate, people. Moreover it would be impossible to deny that there is a didactic strand in the extra-mural mathematical praxis in question. Many private research organisations support important scholarly publications²⁰. Most mathematical texts can be and are used to learn from. Whenever new content is learned, new aspects of the mathematical communication system tend to be learned too - beginning with notation and nomenclature. The mathematically mature public is part of the intelligentsia. Insofar as these social roles have economic ramifications, and that extent is considerable, so we must recognise that the reader and the writer of mathematical text have a social position in which their activity occurs. Inevitably then that activity is tied to a structure of interests which cannot be regarded as common to everyone in society, but is on the contrary defined by differentiations centred on the division of labour.

In philosophy of mathematics it is generally accepted as reasonable to abstract from differences between different knowers, an impersonal "we" is sometimes affected, as in Dummett's discussion of what "we" intuitionists know, presuming that knowing may be taken as uniform. But really, there is only a highly

¹⁹ Cf Schmandt-Besserat, Goody and related literature.

²⁰ For example, IBM and Bell Telephone. Or look at the Journal of Finance, largely supported by an investment house.

stratified mélange of partially overlapping communities of knowers, whose understandings of the various parts of the archive are extremely various and should not be assumed to be "in principle" the same. For diversity of uptake is actually essential for change and discovery. If two people never took a sign differently there would never be a new thought.

§101 The mathematical context is product and precondition of mathematics.

It is evident from the foregoing description, that the mathematised world would not exist without mathematics. Whatever the origins of mathematised society, it is hard to deny that the perpetuation of it relies on mathematics to a considerable extent. Not only is it obviously massively implicated in the physical reproduction and development of the material infrastructure, it is less obviously but equally massively involved in the reproduction of the social stratification which goes with it: the salience of mathematics in education and in job differentiation is univerrally admitted (and generally disliked). Thus mathematics underlies stratification in making possible the material and organisational infrastructure of such a society (accounting, engineering); presupposes it in relying on a constantly replenished readership which needs such a context for its formation; and reproduces it in the ways in which it is involved in the ongoing quotidian renewal of social and economic relations.

The reverse is true too. There would be no mathematics, in the sense which I am elaborating in this work, without the mathematised world to support its production, consumption and preservation. Whether in some attenuated sense the things we know of in mathematics would still exist or still be true if the mathematised world vanished is a question which is of limited interest (but see Part III: I shall answer that they would, but only in an unexciting sense which excludes platonism). The mathematised world is a precondition of mathematics for three main reasons. First, the signs of mathematics are given sense through praxis which involves the context, and that sense involves aspects and elements of the

mathematised world. So readers for mathematical text, and therefore producers of it (writers are readers first), could not be formed in a world which was not mathematised. Second, the point of mathematical activity would be lost if it never issued in formative action in the world. Third, there simply *could not be* texts without the social infrastructure I have identified. Texts require writers and tools for writing. The manufacture of the tools (paper etc) demands a certain kind of society.

An example of the way the mathematised world is needed to give sense to mathematical signs is the gradual acceptance of complex numbers. The crucial need in making these mysterious entities acceptable to the mathematical public was not, contrary to common opinion, the Argand diagram, but rather the possibility of finding physical phenomena, and then making objects which rely on them, for which complex numbers are a good model. Why else was the "Argand diagram" invented half a dozen times over half a century²¹? The theory of electro-magnetism made complex numbers acceptable, and it did so because it was a theory with applications: its terms acquire sense from physical objects, objects made in accordance with the theory. Some individuals acquire knowledge of complex numbers in this grounded way too (engineers mostly) - but most people, I believe, never really understand complex numbers simply because they have no use for them - just as many people don't really understand or believe in irrationals. For most people the formalist interpretation of mathematics is literally correct for complex arithmetic - it is just a game they play with marks on paper. They learn to play it (we have good algorithms) but it remains *empty*.

When I say that the point of mathematical activity would be lost without applications, I do not at all mean that pure mathematics is meaningless activity. Nor do I mean that the point of work in number theory is a disguised form of the point of applied mathematics. Rather I mean that if there were no applications, then the activity of number theoretic research would lose its present point. What

²¹ Cf Boyer [1968].

exactly that is, is not completely clear, but the generally accepted idea is that it is connected with "discovering truths" and "solving problems". These would have to be re-interpreted to something like what similar statements mean about chess, if mathematics had no applications. We say such things about chess, and they make sense, but a different kind of sense. The point of number theoretic research is "important" because it has connections, no matter how distant or tenuous, with uses of numbers which are *clearly* important: taxation, birthdays, shopping etc²². Abstruse purities also have a habit of turning out to be useful outside mathematics, because they are connected with general and basic problems of mathematics which already have a clear sense in applications, even if for the working researcher they are conceptualised as highly specific and "pure" problems in a highly refined subfield. Narrowly focussed vision is necessary for tying individual threads - but one walks on the whole carpet.

The physical necessity for mathematised society to support the *physical* making of texts at all is clear enough as long as we don't limit mathematisation to the use of explicitly formalised mathematics, that is Greek mathematics and after. Pre-Greek mathematics in the form of mensuration, calendrics, accounting is undoubtedly part and parcel of the complex process of civilisation within which technologies like writing and printing arise. This is illustrated by the studies by Goody, Ong, Schmandt-Besserat and others which have already been cited.

§102 How context affects the understanding of mathematical text.

In the next chapter I shall discuss several accounts of mathematics whose strengths and weaknesses derive from their focus on mathematical process, which involves some appreciation of mathematical context. I shall discuss examples dealt with at length in the writings of Lakatos, Kitcher, and Wilder. But first I will

²² In any case, most mathematics is directed towards application, though towards application within mathematics!

use here a simple example, that of Euclid's algorithm (Elements VII.2) for finding the greatest common factor of two numbers, to show how context changes the way a text is used and understood. We cannot proceed by describing the "meaning" of the text, as though it were possible to state the content of a text in a neutral way which captures its "essence", and then add the "garnish" of context. We can only proceed by contrast. So I shall compare the original and the present contexts and understandings of texts about this idea.

The method for determining the "greatest common measure" of two numbers, described in Euclid Book VII proposition 2, (exhibit #12), is usually known as Euclid's algorithm. In its original setting it is a construction, a procedure for producing a number from two others. Let us consider how understanding this by Euclid's original readers depended on context. First, the dependence on other text. Possibly the Elements may have been used as an educated Athenian's introduction to mathematics, so that we might suppose the reader to have read no other mathematical text. Still, successfully reading this one (i.e. VII.2) does have textual presuppositions. The definitions which appear at the start of book VII must have been read, and perhaps some earlier parts too. These definitions are concerned with giving an organised sense to the terms 'number', 'measure' and so on. These terms must have been read previously in other, perhaps non-mathematical, text. Or suppose not - the reader *might* have been taught to read with the Elements too! Still *Euclid* must have read other texts to be in a position to write this one! And of course we know that it is in fact a distillation of many earlier presentations of elementary mathematics.

Second, we should consider the material context of the Athenian reader of Euclid. It is highly significant that the term 'measure' is used in Euclid's presentation of the theory of divisibility, though it seems to us quite irrelevant. It shows the close connection which subsisted for the Greek mathematical reader and writer between this mathematics and the repeated application of a unit to form a multitude, as in the measurement of length. Indeed the definition of 'number' given

by Euclid is just that: a multitude composed of units. This is a highly concrete and highly alien conception of number to the modern mathematical mind. In Greek culture there were diverse applications of number, to be sure, but a few were paramount and there were in any case very few compared with our own. The measurement of land and building, and the counting of possessions and persons were the activities which grounded mathematical terms in praxis²³. But despite this grounding being specific in its narrow and particular focus compared with our own, it is nevertheless *there*: Someone from a culture essentially lacking in mathematics (say Pitjantjatjara) could not make sense of Euclid's definitions at all, for lack of any way of interpreting them. What can 'unit' mean to someone who has never engaged in a practice which could be described as counting, composing units? In some such cultures there is not even a single sequence of content-neutral number terms, and no terms more specific than "many" for groups of more than a few (more than 5 for Pitjantjatjara, apparently).

The third facet of the context is the readership, and of course I have already adverted to it in my discussion of Archive and inscribed world. The three facets of context are mutually interdependent and reproductive. Nevertheless, let us consider briefly the possible readership for the Elements as originally composed. The original readers of texts like Euclid's were a leisured class, dependent on the ancient slave economy. This both provided the surplus for their opportunity to study mathematics and was reproduced by the political organisation and engineering capacities which a mathematicised culture made possible. In the social structures of antiquity, the readership for mathematical text is clearly descended from the scribal classes of the hydraulic civilisations of Mesopotamia and Egypt, which were formed at the very source of the urban, stratified way of life.

Now let us turn to the understanding of this text by a modern reader such as myself. We must begin with the fact that we cannot be talking about the same text token: we have no surviving manuscript from Euclid's time. In fact we have no copy

²³ Cf Szabo [1978], Klein [1968].

from anywhere near Euclid's time. We have only variously doubtful copies which are not of Euclid "bare" but include "corrections", glosses, commentary etc. So the simple presumption that we are talking about a single text is problematic already. In any case, there is no possibility that my understanding of a supposed original text could be the same as that of some personal student of Euclid's, for the simple reason that I cannot read Ancient Greek. I am not part of the readership. I could learn to read Ancient Greek, but this would not put me into the readership, since the process of acquiring literacy in that now dead language would be grossly different from anyone's in 250 BC. In any case I already know a lot about the algorithm which has been discovered since then. For example I have written programs to implement it in computers, which surely differentiates my understanding of it from that of any original reader's of Euclid.

If we consider the Archive of Euclid's original readers and my own, great contrasts are evident. For example it is impossible for one to read a modern version of Euclid's text without a context of commentary and translation. My own copy is Heath's edition in three volumes - in which Euclid's words start on page 152. The enormous literature which has grown up around the Elements, and the history of its immense prestige and influence in Western thought can hardly be ignored totally by someone reading it now. All that other text is present as context - particularly for me, thinking about these things. Whereas for some original reader of Euclid's the Archive consisted of texts of which we mostly have no trace and little conception, as well as being much smaller and closer to the beginnings of the whole world Archive in absolute terms, for textual accumulation began slowly and weakly in the ancient world.

For a different perspective on the context of this text, consider the treatment of the "same mathematics" in Hardy and Wright's *Introduction to the Theory of Numbers*. It occurs²⁴ after extensive developments of a sophistication way beyond Euclid's and it is subordinated to the fundamental theorem of arithmetic, which

²⁴ Hardy and Wright [1975: 179ff]

Euclid did not state and according to Hardy and Wright *could not have stated, because he did not have an adequate notation to permit its statement*²⁵. There are two important things to note about this observation of Hardy and Wright. First, it is a specific example of the general point that notation creates new possibilities of thought, the Klein point about notation²⁶ which I have not yet discussed in detail. Second, and this is the point relevant to our present comparison, the text in question occupies a different textual location vis-à-vis other mathematics now. Now it follows from the fundamental theorem in short order; for Euclid's first readers, it had no such close cognitive relation. Indeed the question exactly what "it" is that is the same here is a difficult question, I think. Hardy and Wright are discussing arithmetic in various algebraic extension fields of the rationals, so that the concept of dividing one number by another, which they have as basic, is very different from the Greek idea of measuring multitudes with units. Of course there are connections - but they are complex mathematical connections, not simple ideal identities.

I have mentioned computers: these are merely the most obvious item in the mathematised furniture surrounding me which Euclid's original readers did not have. In fact, there are a vast variety of items which depend on divisibility and the theory of Euclid book VII for their construction in my environment: money, the scheduling of my time, multiplexed communications, gears and printed circuit boards²⁷ etc. These devices and applications differentiate my understanding of division from that of the original Euclideans. Thus *all* the terms involved in Euclid's algorithm are understood very differently by me and by Euclid, because of the vastly changed contexts we live in: 'number' 'subtraction', 'division', even 'repeated' are all terms whose content is different now.

§103 The myth of the common content (again)

²⁵ Hardy and Wright [1975: 182]

²⁶ Cf Chapter 1, §11 above.

²⁷ Cf Schroeder [1986: 22ff].

Yet is there no sense in which there is something common in Euclid's and Hardy's texts, a "common core" of mathematical content which has been extended over the centuries but endures? After all, we are still using the same name - Euclid's algorithm - for what is surely the same thing. There is strong seduction in this idea, but it is mistaken. Indeed it is the simplicity of this idea which is its fault. Signification is not, generally, naming and indeed the image of naming that we have is itself overly simple. The one term/one referent conception of signification is a persistent temptation whenever we find no obviously different referents to help us resist it. Even in cases where we seem to have one "continuant", as in the ship of Theseus "puzzle", the metaphysical impulse to identify some timeless unchangeable item as the true reality prevents good sense prevailing.²⁸ In the case of mathematics where physical referents are not concerned the temptation is constant.

Perhaps then I should repeat the basic argument against the idea of common content. It is this: the only way to make out that two texts have a common content is to state it in a third text. *Principia Mathematica* is an attempt to do this for large parts of mathematics. Such a third text is either a pure formalism, uninterpreted, in which case it expresses no content at all; or it is interpreted and therefore itself context-dependent and the question arises about the identity or otherwise of its content with that of the two texts which it models. Its grounding, its *launch* as I call it in Chapter 7 above, achieves this and provides the evidence for it. I would rather say that it is simply a third text, which may illuminate the other two. And the subsequent history of *Principia Mathematica* I believe to support this view - it has hardly been much used in place of Euclid or Hardy! To say that a certain N-sign in *Principia Mathematica* states the essence of Euclid's algorithm is to abstract from the context not only of the two signs I have been discussing in this illustration, but also of the context of *Principia Mathematica* itself. For it is actually impossible for anyone to understand the sign in *Principia Mathematica* who does not approach it

²⁸ See Heller [1987] for a recent instalment in this sorry saga, and Wiggins [1980] for the classical morass.

through a process of mathematical maturing, ending with reading large parts of *Principia Mathematica*, including all the preliminary notational explanations. If recognised properly, this process would itself be a complete confirmation of what I have been arguing in this section. There many jokes about the 12 people who have read all of *Principia Mathematica* and the difficulty of deciding who was the one who understood it!

In this chapter I have explained how I conceive of the context in which mathematical text is produced and used. In the next, I shall consider some philosophies of mathematics which I regard as superior to those on which I have so far commented, which are all dominated by idealised conceptions of the products of mathematics. The superiority lies in their attention to the processes involved in mathematics, but I shall be arguing that their understandings of context are inadequate and that this reduces their value considerably.

Chapter 10

Praxis

§104 The concept of mathematical praxis

A variety of writing on philosophy of mathematics recently collected together by Tymoczko¹ under the title *New Directions in the Philosophy of Mathematics* begins with a section of pieces challenging the foundational "paradigm" which is part of the logico-formalist hegemony which I identified in Chapter 2. The rest of the collection is devoted to mathematical practice. By 'mathematical practice', Tymoczko means the activity of

*"mathematicians and those who use mathematics. If we look at mathematics without prejudice, many features will stand out as relevant that were ignored by the foundationalists: informal proofs, historical development, the possibility of mathematical error, mathematical explanation (in contrast to proofs), communication among mathematicians, the use of computers in modern mathematics"*²

It transpires, and this is confirmed by the section called "What is mathematical practice?", that Tymoczko really means by the term 'mathematical practice', mathematics as it actually is, rather than as idealised by foundationalists. But there are several things involved here. There is mathematical activity, what individual people actually do in mathematics. There is the social, collective historical process or processes of mathematical change. There are the individual actual, as opposed to ideal, results of this activity and these processes, such as texts and mathematised objects. There is the structured social formation of these artefacts collectively together with the institutions which support those processes. Any or all of these might be called mathematical practice. I propose to use the term **mathematical praxis** to refer to the dynamic social formation composed of people working on mathematics and the products of this activity. I do not exclude abstract entities such as learned societies and theories, but I emphasise that the people and their self-created material environment are primary.

The few writers who have used the term 'mathematical practice' have done so without a great deal of attention to just what is comprised: the important thing

¹ Tymoczko [1986]

² Tymoczko [1986: xvi]

for them is to contrast practice with theory³, the process of mathematics as opposed to its product. Previous chapters have made clear that current ideas about the product of mathematics are inadequate, and I believe the same applies to the process, the practice. I have chosen to use the term 'praxis' for my own conception in order to contrast it with these current uses. I do not intend any specifically marxist implication, but to signal the material emphasis which I believe distinguishes my conception from others.

The writers whom I shall discuss in this chapter have all made a significant contribution to creating a concept of mathematics as a social practice. Lakatos makes vivid the fact that mathematics is a process in time, by providing us with detailed reconstructions of the process of conceptual development and some important suggestions about the nature of that process. But he does not relate this process to the other social processes with which it is inextricably intertwined. Wilder's writings do make this connection, by taking the mathematical system as one amongst other cultural systems, but his attempt to remain neutral on classical philosophical issues leaves his account with unsatisfactory, implicitly idealist answers. Kitcher's account of mathematical practice, while attempting to assimilate it to scientific practice along Quinean lines, in order to provide an empiricist answer to the philosophical questions, also has unacceptable idealist aspects. Moreover its philosophy of language is inadequate. In fact this is true of all the writing in these new directions. Significantly, Tymoczko's collection contains nothing by Wittgenstein. In the last Part of this thesis I shall return to the question of language and argue that it is to Wittgenstein we must look for a serious attempt to escape from the positivist conception of language and for the development of a conception of mathematics which coheres with an alternative one. His philosophy of mathematics never achieved a coherent formulation, and the expositions we have are marred by his mistaken strategy of focussing primarily on the simplest mathematics. His discussion lacks any connection to the historical and social

³ As in Wang [1974: ch7]. (Some of this was written before 1960.)

context of mathematics as it has developed, but his conception of *language as praxis* is a necessary one for us since it makes it possible for us to accept the variety in linguistic function which I have been at pains to demonstrate in earlier chapters. I shall draw on and improve his account of necessity, to sketch an account of mathematical knowledge which shows why it is peculiar in the ways which were discussed in Part 1.

None of these writers shows much sensitivity, either, to the extremely characteristic material context of mathematical praxis which I have pointed out in the preceding chapter. And none shows any interest in the involvement of mathematics with social structure and control, with power and its distribution, with the pragmatic presuppositions of the use of mathematics. Later in this chapter I shall take up this discussion by introducing the concept of mathematical work. In the next chapter I will return to the question of the social presuppositions of the use of mathematics.

§105 Lakatos on proofs and refutations

Lakatos' tour de force *Proofs and Refutations*⁴ has been curiously infertile; strange that such a wealth of ideas should have such a little impact on thinking about mathematics. Partly because of its stylistic inimitability perhaps; partly also no doubt because Lakatos himself turned away to the philosophy of science, never to return. Even the recent book by Kitcher⁵, which I will discuss shortly, makes only passing reference to Lakatos. Tymoczko's collection has two essays of Lakatos', but no selection from *Proofs and Refutations*, which seems a little strange. This lack of impact is not to be explained by any fundamental poverty in the ideas put forward in *Proofs and Refutations*; On the contrary I believe that there are a number of profound ideas, and some considerable overlap with ideas that can be mined from Wittgenstein's writings on philosophy of mathematics, which I come to in Part III.

⁴ Lakatos [1974]

⁵ Kitcher [1983]. Hallett [1979], [1980] is one attempt to follow up Lakatos' ideas.

Lakatos' position has obvious debts to Popper's methodology of science, but there is also a considerable whiff of Hegelianism in it too. Perhaps the importance of Lakatos' philosophy of mathematics has been masked by the waxing and waning of the Popperian philosophy of science, to which it sometimes seems to be a footnote.

Lakatos' view about change in mathematics is that it consists in the improvement of conjectures and concepts under the impact of the mathematical criticism, which is governed by the logic of proofs and refutations. His most detailed example concerns the concept of polyhedron and the Euler formula. The basic sequence of events is⁶ primitive conjecture ($E-V+F = 2$), proof, global and local counterexamples, proof improvement by lemma-incorporation to form a new conjecture. Global counter-examples are cases where the conjecture is false, which are dealt with by qualification of the conjecture; local counter-examples are those for which the proof fails, which require a new proof to be invented. During the process of development, there are several strategies which workers adopt towards the mutual adjustment of concepts and examples: surrender ("the conjecture was wrong") monster-barring ("that's not a polyhedron"), monster-adjustment ("that polyhedron has hidden sides"), and lemma-incorporation ("polyhedra can be planarised after a face is removed"). *The method of lemma-incorporation is the method which leads to mathematical knowledge.* Lemmas to incorporate are found by proof-analysis, in which that lemma is identified which the difficult examples falsify. When lemmas are incorporated, concepts are *stretched*. Concepts are not merely extended or generalised by shifting the frontier of inclusion, they are replaced - so that even one's problem (from which enquiry starts) may get changed in the process of research. The limit of concept-stretching is the destruction of meaning and truth in formalisation. A theory may reach the stage of formalisation, at which point its theorems are drained of content but are fully certain. However this is only relative to a background theory of logic and the process can break out

⁶ Lakatos [1974:127]

again if logic is put in the melting pot.⁷ This very brief sketch of Lakatos' account cannot do any kind of justice to its vitality and plausibility. Lakatos' method is to present an imaginary dialogue between mathematicians representing various opinions about methodology whose debate constitutes a rationalised reconstruction of the rather more confused real history which he documents in footnotes. He modified this technique in his papers on philosophy of science by abandoning the dialogue but retained that of the two levels, rational reconstruction main text and irrational historical footnotes.

Lakatos' conception of practice is *criticism*, as one would expect of a student of Popper. Notice that although he cites a lot of real texts, his dialogic presentation falls into the trap of allowing that mathematics could indeed be basically spoken discussion⁸. I have already given reasons to deny that. The emphasis on criticism, which is an interesting cross between Popper and Polya in fact, certainly brings back one element of praxis which the logico-formalist concentration of the finished products ignores. And it gives some indications of how the questions about *choices* are to be answered, which I posed against Logicism at the end of Chapter 6. However there are a number of shortcomings which must dampen the enthusiasm which reading Lakatos tends to provoke. I will suggest some in the sections after next.

§ 106 Lakatos on notation

What does Lakatos say which will cast light on the role of notation? Little, explicitly, but perhaps we can extract some views by approaching obliquely.

⁷ [1974:99] In [1974] Lakatos appears to grant something to Popper's [1947] as showing the limits determined by classical logic, but in other places it becomes clear that this was a misguided generosity and that Lakatos was not unaware that logical theory is just as nonultimate as any other (e.g. [1974:123n], [1962:22-23]). The flowering of nonstandard logic since the time of [1974]'s writing in the late 1950s opens this whole issue up, and it is a pity that there is little hint in Lakatos' writings as to how he would react.

⁸ The role of mathematics lectures is interesting. No mathematics lecture is just spoken words. Good ones use speech to clarify written text (written on blackboards, usually).

through Lakatos' views on concept-formation. What is made vivid by Lakatos' presentation in *Proofs and Refutations* is that in mathematical research signs (theorems, definitions, figures of counter-examples etc) are put forward and accepted or rejected. The way in which they are received into the literature determines what is happening to the concepts involved. Lakatos implies⁹ that in the motley of mathematics, formalised theories can be simply part of the material which is dealt with in informal (or post-formal) proofs. Thus notation, insofar as it figures in informal mathematics, is part of the subject-matter. For example, the Duality theorem of projective geometry concerns not only lines and points but also theorems about them. But in the mathematics discussed in *Proofs and Refutations*, notation is not mentioned. Yet it is apparent that it plays an *increasing* role in the various formulations of the conjecture. In the early stages, notation plays a minor role; the original $V-E+F = 2$ is little more than abbreviation of a sentence - though we should beware of thinking that the terms in such a sentence are devoid of notational support¹⁰. But later discussion about alternative formulas relies on their notationality for the production of alternative possibilities - the classic use of notation for many writers, letting it "think for us". In the process of giving the Poincaré theory¹¹, the sequence of definitions remains apparently verbal, though the terms 'edge' and so on are really misleading, since they are undefined and do not carry their usual meaning. But eventually we reach the need to define 'boundary'. Since this involves addition modulo 2, it becomes necessary for Lakatos to define it through notation; similarly for the concept of incidence matrices; these are not abstractions from perceptibles but items needing to be defined notationally. The

⁹ In Lakatos [1978b], for example. This paper was not published by Lakatos (though Tymoczko includes it) and is very spotty, but it makes some good points. For example, there is no real problem about how to choose the interesting formal systems: the only ones worth considering arise naturally from informal mathematics and science. He also gives the trivial formalisation as a refutation of formalism. But the real point of formalisation, its value and its ongoing role in mathematical thought, are not really addressed.

¹⁰ Cf Chapter 7, above, on this.

¹¹ This is the final form of the theory of polyhedra which is discussed in Lakatos' dialogue.

proofs which follow are calculations, in which geometrical intuition of polyhedra is replaced by the more powerful purely graphic intuition of notation manipulation (which is still geometrical).

How would Lakatos explain all this? I suggest that it is consonant with Lakatos' position to suggest that formalisation relies on notation, theories being formalised by their presentation in notation which permits the new usages of the terms to deviate as wildly as we will from those of the original terms; that notation thus allows concepts to be weaned from their empirical roots and stretched by their "translation" into these formal systems; and that notation thereby allows for insight to be replaced by calculation in the verification of conjectures (change in the nature of proof from informal to formal). This suggestion is of the nature of a supplement to Lakatos' picture of the progress of a mathematical theory toward formalisation. But it may also enable us to address some of the shortcomings of his position which I indicate in the next section.

§107 Some limitations of Lakatos' position

Lakatos' discussion is wonderfully refreshing after a diet of logico-formalist hegemony, but it has some significant shortcomings which appear when one tries to generalise from the situations he discusses to an overall perspective.

I pointed out already his insufficiently radical attitude to logic (cf note 4). A revealing aspect of this is that although he exemplifies many kinds of criticism in his dialogue, Lakatos does not thematise any other than counterexample. All his conflicts are presented as logically of the kind: do you or don't you accept X as an example of Y? But many of the protests, both those of his disputants and those which he reports from the real mathematicians, concern the value of generalisation or the rationality of a method or the rigidity of logic. He also says that change in the concepts of truth and proof follow those in criticism. The new kind of refutation which marks the 19th century revolution is the acceptance that there is no limit to what is acceptable as a concept-stretch. Any kind of substitution may be acceptable,

depending on its fertility. But the way this mostly seem to come out is that decisions about whether to accept X as a case of Y can have wilder and wilder justifications, whereas it is also suggested that these decisions follow from decisions about the justifications. Lakatos does not make clear whether there is a fixed priority here, or a characteristic one for a given stage of a mathematical theory, or for a given time in history, or just chaos. The overemphasis on the logic of the counter-example is no doubt the influence of Popper, whose philosophy of science could be unkindly satirised as a prolix way of pointing out that counterexamples refute generalisations, but examples don't prove them.

Lakatos does not clearly articulate the real *point* of formalisation, its value and its ongoing role in mathematical thought. Lakatos does not sufficiently consider the implications of the coexistence of the old and new concepts, in particular he does not articulate the relation of the informal and formal formulations of a theory. He does not really remark on the fact that the theories and concepts which he has extracted from history do not replace one another, except as favoured by the avant-garde: the very possibility of his work depends on the continuing presence of these old formulations in the archive. I have argued in earlier chapters against this replacement idea. An interesting illustration of my point is provided by the recent book *Surfaces* by Griffiths¹². This provides an exposition which covers the same theory of which Lakatos has followed the history. Griffiths' aim is to help students who may be relatively uneducated mathematically, or who may have to teach young people. In either case the overriding aim is to develop the reader's three-dimensional intuition. This will not be achieved by an axiomatic approach. Griffiths presents the material informally, along the lines of the earlier stages in Lakatos' discussion. But he also includes in an appendix a "rigorous" version of the same material. This is a wonderful example of the necessary retention of old formulations side-by-side with new ones, and of the explanation for it: different texts are needed for different purposes.

¹² Griffiths [1976]

Lakatos has been criticised for his lack of an explanation for the invention and acceptance of the method of proofs and refutations in the 1840, or of what, if his claim is correct that this happened, is the logic of mathematical discovery in earlier and later times. The latter is simply a demand for more of the same, but the former complaint indicates what I regard as a most serious shortcoming of Lakatos' position. Bloor has suggested¹³ that an explanation is available, in terms of the developments of the social position of the Prussian professoriate at the time. One can easily be sceptical about this, particularly since his argument involves a very crude classification derived from the relativist anthropologist Douglas. But scepticism is only useful as a prophylactic, and it is necessary to recognise that Bloor is asking good questions, even if his particular answers are not good enough. (Indeed his later volume¹⁴ shows some improvement in point of plausibility - significantly, because he has abandoned Douglas for Wittgenstein for his basic notion of praxis.) The point which Bloor has noticed is that mathematics as a practice is pursued by people who have to live. It is necessarily enmeshed with their other practices, and cannot be understood in isolation from them, in particular economic practice. People *are* motivated by curiosity and the desire to know the truth. But this is insufficient as an explanation for the directions research and its reporting take - and ludicrous as an explanation for the 95% of mathematical practice which is not "pure" research.

§108 Wilder's cultural evolutionism

What is good about Wilder's work¹⁵ is the basic plan of recognising mathematics as a *cultural system*, in interaction with others. Where Wilder improves on cultural histories of mathematics such as Kline's¹⁶, is in *attempting* to describe more than just the content of mathematics and its relation to scientific

¹³ Bloor [1978]

¹⁴ Bloor [1983]

¹⁵ Wilder [1968], [1981]

¹⁶ Kline [1975]

ideas and social developments, in recognising that mathematics is a process of interaction of cultural artefacts. The concepts and theories of mathematics are *cultural elements* with a history and a development in specific circumstances; and they form a *cultural system* whose internal relations and interaction with its environment can be described and explained. Wilder's approach is anthropological, in the tradition of Spencer, Kroeber, White and other cultural evolutionists. A cultural element is an item - Wilder's prototypical example is of set of beliefs about the behaviour of a kind of bird - which is shared by individuals. It arises from various individuals' beliefs by a process of communication which traditionalises it through consensus formation. A cultural system consists of many such elements interacting through communication. In such a system there are internal stresses, and acting on it there are also external ones. These lead to selection among the various elements; the system evolves as selection alters its composition.

Put so briefly, the value of Wilder's account is hard to judge. Unfortunately the detailed exposition does not remove this doubt. One good example of the value and the deficiencies of Wilder's perspective is in his discussion of *consolidation*, the process of unification of theories and concepts in different areas of mathematics by a process of generalisation. One of his fifteen examples is that of Descartes' consolidation of geometry and algebra. Fermat made much the same consolidation at about the same time, which is explicable in terms of hereditary or internal stress ("the theory was in the air"). In addition, important practical problems (optics etc) made use of the new techniques, showing that there was also environmental or external stress. The comparison of the various examples is illuminating but one is left feeling doubtful as to whether anything has really been explained. Does it really add anything to call the possibility of applications "environmental stress", or the existence of the needed preconditions for the consolidation "hereditary stress"? Is selection at work in consolidation? It is not mentioned in that chapter at all; rather, diffusion is referred to.

This is objectionable if Wilder is serious about the "evolution" in his framework, but it seems that the tradition to which he belongs regards biological evolution as merely a particular case of evolution, which need not be mimicked in all respects in social evolution. Wilder specifically denies that the idea of evolution of culture arose from that of biological evolution¹⁷. But though true, it is irrelevant since all the explanatory force in the idea of evolution is contained in the biological theory. Actually, the denial which I have just cited comes at the end of a rather confused discussion of the relation between history and evolution, in which Wilder conflates the contrast between a record of events and the events themselves¹⁸, with the contrast between changes in individuals and in forms¹⁹, and also with that between long-term changes long ago and short-term changes in recent times²⁰. It is passages like this in Wilder's work which reinforce one's scepticism about cultural evolutionism. One must suspect that the long neglect of his work derives from the prominence of this important but not fully worked out evolutionary analogy in it²¹. He emphasises the evolutionary character of mathematical processes, but it turns out that there is no real attempt to transfer the explanatory force of evolution theory to mathematics despite talk about 'selection pressure' and the like.

What is missing from Wilder's account is any conception of how the social process is related to the individual events. In real evolutionary theory this is provided by the statistics of variation and natural selection. But Wilder uses 'selection'²² in a way which is ambiguous between conscious choices of individuals and blind processes of elimination at a social level. Really his discussion is only

¹⁷ Wilder [1981: 20]

¹⁸ Ibid, p18

¹⁹ Ibid, p 19

²⁰ Ibid, pp19/20.

²¹ In his more recent formulation of his account, Wilder supplements the evolutionary metaphor with another, that of vectors. But once again, no real use of the explanatory force of the theory of vectors is made, it is just a useful terminology.

²² [1981:51ff]

nominally evolutionary ; for instance, there is no mention of reproduction or variation.²³ And this is because he does not take seriously the material basis of culture. The material context of mathematics which I have elaborated in the previous chapter gets little attention. Although in his definition of cultural element Wilder includes tools, he only pays real attention to beliefs, theories and other propositional items. His first example of a cultural element is a shared belief about bird behaviour; he mentions that such knowledge becomes traditionalised, but does not consider how this happens. I have already put forward the thesis that in mathematics it depends upon the fourfold written sign-system and the associated context. It is not that Wilder has an alternative to my view, he doesn't consider the question. For Wilder, mathematics consists of sets of beliefs. But the theory of evolution (like that of vectors) concerns *material* objects and their forces. And there are material objects in mathematical praxis - people, texts, artefacts. The basic problem with an approach of Wilder's kind is that it attempts retain the sociological insistence that social effects (e.g. changes in theories) be explained by *social* categories, while leaning on the explanatory aura of evolution theory, which actually achieves its explanation of social effects (e.g. speciation) in terms of the statistical properties of *individual* objects (nucleotides, ultimately). Evolutionary epistemology generally suffers from this combination of the fear of reductionism²⁴ and the lust for scientificity.

Returning to consolidation, I can illustrate these general complaints. The chapter is called "Consolidation: force and process". Wilder says that it is immaterial (!) whether consolidation is termed the one or the other. Considering the care he takes to clarify the meaning of his term 'culture' in his first chapter, because he has found that misunderstanding is easy, it is amazing that such

²³ Incidentally , there is a persistent misunderstanding of the Appel/Haken proof, never mind its significance (no computer *concepts* are involved, as Wilder seems to think).

²⁴ Not that I think evolution theory is reductionist. But this is too long a digression to go into now.

insouciance is possible here: surely forces and processes are radically different things. To ask whether natural selection is force or process may explain his ambivalence, but it surely only shows that neither term is appropriate, not that both are. Later, Wilder mentions that there is both a social and an individual aspect of consolidation, but for the individual consolidation seems to be mere borrowing of techniques or concepts from an alien field²⁵. How this changes the concepts at the social level is not explored at all. One of his parallels raise the question implicitly: the steam boat is a consolidation of two prior ideas, and one can see roughly how the concept can arise when individuals create instances of steamboats by their individual action. What is missing from Wilder's account of the consolidation of algebra and geometry by Descartes is any indication of how the rather complex *concept* of analytical geometry is related to such objects (or acts) as individual writings by Descartes and Fermat. Surely he does not think that these are simple *instances*, like steam boats? Then again, perhaps he does, for his conception of language is very naïve: he clearly²⁶ takes language to be essentially naming.

I conclude from this brief discussion that Wilder's descriptions of mathematics are valuable as making plausible the identification of mathematics as a cultural system, of praxis as the concerted social action which reproduces and extends that system and brings it to bear on the rest of culture. But his rather half-hearted attempts to be "scientific" about this and to make use of evolutionary and vectorial metaphors to produce "laws" don't work because he ignores the material infrastructure of mathematical practice. Despite the vacuity of the "laws" presented and the unsatisfactory level of rigor in the conceptual elaboration, the basic viewpoint - "a way of looking" - deserves some applause. But it is a step in the right direction rather than an arrival anywhere.

²⁵ Ibid p88.

²⁶ Wilder [1981: 8]

§109 Kitcher on mathematical practice.

While Wilder's attempt to use evolution to explain the development of the mathematical cultural system is counterproductive, Kitcher's recent account of mathematical knowledge avoids this criticism. Because although it claims to be evolutionary, this carries none of the explanatory weight of the book: his account is historical but in no explanatory sense evolutionary. On the other hand, the concept of mathematical practice is given a detailed characterisation and does carry explanatory weight. Kitcher's book²⁷ is a welcome attempt to rescue the philosophy of mathematics from the deathly grasp of apriorism (together with apsychologism and ahistoricism) - his version of what I call the logico-formalist hegemony. His arguments against these views are workmanlike but not very novel. He opposes to the a priori conception of mathematical knowledge the argument from long proofs²⁸ and problems with the concept of mathematical intuition, on which apriorism frequently calls. He poses objections about the essential character and exactness of presentations in intuition, and the problem of practically impossible presentations. These objections point out that any cognitive power of mathematical intuition must suffer from the limitations of our outer intuition - which are what pose the problems of mathematical epistemology in the first place! His positive theories are less convincing, but more interesting. He puts forward the following view of mathematics.

For Kitcher the alternative to the apriorist account is an empiricist one. So he argues for a strong version of the Quinean holist view, that mathematics cannot be distinguished from science.

"we can think of mathematical change as a skewed case of scientific change: all the relevant observations are easily collected at the beginning of the enquiry; mathematical theories develop in response to these and all the subsequent problems and modifications are theoretical."

²⁷ Kitcher [1983]

²⁸ Such as the classification of the finite simple groups.

However, mathematics is clearly not about the same things as physics, so Kitcher has an account of "mathematical reality". Mathematics is the theory of possible operations of an idealised subject, operations which consist of various levels of collecting, correlating and ordering. In this sense it describes the structure of the world. Because it idealises, it is not empirical knowledge in the sense of perception, but it remains empirical, like scientific knowledge, with a perceptual basis. Mathematical knowledge is obtained by building on a (Mesopotamian) basis of empirical observations with rational transitions in *mathematical practice*. Changes in mathematics are of the same nature as changes in science; change occurs through the (rational) modification of practice. Mathematical practice for Kitcher has five components, which he writes as $\langle L, M, Q, R, S \rangle$, the language, metamathematical beliefs, questions, accepted reasonings and statements of a particular moment. Various kinds of change occur depending on which components are affected. An individual's mathematical knowledge is obtained mostly through teaching, from her community's authorities, who grasp and extend current practice. Their knowledge is explained by their own teachers' knowledge and their own rational changes to practice. So the explanation of mathematical knowledge consists largely in explaining the development of the expert practice. This is an historical ("evolutionary") task.

There is much of value in this account, but I intend to present in the next three sections three criticisms of Kitcher's conception of practice. First, Kitcher's semiotics is conventional and inadequate, relying on referential semantics. Second, his ontology is mysterious and not integrated with his holism. Third, his pragmatics is insufficiently attentive to the implications of his own observations about actual mathematical activity, to wit its dependence on writing. These criticisms are explained in the next three sections.

§110 Kitcher's conventional semiotics

Kitcher's general view of language is disappointingly conservative, and it is

here that his account is most obviously the worse for not attempting to come to grips with Lakatos and Wittgenstein²⁹. He simply assumes that semantics for mathematics must be referential. That is why he invents his theory about ideal operations, to enable mathematical sentences to rely on covert references to operations instead of their apparent overt references to numbers etc. Kitcher treats the meaning of a term as its reference, and considers that the only semiotic problem is how that reference is fixed. He does not discuss other kinds of terms, implying³⁰ that there are none. The main difficulty he sees is the case where the language component is changed by the introduction of a new expression which "violates previously accepted theorems". (Most notational changes are just abbreviations.) To account for these cases Kitcher introduces the idea of reference potentials³¹.

The *reference potential* of a term type is the set of events which can serve as "initiating events" for tokens of that type. The initiating event is the "ceremony" through which the reference of a token is fixed. Kitcher distinguishes three kinds: conformity (to other speakers use), present paradigm (ostensions) and stipulational (descriptions). His main point is that the reference potential of a term is a large and changing class. As it changes, so does the concept we identify with it. Kitcher uses these notions trying to clarify changes in mathematical concepts such as function, ω and i . What passes undiscussed is his conflation of the use of a term, with the supposed *reference* of it about which he continually talks. What is clear from his examples (and many others) is that a particular term of mathematical language has a changing and complex range of constraints on its acceptable uses. To see that this can explain various kinds of change in mathematics does not require any assumption that such terms have "references" which are "fixed" by these constraints. What does he really show about the term 'function', for example? That in its early uses (Leibniz), the term is constrained to certain *legitimate substitutions*

²⁹ His sole reference to Wittgenstein is a footnote citing Wittgenstein's worry about the cogency of fully written out formal proofs. In another footnote (p5) Kitcher gives some rather weak reasons for not engaging with Lakatos' ideas.

³⁰ "I suggest that we identify concepts as reference potentials." [1983: 168]

³¹ This idea has some affinity with that of "partial reference".

(the functions of a curve are the tangent at a point etc); later, Euler legitimated calling a function any of a class of *expressions*; later still (d'Alembert) some items are called functions which are determined by *descriptions* alone. The idea that a description achieves a reference in the same way as a name is highly contestable, yet despite Kitcher's talk about baptismal ceremonies, there are no names involved here. Of course 'function' is a kind term so this is hardly surprising. What is surprising is the equanimity with which philosophers accept the obvious inappropriateness of talk of baptism and references in such cases. The term 'function' *applies* to various items, it doesn't refer to them at all. In the cases of ω and ι , talk of naming and reference may seem more apposite, but these cases are really even less convincing: no ostensive acts are ever used in specifying the use of these terms, on the contrary they are defined in context from the start. The idea that they have references whose fixing is in question is gratuitous. The proof of this is that the accuracy of Kitcher's accounts of the introduction of these notations does not rely on their existence. It is ironic that Kitcher mentions in a footnote here³² that up to the eighteenth century the term 'number' was constrained by the requirement of possible concrete construals, which was dropped to make way for negatives and imaginaries; yet philosophers like himself very commonly retain just this requirement in demanding that meaningful terms must refer!

§111 Kitcher's platonist mathematical ontology

Consider now the notion of the "ideal operations of the ideal subject". Kitcher claims that these are the actual referents of our mathematical theories. These theories are idealised descriptions of powers which we actually have to perform collectings and collatings. There are many objections which spring to mind about this suggestion which I shall not pursue. I am interested to show that Kitcher's detailed discussion actually leads toward a rather different ontology, one

³² Kitcher [1983: 176n]

more consistent with the comparison he makes of his view with Chihara's Mythological platonism. "The common theme is that mathematical statements owe their truth to the stipulations on mathematical vocabulary which are laid down"³³. This kind of textual conventionalism will be adopted in Chapter 12 below: it is in the spirit of Wittgenstein's philosophy of mathematics. But it is far from the ideal subject.

In order to make plausible his account of mathematics as about possible operations of an ideal subject, Kitcher must explain how there can be collectings of collectings and so on, analogous to sets of sets. To do so he claims³⁴ that

"To collect is to achieve a certain type of representation, and, when we perform higher-order collectings, representations achieved in previous collecting may be used as materials out of which a new representation is generated."

To make this plausible he makes the point that

"mathematical language plays a dual role. Not only do the sentences which occur in mathematics books describe mathematical operations...but in producing those sentences, the mathematician may be engaged in performing those operations. In inscribing the token '{a. {b}}' I may be achieving that representation which constitutes the performance of a collective operation on a and the prior operation which collects just b."

Now this won't do, since it is not "the operation which collects just b" (an achievement), but its *result* which is collected along with a. Kitcher tries to deny this by explicitly claiming that we collect operations not symbols, just prior to the quote above. But saying does not make it so! The token of {b} in the token of {a,{b}} in the quote is a *representation* of the simpler collecting, not the collecting itself; so the token of {a,{b}} though it represents, indeed is, a collecting of a and {b}, is not a representation of a and the collecting of b. For if this were so, we could be nonplussed by the discovery that a were {c}, so that it was really a collecting too! If this is denied then it is hard to see how the inscription can be an instance of the operation it "describes", since it is symbols not collectings which are collected by it.

³³ Kitcher [1981: 110n]

³⁴ [1981:]

I cannot see how Kitcher can call the performances of writing "ideal" either. Nor is it evident how truth be guaranteed for a token, by its being one of a kind attributed to the "ideal subject" (except perhaps for some existential ones). But this recognition of the "duality" in the role of language, and Kitcher's connection of it to the value of perspicuous notation in facilitating performances of a kind we want to discuss, is important. As he says, what mathematicians do is not to peer closely at mathematical reality. Rather they manipulate symbols - or more exactly, I would say, they *play with signs*. Does this concession of Kitcher's reduce his account to the kind of formalism which has been often jeered as reducing mathematics to marks on paper? Well, yes, I think it does! Not that this is so bad, properly interpreted. But Kitcher clearly doesn't like the look of it, and he excuses his dropping the issue on the ground that we need more data from the psychology of mathematical research. Kitcher approaches the point that the operations in question are really inscriptions, but does not address it squarely. He prefers to gesture about possible operations of imaginary ideal subjects³⁵.

If we examine how Kitcher's ontology actually figures in his accounts of mathematical change, we find that as far as possible he keeps it out of them³⁶. The case where he claims that it is useful is in explaining why complex numbers were gradually accepted when they were. Kitcher's claim is that the invention of the Argand diagram provided an extension of the reference potential of 'number' by allowing its referent to be fixed by new vectorial operations as well as the old operations of counting and measurement. Imaginaries could now be accepted as numbers because they can be specified through these newly acceptable number

³⁵ I think this is all a muddle caused by fear of the usual bogey questions of the philosophy of mathematics. We dare not say that mathematics is about numbers and triangles, because these have proved to be too troublesome for our epistemologies and ontologies. So we take refuge in fashionable havens such as *possibilia* (which though controversial, are at present ontologically OK). It's amazing how many people are prepared to "countenance" possible worlds or their equivalent (structures, for example - cf Chapter 2) in order to avoid "countenancing" numbers. Or equally spaced regions of space-time (transcendentally deduced rulers) - cf Field [1980]

³⁶ Kitcher [1981: 171].

fixing operations. But in this I find nothing about the *ideal* operations of *ideal* subjects: rather, this explanation use the *actual* operations which *real* subjects use to specify such newly accredited numbers, by writing down expressions like $6cis(\pi/5)$ whose meaning is indeed determined by the vector geometrical operations. In the same section, Kitcher claims that Cantor's theory of infinite numbers was accepted through a change in the reference potential of number, namely by removing the "theorem" that nothing followed an infinite series. Later, in chapter 9, he claims that Cantor's theory was accepted where Bolzano's was not³⁷ because it throws light on ordinary arithmetic, and Bolzano's does not; thus it provides a rational extension of the description of the powers of the ideal subject. But I would say that this idealism is unnecessary; the real difference is that Cantor's theorem allows us to compare infinite sets of different power, which turns out to be very useful, but Bolzano's theory did not. Kitcher even mentions this contrast, but does not see that this extension to the *actual* powers of *real* subjects is enough to justify the different treatments which the theories received. Notice that here too, these new powers are exercise by notational means: such a comparison is effected by specifying a 1:1 correspondence, or proving that one does not exist.

In this section I have argued that Kitcher's ontology is platonist, with internal difficulties and little explanatory force. *This* is not mathematical reality.

§112 Retreat from pragmatics

There is a second point in Kitcher's theory where he almost recognises the need for writing in mathematics. In arguing that the a priori notion of mathematical knowledge fails to the worry about long proofs, he develops the "Storage model" of proof advanced by Descartes. The point here is that somehow the truth of every proposition in a proof must be evident to the cognising mind at the one time, yet it is difficult to keep many things in mind at once. Long proofs throw doubt on the suggestion that running over proofs repeatedly helps extend our grasp.

³⁷ Kitcher [1983: 209ff]

Later he makes some changes to his conception of proof, making proof a functional term. A proof has the function of generating optimally new knowledge from old, or else of increasing our understanding of old knowledge. If a "sequence of statements"³⁸ does this it is a proof. Proofs are not characterised formally. I agree with this point of view, because it accords with actual mathematics.³⁹ My point here is that this welcome change to the class of sequences of statements which count as proofs does not affect the difficulty created by long proofs, nor the need for an account of the way proofs are used which copes with them. And the fact is that we rely on *writing* to support proofs which enable mathematics to get to a reasonable level of complexity. There just *is* no mathematics in an oral culture. These two points together come to this: Kitcher has noticed without realising it that mathematics, *doing* mathematics, means reading or writing mathematics. Mathematics does not refer to so much as exhibit its subject matter; and gaining mathematical knowledge or understanding by the use of proofs is not consulting the mythical platonic realm but following and grasping written exhibits. Mathematical praxis is *textual*.

Yet Kitcher's notion of a practice is completely idealist. It consists of a language, i.e. a syntax and semantics (whatever they are), a set of beliefs about mathematics, sets of accepted statements and reasonings and questions. This is somewhat platonist! Just as does Wilder, Kitcher thinks of mathematics as sets of propositions and the concepts of which they are composed. One might have expected a *practice* to include methods, techniques, tools, systems, institutions and artefacts. And in fact of course, mathematical practice does. Even Kuhn, from whose paradigms Kitcher has derived the idea of practice, includes exemplary *texts* in a paradigm. In mathematics as in philosophy, the main tool available for the worker

³⁸ [1983:181]

³⁹ But I think Kitcher could address its divergence from what mathematicians commonly say about proofs. He says [1983:190] that such remarks are material for philosophical analysis, not contributions to epistemology, but there is surely some tension in taking an empirical attitude to what is mathematics at a time, allowing that the metamathematical component varies and therefore so does what is accepted as proof, while dissenting from the present consensus that real proofs are formalisable in ZF! (Davis and Hersh [1979,1985 *passim*] indicate that it is not a thoroughly heartfelt one.)

is the textual inheritance. (How do tables and diagrams fit into Kitcher's notion of practice? Are the diagrams in Euclid and Newton part of mathematical language?)

A rather unconvincing argument of Kitcher's, is his attempt to assimilate mathematical change to scientific change. He attempts with some success to play down some of the obvious differences, such as the role of observation in science and the persistence of truths in mathematics. But he is forced to admit that there is a difference he cannot explain away: rival theories in mathematics are not killed off (replaced) but reinterpreted, as are rival geometries and set theories. His explanation of this difference is that "the task of the mathematician is to unfold the possibilities of theory construction....." But this does not seem to me to make any difference. There is still a gross difference here: in science there are rival theories, in mathematics merely theories in need of consolidation or generalisation. Kitcher's theory of reference potentials, an attempt to avoid the perils of incommensurability arguments for mathematical transitions, relies on the idea that mathematical terms refer in the simplest sense, and this is an idea which just won't do. In any case it is motivated by his wish to avoid the "incommensurability problem" being transferred from science to mathematics if his argument for their continuity is too successful. The other main line which Kitcher has on blurring the distinction between science and mathematics is to emphasise the complexities which have been recently found in the idea of refuting scientific theories, using the related point, made by Putnam particularly (cf Chapter 2), that quasi-empirical methods - evidence and experiment - do seem to play some role in mathematics. But one can grant that there are complexities here without abandoning the sense that mathematics is necessary, physics not. Perhaps putting it like this is too provocatively naïve; but I think that there is a strongly felt cognitive difference between mathematics and science which needs accounting for. Quine's celebrated attack on the analytic/synthetic distinction is not merely far less convincing now

than it was in more positivist days⁴⁰; even if correct it would leave a large space for an explanation of this felt difference. Actually this blurring attempt of Kitcher's is curiously ambiguous. It is hard to see how one could justify distinguishing mathematical practice from scientific practice if his argument were correct. And scientific practice is even more obviously organised around material artefacts (instruments, for example) than is mathematical. But in fact when Kitcher gets to discussing mathematical practice, its continuity with scientific practice is quietly shelved and although applications are mentioned occasionally as motivating the mathematicians involved in the development of classical analysis, they do not come into the explanations, and mathematised artefacts, of which science is full, are not mentioned at all.

So at the individual's level, Kitcher does not recognise the importance of the material pragmatics of mathematical activity; and at the social level he does not recognise the importance of various material aspects of practice - the mathematical sign-system, mathematically literate people, the archive, and the mathematised world which I have described in preceding chapters. His account claims to be empiricist because it is not apriorist; but really it is idealist, as one can only expect given his account of mathematical reality. He has recognised that mathematical knowledge has empirical content, but not given an empiricist account of how it is acquired or how it changes.

§113 Consolidated criticism of these writers

It may seem churlish to have adopted such a critical stance toward the writings on mathematics which I regard as the most worthwhile attempts to transcend the logico-formalist hegemony. But apart from the fact that we have not yet come to grips with Wittgenstein, whose work I think is more important still, I

⁴⁰ For example, "Two dogmas" relies on the idea that if we cannot give a clear necessary and sufficient characterisation of this distinction, then it must be unreal. Thereby it is assuming that all concepts have explicable sharp boundaries, an assumption which Wittgenstein attacked in his later philosophy of language, successfully in my view.

think it inevitable that since these attempts are so recent, and the inertia of a century of logico-formalism so great, we must be a long way from a satisfactory account of mathematics. I will draw together here what I think is good in these writings on mathematical practice, and summarise my criticisms of them. This will identify what I think still needs to be added to conceptualise mathematics as it actually is. So I will end this chapter by indicating what I think a sound concept of mathematical praxis involves.

If we distill from the writings which I have discussed in this chapter a concept of practice which takes what is strong from each of them, it would be something like this. Mathematics is a cultural system, consisting of cultural elements related by communication processes. This system has a number of important components, $\langle L, M, Q, R, S \rangle$, each of which is subject to change by a variety of processes. The overall logic of this dynamic system is that of the criticism of attempted proofs by the method of proofs and refutations. The three main weaknesses which I have identified in these conceptions can also be listed shortly. The conception of language on which they are based is essentially logico-formalist, accepting referential or representational semantics as in principle adequate for the essentially fact-stating nature of mathematical expressions. The conception of mathematics which is being put forward is still at bottom an idealist one, taking the object of our investigations to be propositional in nature, ignoring the material dimensions of mathematics as it really is. The relationship between individual activity and social consequences is not really addressed, mainly because the concentration on the abstract level precludes any serious attention being given to actual activity. Lakatos' dialogical experiment purports to simulate mathematical activity but is really simulating talk about it. The other accounts do not investigate real activity at all - and nor does anyone else. I have pointed out in Chapter 2 that there are methodological difficulties here, but difficulties do not justify such a glaring omission.

That omission is connected with two others which are also easy to point out but much harder to do anything about. The first is that these accounts do not articulate the *point* of mathematical practice: why is this cultural system important and why do we have it? The second is consequential on the obvious answer that mathematics gives powers which can be applied outside the mathematical system. How is the mathematical system related to its social environment, the scientific, technical, political, economic, artistic, educational, and other systems in culture? This is not a demand for the philosophy of science, of technology etc to be incorporated into mathematics, but for the application problem to be recognised in the very conception of mathematical practice.

I have indicated already what I think needs to be done about the reliance on positivist conceptions of language: Wittgenstein's critique gives us the basis for a reorientation of our philosophy of language, and my own account of mathematical text extends it to the specifics of the mathematical communication processes. This leaves a lot still to be done however in tracing specific conceptual changes in mathematics through their textual evolution and technical applications. Further conceptualisation of the mathematical context along the lines of Chapter 9 is needed to get away from the idealist concentration on theories and concepts. As to the articulation of the point (or points) of mathematics and of its relation to other cultural systems, I will suggest in the rest of this chapter that the key to this is to develop a notion of *mathematical work*. The relation between individual action and social process is a very difficult topic and I do not pretend to have a satisfactory approach to reveal. Indeed I think that at present our descriptions of the phenomena, mathematical practice, are too weak to permit any reasonable attempts at explanation which might pretend, as does Wilder and perhaps Kitcher, to some kind of scientific status. Wilder talks about prediction, but Kitcher is not so naïve. Yet Kitcher's naturalised epistemology like Quine's from which it descends, must surely carry this consequence if it is genuine. My Wittgensteinian attitude is that good enough description will help to resolve philosophical perplexities about

mathematics rather than that it will found a science of mathematics planning. Only I think, against Wittgenstein, that such resolution must come through understanding actual historical events as much as as understanding how we use language.

None of the writers I have discussed in this chapter shows much awareness of the fact that mathematical praxis is prosecuted by material creatures with specific physical cognitive capacities using techniques which allow them to support specific cognitive capacities. I have argued that the technique which is necessary for the prosecution of the mathematical praxis is the particular form of writing system found in mathematics with its four functionally independent codes, and the threefold context which has arisen through the praxis itself. These material aspects of mathematics been largely ignored in the literature and that is true of these writers too. In the next section I will try to fill the gap in these conceptions with the concept of *mathematical work*.

§114 Mathematical activity is work

Any individual activity of mathematicians, such as the borrowing of concepts from one field to solve problems in another or calculating cases of a number theoretic conjecture looking for a counterexample, is *work*. It is work because it requires great efforts, is part of a socially organised enterprise, is demanded by others and paid for⁴¹. This fact about mathematical activity is the source of its inadequate conceptualisation in writing about mathematical practice. I will briefly justify these claims in this section.

The effort required in mathematical activity is quite obvious to anyone who has tried it. It is even more obvious to anyone who has taught it or tried to help

⁴¹ Sociologists might dispute whether these marks *define* work. But they *characterise* it sufficiently for my present purposes. Recall my remarks on definitions at §21 and §60. Rossi-Landi [1983] and Sohn-Rethel [1978] have some relevance to what follows.

children learn it. Newton is said to have attributed his great discoveries to "thinking on them constantly" - and is known to have worked 20 hours a day, with legendary concentration. Mathematicians commonly work for years trying to solve problems; Einstein spent several decades looking for the unified field theory he never found. Such efforts are necessary because mathematics is difficult. There are many reasons for this, some of which will be discussed below, and some in Chapters 11 and 12. The difficulty is not something created by capitalist schooling, as is sometimes suggested by progressive educators. The fundamental reason is that mathematical activity is work on one's own concepts.

The organisation of mathematical work is also readily apparent if we think about where it is done. Mathematical learning is largely done in highly organised educational institutions, where much of new mathematics is made too. The provision of the materials of mathematical work is mediated by large organisations such as scholarly societies and the journals they edit, education ministries and their bureaucracies which produce curriculum material, publishing and distribution systems, libraries and the complex organisational infrastructures they presuppose, laboratories and the highly organised technological economy that stands behind the instrumentation which they employ. The prosecution of mathematical work is also highly organised: most learning is orchestrated by curricula, teachers, accreditation and award systems, in social groups ("Applied Statistics 3", etc) formed for that specific purposes in peculiar communicational relations with other social groups. (The apparent *disorganisation* of the complete autonomy which researchers have to choose their subjects of study is largely delusory, as I shall argue in a moment.) Much of the application of mathematics goes on in large organisations such as economic enterprises, government departments and instrumentalities, applied scientific research institutions. In all these, the mathematical work which is done is highly directed by the aims of the organisation and its choice of problems, methods, budget, time constraints, shareholders or constituents' demands etc.

Implicit in the organisation I have just pointed out is the fact that much mathematical work is *demande*d by others. It is not freely taken on as task by the worker but is wholly or in large part required as part of social or economic obligations. In societies like ours everyone has to do mathematical work at school or nine or ten years like it or not - and most don't. Very large numbers of people have to do more of it as part of the requirement for some real aim of theirs - as prerequisite for entry to some course, or as part of the course ("statistics for psychologists" is the classic case). Most do so unwillingly. Many people have to cope with simple mathematical work as part of their work in banks or shops or factories or offices, and cope unhappily and/or not too well. Even the tiny minority of mathematical workers who have "freely" chosen to become research mathematicians have a wide range of obligations on their activities: in order to get attention for their own work they have to give attention to that of others, which requires work for its understanding. To pursue their own work they often have to learn theories in which they have no particular interest, which may be relevant to their aims. Even their aims are not chosen freely but can only be selected from the objective situation in a particular field of mathematics at the time. If problems on offer are not particularly interesting, there is still an obligation to work on things - indeed too much so, for there is on many academics an unhealthy pressure to produce results - which usually means work needs to be done.

The motivation implied by my last remarks is of course that most research mathematicians or applied mathematicians such as statisticians are *employed*. They are *paid* to work on mathematics. There may be great freedom in what they work on, but the extent of this freedom and the numbers enjoying it are inversely related. Some employees of the Australian Bureau of Statistics have some freedom to pursue mathematical work not immediately needed by the organisation, but most of the tasks of most of them are dictated by the ongoing and pressing tasks with which it is charged.

I anticipate that all the considerations of this section may seem easily brushed aside as merely empirical facts about our present society. Did not Greek mathematics flourish pursued by gentleman amateurs motivated by pure considerations such as the love of truth and the desire for wisdom? Well no, not entirely. Even in classical Athens there was competition for pupils; Archimedes was often pressed into military service designing war-machines; the direction of interest of classical geometers was strongly influenced over centuries by the origins of geometry in practical problems; and gentleman amateurs aspired to power then as now - Plato's courting of tyrants is merely the example most familiar to philosophers. In the early centuries of modern mathematics the gentleman versus practitioner division of mathematical labour underlay the continuing honorific use of "geometer".

We have an image of mathematical work which is free of base motivation or pecuniary interest but it is wildly at variance with the general nature of mathematical practice. I shall offer some argument about why this is so in the section after next. First I shall point out the textual nature of mathematical work.

§115 Mathematical work is textual

The actual activity, mathematical work, is enmeshed in economic relations then. But what form does it actually take? On the basis of the analysis I have given in preceding chapters, I say that mathematical work is textual. The work of application of mathematics is textual (though not only that, of course), the work of creating mathematics is textual, the work of learning mathematics is textual.

The textual nature of applying mathematics can be seen in the massive documentation which is generated by enterprises which are making such applications, such as the Bureau of Statistics. Indeed there is a wonderful study of medical research⁴² in which mathematics is obviously being applied in a routine way all the time. One of its most striking findings might be formulated as that such

⁴² Latour and Woolgar [1986]

research is essentially (or can be usefully seen as) a process of *the production of documents from documents*. Again, it is apparent from my discussion of diagrams that many uses of mathematics lean on the diagrammatic component of mathematical text for mediating the application of mathematical results, as in the use of plans and maps.

The textual nature of the creation of mathematics is a matter of personal experience. Trying to solve mathematical problems is a matter of playing around with the existing literature or various textual formulations of the problem and a welter of half-coherent scribbles on scraps of paper until, if one is lucky, something suddenly looks right. Whether this is the general nature of the activity is hard to judge, and I return to this topic in the next chapter. In any case, whatever the working methods of creative mathematicians may be, there is no doubt that students, who are often being forced to re-invent already existing mathematics in the process of learning it, do work like this.

As for the work in learning mathematics, if I am to learn why Riemann's hypothesis is true after all, I must come to understand a proof of it. There is no other way, no royal road⁴³ in analysis any more than in geometry. Now how am I to gain this understanding? By working through a proof. 'Working' is correct, coming to understand a genuinely new proof always requires work, it cannot be read like a newspaper report about the cricket. And of course it must be read, proofs require a sustained attention which can only be given to signs which will stay still for long and detailed inspection. The verbal proof-sketches which workers may give one another are mere promissory notes on real proofs. The reason why understanding a new proof requires work is that proofs change concepts by novel constructions, as I will argue below in Chapter 12. Just as new muscle is built on and through the exercise of old ones, our concepts are developed by putting them to work in difficult contexts. When unusual things are said about items about which I already know a

⁴³ Euclid is said to have responded with such a denial about geometry to some lazy monarch who wanted the knowledge without the learning.

good deal, I must actively move my conceptual counters around in order to accommodate the novelties into the overall picture. Since, like everything else, mental activity occurs only against a gradient of inertia, work is required to achieve this accommodation.

The process of actually acquiring new knowledge by working through a proof has no doubt been the object of some psychological investigations, but I wouldn't hope for any great enlightenment from that quarter. Only, a corrective to taking one's own experiences as general. My experience is that no matter how well organised a proof, one's assimilation of it is obtained by a minute scrutiny of the parts of it and their connections to each other and to other earlier texts. This scrutiny deviates from the actual sequence of presentation of the proof in order to establish content for all the empty intuitions which are inevitably invoked by it. If the proof uses a property of cyclic groups, I must reactivate my apprehension of what cyclic groups are, verify that the property claimed is indeed true of them in the circumstances given, perhaps referring to a text on group theory. All this is highly textual work.

§116 Mathematical work is phenomenologically misleading

There is very little literature on the making of mathematics, because the prevalent idealist view is that it is a private process going on in the recesses of consciousness in which the individual grapples with ideas in an ineluctable isolation, and in which individuals (men) of genius create new mathematics (though only recognising old realities) through strokes of insight not amenable to rational explanation. This is bosh! I will argue in the next two chapters that although the process certainly involves "pure thought", it does not make sense and could not be prosecuted outside the context which I have been setting out as the locus of applications. The creation of pure mathematics in the clarity of pure thought and the application of mathematics in a sea of mathematised confusion are merely two different aspects of one overall process, which cannot be understood or go on in

isolation from one another. We have to get away from the picture "here am I intuiting number facts, over there are the apples. How are we related?"⁴⁴

This "Hardy" conception⁴⁵ of the work of the mathematician has a strong attraction for practitioners. Some sources of this attractiveness are the concentration on culturally novel mathematical making; the reliance on anecdotal reports of the phenomenology of leading mathematicians⁴⁶; the division of intellectual labour and the phenomenology of mathematical work. One seems to be working in the mind's eye only, as contrasted with the artisans making the tools or taking the readings.

This phenomenology is misleading in at least two ways. It is tempting to think that mathematics can be done entirely in the mind's eye because undoubtedly much is. One thinks about problems at odd times and in odd places, or even just staring into space. But I would say that this facility of envisaging the problem is a power which has been acquired through mathematical work, deriving from textual work proper and not possible for someone without such experience⁴⁷. The delusion we must avoid here is analogous to the idea that text acts are all derivative because they can be carried over into speech act counterparts. The other aspect of "mental arithmetic" which is misleading is the sense of objectivity certainty which arises from it. I will discuss this at length in Chapter 12.

§117 What mathematics comprises

I will now sum up how mathematical practice seems to me. I will use **praxis** instead of 'practice' to signal the distinction of my own conception by the inclusion of the concept of mathematical work. Mathematical praxis is a cultural system founded on texts written in a fourfold sign-system consisting of Word, Diagram,

⁴⁴ This is only mildly a caricature of the scene discussed in the Craig/Wright debate which I will discuss in Chapter 12.

⁴⁵ Hardy [1967: 127]

⁴⁶ Hadamard [1954]

⁴⁷ The experiences of blind mathematicians would be of interest here, but no-one seems to have studied them.

Notation and Paragraphy, in a context with three main aspects, an Archive, a mathematically educated Readership and a Mathematized world, maintained and extended by processes of mathematical work, work whose evaluation is deeply ambivalent, dispersed widely in the society and interfused with most other forms of work and production. The over-riding modality of this work is *correction*, ranging from simple marking of school work through the technical and scientific confirmation and disconfirmation of applications of mathematics, to the logical criticism of mathematical discourse in the professional literature⁴⁸. The work is effected largely through text acts of a variety of kinds.

Mathematics is generally taken to consist of concepts, definitions, theorems and theories. I do not deny this; on the contrary, I myself use these terms freely. My description of mathematics by no means excludes the more traditional elements anyway: readers have concepts and work on them, in texts they propose definitions and expound theories, and so on. What I reject is the attempt to deny the importance of the material embodiments of these forms by ignoring them, which leads to the prevalent platonism in philosophy of mathematics and which is the source of many of its problems. Curiously enough, in the logico-formalist description of mathematics these terms do not figure and it is only in the unofficial supplementary remarks, such as the launches which I have pointed out in classic expositions, that they are mentioned. This is an interesting rhetorical technique which I shall consider at length on a later occasion.

In Part III I shall return to the problems of philosophy of mathematics with this conception of praxis.

⁴⁸ I do not propose to designate mathematical praxis in set notation along the lines $\langle\langle W, D, N, P \rangle, \langle A, R, M \rangle, \langle \dots \rangle\rangle$ because these components are not usefully thought of as sets, and because nothing would be gained. What use does Kitcher make of his quintuple representation of practice? None.

Part III

A conception of mathematics

Chapter 11

Making mathematics

§118 Creation and application in praxis

In this chapter I discuss the application and creation problems in the light of my description of mathematics developed in Part II. My conception of mathematics draws on Wittgenstein's later views, so in the first two sections, I sketch those features with which I agree of his mature philosophy of mathematics and explain how his views will be developed, corrected and extended throughout these last two chapters using my analyses of text, context and praxis.

In §§120 and 121 I will suggest that inventing mathematics must be seen as a textual process: I oppose an account based on mathematical praxis to Hadamard's idealist one. In the rest of the chapter, I discuss the application of mathematics. I contrast Wittgenstein's view that mathematical sentences are rules for the use of mathematical terms in non-mathematical contexts with the view that application concerns the relation between vague and exact concepts made through measurement. I demonstrate how formal descriptions of the logic of measurement conceal several crucial pragmatic presuppositions. I argue that these presupposed social practices are the sources of the normative force which Wittgenstein correctly identifies in the application of mathematics to the world. This argument provides a way to resist some recent attacks on Wittgenstein's notion that mathematical signs are used to regulate applications. Pointing out that applications form a highly varied spectrum in terms of their reliance on the pre-existing mathematical praxis, I conclude that the normativity in application is brought to bear in the way N-signs are treated in these social practices.

§119 Wittgenstein's later philosophy of mathematics

Wittgenstein published his early views in the *Tractatus*, but when he came to reject some of these views in the 1920s, his new conception of philosophy precluded the proposing of distinctive theses. In consequence he had insurmountable difficulties in creating a satisfactory presentation of his later views and methods. There is, however, a growing recognition that there are many points in common

between the earlier and later views. The general Wittgenstein literature is now very extensive, though that specifically on his philosophy of mathematics is still not large. This is rather surprising since it has become apparent, with the publication of more and more of his remarks and notes on his lecture courses, how central and basic his thinking about mathematics was for his whole later philosophy. When the first edition of his remarks on mathematics was published in 1956, it was poorly received. In retrospect it seems that the novelty of Wittgenstein's later philosophy was not immediately understood. This is no longer so as far as the philosophy of language and mind are concerned, but the philosophy of mathematics is still largely ignored. The literature consists of a few unsympathetic early papers¹ which don't engage with the real challenge of Wittgenstein's radical views, a series of discussions of one particular theme, the idea of alternative mathematics, which I shall discuss below, two short monographs and a long study², and then a variety of more recent articles on particular themes³. Wittgenstein's later philosophy of mathematics is distinctive in a number of respects, but it lacks a comprehensive and authoritative statement. One reason why the secondary literature is less successful in repairing this lack than it has been for the similar lack for his philosophy of mind and language is the interdependence which holds among all these areas of Wittgenstein's thought. It is a consequence of his later views that philosophy of mathematics cannot be hived off from the rest of philosophy in the way which has been customary this century.

The novelty of Wittgenstein's philosophy of mathematics is his anthropologism. It begins from his view of language as the life of society, language is an expression of the will, an instrument of control. This entails that agreement in action is the basis of life, in particular of logic and of mathematics and that language is a toolbox not a single tool. That is to say, meaning is made by use and

¹ For example Kreisel [1958], Bernays [1959]

² Klenk [1976], Kielkopf [1970], Wright [1980]

³ For example Hacking [1985] on necessity; Fogelin [1968] on Intuitionism; Ambrose [1972] on generality. Further citations are given later.

the uses of language are very various. In philosophy, he thinks, we are led astray by pictures and analogies deriving from one use of language misapplied to another. In particular we wrongly think of mathematical sentences as fact-stating because of their superficial similarities to other kinds. But really the point of mathematical statements is to guide the use of mathematical terms in all the variety of real life: mathematical sentences are rules. Applicability, on Wittgenstein's view, is built into mathematics⁴.

A recurrent topic of Wittgenstein's remarks is that calculation is not an experiment. This is because proofs and calculations are geometric. It is the key to the fact that mathematics is normative. Wittgenstein makes his sharp distinction between mathematical and non-mathematical statements based on this difference: the former are rules for the latter. These rules, i.e. conventions, are made not found. Mathematicians make essence, they don't find it. Mathematics like logic is *conventionally* necessary⁵. What this means I shall discuss at length.

Proofs are creative in Wittgenstein's view: it is in proofs that mathematicians make rules and thereby modify concepts. Wittgenstein takes it that the *informal* proof is the real proof, he is not impressed by formalization⁶. A proof gives us a technique; definitions are transitions, not abbreviations. We must look at the proof not the proposition proved to see what it does for us⁷. The voluntarist element in Wittgenstein philosophy of mathematics is that proofs must be adopted: necessity is created not recognised. We make statements into mathematics by regarding them as rules - but we do it to well-motivated generalisations ("the deep need"). This puts the requirement on proofs that usable proofs are surveyable⁸.

⁴ Wittgenstein [1956; part 5]

⁵ Ibid, part I, §§ 75-103, part II, §§ 65-76;

⁶ But he is not *opposed* to formalisation, only to misinterpretation of what it can achieve. My attitude is the same.

⁷ Here Wittgenstein anticipated an important insight of Lakatos'. This doctrine that "proof precedes truth", as some of Wittgenstein's commentators put it, is a bad formulation, given his notion of mathematical truth.

⁸ Ibid part I, §§24-74; part II §§1-44; part III, §§29-41; part IV, §§41-46.

Consequently, alternative mathematics makes sense, in a sense - in a different form of life. We are not forced to have the concepts we do have. The open texture of all concepts show that choice is involved in the linguistic action which builds concepts up. The "rule-following considerations" which Wittgenstein frequently raises apply this idea to calculation and proof; I will discuss them shortly too. The possibility of choices being made other than those which we usually do make entails the possibility of "queer arithmetics" (as Lehman calls them) and other deviant kinds of mathematics⁹.

Wittgenstein's attitude to platonism in mathematics is one of *therapy*. The practice of discussing mathematical subject-matter as objects leads to endless difficulties because of the grammar of the word 'object', which entrains a whole picture from whose implications we find it hard to break free. The problem is not to explain whether mathematics has objects and what they are like, but to see that the whole question is basically senseless. It is a consequence of taking mathematical sentences as descriptive and not normative¹⁰.

Wittgenstein sharply distinguishes mathematics from philosophy of mathematics. The philosopher is not to criticise the mathematics, or urge revisions of it - as the leading schools of thought of his time, Logicism, Intuitionism and Formalism all allowed - but to enable us to command a clear view of it. Moreover, since he thought that mathematics was not in need of foundations, and logic could not supply them anyway, he was sceptical about the philosophical significance of work on "foundations". He did not consider the consistency proof hunt of particular importance. But he did think that the way many mathematicians talked *about* results such as Cantor's and Gödel's to be the source of great confusion¹¹.

⁹ Ibid , part I §§134-169; V, §§27ff.

¹⁰ Ibid part IV, §§52-3

¹¹ Ibid part I I, §§77-90, part V §§27ff. I have cited illustrative passages in the *Remarks on the Foundations of Mathematics*; similar ones, and others of great independent interest, can be found in the other Wittgenstein items listed in the bibliography, but I do not wish to become involved in Wittgenstein exegesis for its own sake; I wish to *use* these ideas.

Mathematical writing about infinity is perfectly in order, but Wittgenstein thinks it must be completely separated from ideas of hugeness, or infinite numbers; when we say that mapping the rationals 1:1 to the reals is impossible, we are saying we have no use for a statement like 'f maps \mathbb{Q} to \mathbb{R} one-to-one'. Since for Wittgenstein all the philosophical problems are in ' $2+3=5$ ', he avoided the discussion of sophisticated, and so complicated, examples where one could be misled by technical irrelevancies.

§120 Combination of the two

I propose to combine these ideas of Wittgenstein's with my description of mathematical praxis in the following way. Wittgenstein has little to say about the creation problem, but my criticism of Hadamard's idealist account and the alternative I propose in the next two sections based on the idea of textual work is in the spirit of his later philosophy, particularly the idea that an inner process stands in need of an outer criterion¹². My contention is that "flashes of inspiration" are not the essence of mathematical invention, because the process of making new work public contains the real criterion for its being accepted as mathematics. (This argument is essentially the argument already made in Chapter 3 against Intuitionism. But now we realise that the public is the mathematical readership.) To address the application problem I intend to amend his idea that pure mathematics constructs rules for the use of terms in applications, by showing that these uses presuppose normative social practices other than the linguistic. In this way I will supplement his claim that the sentences of pure mathematics are not descriptive by identifying the sources of their normative content. I will concentrate on measurement in this part of my discussion.

The peculiar truth of mathematics, its necessity, is explained by Wittgenstein through his highly controversial social theory of necessity. I shall defend this view against several recent attacks in the last part of this chapter by

¹² Wittgenstein [1953: §580]

arguing that an adequate conception of mathematical praxis, mine, shows how the normative force is put into mathematics through social processes. I am able to distinguish this kind of necessity from other kinds with which Wittgenstein conflates it, by exhibiting its textual basis. (If this view is conventionalism, the conventions are not like those usually thought of in connection with this label. The apparent relativism of the view is real but harmless, though hard to accept. It does not in any way impugn the necessity of mathematics to explain it.)

Continuing with this program in the next chapter, I shall explicate Wittgenstein's idea that a proof is a design through the characterisation of text which I have given, in particular through the identification of specific new functions for Notation, which extend while preserving those of Word and Diagram. I will explicate Wittgenstein's insistence on the perspicuity of proofs, and justify the Formalist's conviction that concrete proofs make necessary truths evident, by showing how in textual work on proofs the creation and the recognition of necessity are fused.

I shall adopt Wittgenstein's idea about the object problem that it arises from our projecting our forms of representation. The object problem is a pseudo-problem, created by inappropriate talk about objects in alien contexts. I will improve his account by clarifying what projection means, and by showing how the specific features of mathematical praxis contribute to the peculiar force of the objectification of mathematics.

Consequently the access problem is also a pseudo-problem. There is no need for special mathematical intuition to explain how we are aware of peculiar mathematical objects, our knowledge of numbers and such like can be explained by careful attention to how we actually use text to create, learn and use mathematics. Our ordinary intuition, educated through mathematical praxis, is adequate.

There being no real problem about objects, there is none about infinite objects. I show how features of textual work make our access to infinite objects possible. However, there are a number of difficult problems about infinity and the

physical world which are hard to disentangle from this one, and I do not pretend to have given a complete account of infinity, which I must postpone to a later investigation.

Let us begin this program with a consideration of mathematical invention.

§121 Mathematical invention

There is very little convincing literature to be found on the nature of mathematical creation. The main work is still Hadamard's book¹³ *The Psychology of Invention in the Mathematical Field*, which is old and, I shall argue, considerably flawed. Investigation of this topic seems to have been inhibited by three factors. First, the prevalent "pure thought" conception, which Hadamard admirably articulates; second, by the logico-formalist hegemony which declares irrelevant to philosophy any questions about origins of our intellectual achievements; and third, by certain methodological difficulties which I have already mentioned in Chapter 2. These latter difficulties are actually exaggerated by the first two attitudes.

For Hadamard, the whole topic of mathematical creation (though he really only considers *pure* mathematical creation) may fairly be said to revolve around *one moment* in Poincaré's life. Poincaré was the pre-eminent mathematician of France around 1900 and his very clear philosophical essays had great influence. In his essay "Mathematical creation"¹⁴ he describes an important episode in his career, in connection with his discovery of the existence of Fuchsian functions. One particularly important idea came to him *as he stepped aboard a bus*. This moment seems like a mandala for Hadamard, it resembles Proust's moment with the madeleine and the tea. Indeed I would suppose that the two things, Proust's description and Hadamard's attitude to Poincaré's moment, are actually connected. In any event, Hadamard's idea of mathematical creation has as its centrepiece *the*

¹³ Hadamard [1954]. Written in the 1930s and 40s.

¹⁴ Poincaré [1956]

moment of inspiration by the man of insight. Other aspects of mathematical thought which he considers (the long preparation for work on a particular problem, the role of the unconscious mind, the different kinds of mind, the nature of the synthesis of ideas) are all treated as the possibly necessary surroundings of this mysterious but godly moment. He does not take very seriously the lack of anecdotes of this kind about inspirations which turn out to be mistaken (nobody reports negative results), or of such reports by less successful mathematicians than Poincaré.

The "great man" aspect of his account is emphasised by his discussion of Fermat's last "theorem" and of Riemann's hypothesis. In each case, a mathematician claimed to have proved something which has eluded subsequent mathematicians to this day. Hadamard does not seem to think it possible that perhaps they were *wrong* about having proofs which have unfortunately not come down to us, while being right in their *conjectures*. No, Hadamard thinks *insight* does not work like that - not for the real genius. But it does for most of us, and I suggest that it might even for them. Perhaps they had fallacious proofs.

Hadamard quotes a number of mathematicians describing what they think about, which in many cases is clearly visualised versions of real mathematical text. But he does not consider that the viewing of real texts and work with them may be part of the necessary basis for the long ruminations and the flashes of insight. In fact scrutiny of his text reveals that he completely neglects the material context of mathematical work, he does not consider applications of mathematics, and does not consider that perhaps physical syntheses, on paper at least, might be part of the explanation of the so mysterious mental syntheses of ideas¹⁵.

I will argue in the next section that this idealist account of Hadamard's could be greatly improved by more attention to mathematical praxis, by using a recent account of another discovery. The possibility of my description will give weight to the doubts I have raised about Hadamard's conception.

¹⁵ Cf my strictures of Wilder on steamboats in Chapter 10.

§122 How Conway found a new simple group

In 1967 Conway discovered three new finite simple groups, usually called .1, .2 and .3 (the sporadic finite simple groups all have strange designations). These are all subgroups of the group of symmetries S of *Leech's lattice*, which is the lattice of positions of the centres of the "spheres" in the densest known packing of spheres in 24 dimensional Euclidean space. The packing problem is the generalised abstraction of the everyday problem of efficiently packing many identical objects into a space. In his engaging historical and expository monograph¹⁶ on these and related subjects, Thompson gives some insight into how Conway did this, by reporting Conway's description to him of how he went about the task he had given himself of investigating the group S . There were various reasons for thinking S might contain new simple groups. At the time, the project of classifying all finite simple groups was a high priority in group theory¹⁷. The points I wish to make about this discovery of Conway's, concern the one evening on which Conway achieved the essential breakthrough. It was indeed a turning point, but one twelve hours long; and made not in the mind's eye but on paper.

It had taken some time for Leech to persuade group theorists to work on his lattice. But when Conway did "*look at it properly*" he became convinced that a new group was there. Thompson describes how Conway locked himself away from his family for the first of what he expected to be many regular 12 hour Wednesday and six-hour Saturday sessions of work on the task. The first thing he did was "*writing everthing he knew about the problem on a large roll of white paper*"¹⁸. He carried out a lot of calculations (on his paper) and became convinced that the order of S (the number of elements in it - an important clue to its possible properties) was a certain

¹⁶ Thompson [1983]

¹⁷ It has now been completed - but the proof of the classification is about *ten thousand pages long*. This result raises lots of considerations about the nature of proof which I will go into elsewhere. It is obviously something of a problem for Intuitionist and similar idealist interpretations of mathematics.

¹⁸ Thompson [1983: 121] My emphasis.

number or else perhaps its double. He telephoned John Thompson, the world's foremost group theorist, also then at Oxford, and dictated the number to him. He says

*"We used to joke that, if you wanted a new group and if you dreamt of an order, all you had to do was to pick up the telephone, dial 61200, ask for John Thompson, and then **dictate** the integer. The results could be quite spectacular."*

The point of this part of the anecdote is that the orders involved are very large. In Conway's case it was 8316553613086720000. Naturally it must be "dictated"! Dictation is a speech act which requires that certain text acts be effected for its felicity. Admittedly this number can be given in prime factorisation, which is more indicative of its origin and slightly more memorable, but it is pretty clear that anyone planning to think about such a number will *write it down* - or risk wasting months thinking about a different number of similar digits but no interest.

From this number Thompson and Conway inferred the orders of three possible new simple groups. But to show that they were more than possibilities requires demonstrating their existence, that is finding concrete descriptions of them. You don't list the elements of such huge groups of course, you give a set of generators.

*"I started **writing down** a 24 by 24 matrix, filling in the entries piece by piece. ..But I couldn't quite **see** how to prove that the group generated by this matrix, and some others, was the group I was interested in..." (my emphases)*

He rang Thompson again to tell him of the matrix. Soon he realised that all he had to do, was to "**write down** a list of forty vectors such that if the matrix fixed those as a whole, then it was all right. .." (my emphasis again). He did so, ringing Thompson twice more in the process, checked them, and was done. "*Twelve hours had changed my life*" (the result made him famous).

Without pushing the details too far, it is pretty clear from this story (and a glance at exhibit #11 emphasises it) that this discovery could only be misleadingly be pinned to a single moment. It was part of an ongoing social process of investigation. The large roll of paper is not without significance too - it indicates

the need to get the various facts together - just so are the ideas brought together whose meeting Hadamard finds so mysterious. The actual process of Conway's invention was textual - it all happened on paper, the phone calls grounding the written signs about which the mathematicians thought.

§123 Novelty, invention and application in mathematics

To develop the viewpoint expressed in the last paragraph it is necessary to escape the cult of the new. From the point of view of the logico-formalist hegemony: whether a result is new is of "only" pragmatic interest. I am happy to concur to this extent with the logico-formalist hegemony, that the process of making mathematics is much the same, whether the result be a radical novelty created by a proleptic genius like Gauss on nonEuclidean geometry, or the faltering attempts of an ordinary schoolgirl to solve a problem which is no more than a routine exercise from the instructor's point of view. (Perhaps the logico-formalist hegemony does not explicitly include this thesis, but it should.) What is the same is that some novel construction must be arrived at by a more or less trial and error process of bricolage with existing text fragments.

The phenomenon of *multiple invention* is another argument for regarding novelty as relative. Wilder has examples and discussion of this phenomenon, which has been well remarked in studies of science as a process. The most celebrated case is the simultaneous invention of the calculus by Newton and Leibniz (and Fermat, and others, depending on what you call 'the calculus'). The point I wish to make is that it greatly diminishes the plausibility of the lonely genius notion of creation to see how necessary in their context inventions generally are. Multiples are not coincidences, but the predictable outcome of a process which is fuelled by the state of its archive, the training of its participants and the setting of its problems as social processes kept at a common "research frontier" by a complex communication system. If Conway hadn't done it, someone else would have.

Despite this plea for a less magical conception of invention in mathematics, I do not wish to deny that some additions to the archive have far more significance, more consequence, more uses, than others. The theme of this thesis is that the greatest of all novelties for mathematics, the founding invention of the tradition, is that of the writing system. Perhaps the second most significant¹⁹ is the introduction of print. Within these great acts of our cognitive drama there are many levels of scene and subplot. One form of division is the analysis of the various codes given above, and a more detailed and chronological account would demonstrate (in enormous detail) how particular contributions to these systems were made in the course of solving mathematical problems, and how they contributed to those solutions. From this point of view the salience of some contributions to mathematics is clear: the Cartesian invention of coordinate geometry is most obviously a novelty of the first order in the sign-system, whose ramifications were felt for centuries. Its notational character is evident when contrasted with earlier uses of coordinates by the Arabs, and the algebraised geometry of the Greeks - it builds on and could not be done without Vieta's algebraic sign-system.

At the other end of the scale of novelty is Hansen's paper, which introduces no obviously enduring new concept. Yet even here there is a novelty in the paper and it rests on a notation: a notation waiting to be introduced. The main concept deployed in this paper is that of a *strongly regular ring with the property that $x^2 \geq 0$ for all x* ; if this turns out to matter, such rings may well eventually come to be called *Hansen rings*. This will happen if Hansen's paper is often cited in the subsequent literature. There are many examples of this process - Galois groups, Dedekind domains, Liouville's number, Avogadro's hypothesis. Each of these, and there are thousands more, is a stone in the enormous mosaic of the mathematical sign-system. Of course, none of them is particularly important and most could be dropped without much loss; but I have warned repeatedly against drawing from

¹⁹ Which I have largely ignored for reasons of space.

such a protasis any apodosis which would impugn their collective importance. This kind of novelty is semiotic too.

Applying mathematics requires novelty too. "Applied mathematics" is full of novelty in the same sense Conway's work is novel. But even mundane application of mathematics, the use of standard techniques, generally requires novelties to cope with specific features of the situation. Consider exhibit #4 again, CS174. What counts as someone attending a rock concert? Do freeloaders outside the stadium count? What about "friends of the band" who don't pay? and so on. How do we decide if the normal approximation is good enough? Such novelty tapers off as the application becomes more and more routine. In the case of exhibit #1, the use of multiplication seems to be novel in only the thinnest sense - a new case for multiplication, but just like lots of preceding ones. But any new case *may* turn out to have surprising and important features which upset our routines, so in a sense we are always *deciding* to use the particular mathematics in the particular case. Wittgenstein's notion of application takes this idea seriously. The alternative viewpoint which I shall contrast with it in the next few sections does not see this essential freedom in our praxis.

§124 Two accounts of application

Körner points out that traditional philosophies of mathematics have little to say about application: Intuitionism ignores the question, (Körner can find only theories of Kant's and Weyl's about rational mechanics), Logicism virtually ignores it ('1 apple plus 1 apple is 2 apples' is just a logical truth about apples), and Formalism relegates it to the limbo of pragmatic issues (various calculi are acceptable for various purposes, but the reasons are beyond the scrutiny of mathematicians and philosophers - bit grubby probably), not really part of philosophy of mathematics.

There are two possible kinds of account of the application of mathematics, distinguished by whether they take pure mathematical sentences to be descriptive.

If the sentences of pure mathematics are taken to be descriptive (of platonic objects, usually), then their relation to empirical statements of similar form is explained as one of *idealisation*. Alternatively, if it is denied that sentences of pure mathematics are descriptive then they may be interpreted as *regulative* of the related sentences of applied mathematics. Wittgenstein's account is of the latter kind, and I shall argue that though largely correct it must be supplemented by attention to the pragmatic presuppositions of measurement.

The crucial link between exact and inexact concepts which this term 'idealisation' conceals is *measurement*. I shall argue that social practices which support the existence of measurement are the source of the normative force which Wittgenstein correctly attributes to sentences of pure mathematics. The prevalence of platonism tends to favour an approximation account, but the assimilation of mathematics to logic and the general interpretation of logical laws as normative tends to favour a regulative one. Logicians being on the whole platonists, it is not surprising that they avoid much consideration of the application problem which would present them with a dilemma. The question how logic is applied itself has no very satisfactory answers anyway. For logicians, formalists and logical positivists alike the underlying picture is that having chosen to apply a certain calculus to a situation, the laws of logic compel us to certain conclusions and not others. For logical laws are for them descriptive really, describing the logical facts about what follows from what. Otherwise, what hold have we on truth? The classic argument that conventionalism (the early form of the regulative idea) cannot support any truths is Quine's "Truth by convention". It is commonly taken (for example by Putnam and Benacerraf) to have completely refuted the regulative idea. Which leaves only the approximation idea (and explains Quine's ambivalence about "revising logic": he admits it in principle but can't really imagine it).

In §§124-7, I will first consider the relation of vague and exact concepts. The detailed analysis of "idealisation" through measurement reveals in spite of itself that the use of mathematics beyond counting rests on specific social practices,

techniques and mathematised artefacts. So my account of context is vindicated on this score. Then I turn to the use of arithmetic without measurement, and evaluate Wittgenstein's account of the normative element in the use of mathematics. It has been attacked in several ways, and I defend it by using my account of Notation to explicate and reinforce the doctrine that mathematics is not descriptive, and using my account of Readership to explicate Wittgenstein's "rule-following considerations".

§125 The sources of vagueness and precision

The approximation idea of application is that we apply a calculus to some phenomena by observing or deciding that the calculus approximately matches descriptions of the phenomena. The match is approximate, because $8.5 \cdot 14 = 119$ exactly, but measurements of time worked do not yield 14 hours exactly. Let us accept for the purposes of discussion that the problem of application can be discussed in terms of the connection between vague and sharp terms ('14' is sharp, 'hour' vague); but I shall argue that this involves looking at the juxtapositions of various kinds of artefacts, practices and texts, not just ethereal ideas. I also reject the assumption of the literature on vagueness, which seems to be that we understand sharp concepts, but vague ones are under a cloud because they support sorites paradoxes²⁰. It seems to me that we should invert the questions usually asked about vagueness. The onus of explanation should be on precision not vagueness, for vagueness is endemic in language, precision a late, local and specialised modification. The question is not whether vagueness is eliminable from language but rather how does sharpness become possible in a language, not is there a special logic of vagueness but rather how is the logic of sharpness special?

Körner's idea of the relation between pure and applied mathematics is that empirical concepts (inexact or vague concepts) are *idealised* into precise or exact

²⁰ In general I reject the idea that the construction of a paradox identifies a point of peculiar importance in our conceptual network; rather, I see it in the fashion of Wittgenstein as a sign of our trying to take language into inappropriate situations.

concepts, which are manipulated within pure mathematics and then concretised back to empirical interpretation²¹. Körner adumbrates a "logic of inexact concepts", as in need of development after his introductory discussion pointing out logical differences between exact and inexact concepts²². But it seems to me that despite various recent efforts to provide such logics, the interesting question is being avoided. The interesting question is, how is the idealisation made possible, and how is it made? Work on the logic of vagueness, while interesting, is irrelevant to the present enquiry. It is the intervention of measurement practices which provides the link between the sharp or exact concepts of pure mathematics and the vagueness (or inexactness) of empirical predicates. And, as I shall show, measurement is a textualised praxis. I am therefore proposing a materialist explanation in place of the idealist account of Körner.

The prevailing tone in discussions of vagueness²³ is one of gloom and regret, a rather strange attitude. The idea is quite general, that vagueness *infects* language from without, spoiling what would otherwise be a nice little tool for doing mathematics. Wright says²⁴ that we no longer have this Frege-Russell view of vagueness as a defect, but I seem to see it in most of the literature of the philosophy of mathematics still. The idea that a language could be entirely vagueness-free is very widely taken seriously by formal logicians, even if "natural languages" are not. Vagueness is a feature of language which has never been taken very seriously. Yet there is no example of a real language without vagueness, but hosts of them without any exactness at all. Why is the onus on the vague to justify its place in language rather than the sharp? I believe that an explanation of this disparity would be very

²¹ Cf Körner [1960:158ff]

²² It is rather extraordinary that his observations are new in 1960.

²³ The earliest are Russell [1923] and Black [1937]. More recently there is work provoked by Körner, such as Mellor [1966], [1967], and Fine [1975]; Sanford has written a number of papers pursuing a program parallel to that of Fine, for example his [1975]. Other interesting discussions concern the sorites paradox: Cargile [1969], Campbell [1974], Unger [1979]. There is also work on fuzzy logic, deriving from Zadeh [1965], for example Goguen [1969], Gaines [1976], Priest [1979], the latter applying it to sorites. Some good general discussions of vagueness are Margalit [1976] and Scheffler [1979].

²⁴ Wright [1975: 325]

revealing, and that it must be sought in the explanation of the power relations mediated by language. But we cannot pursue it here. I will restrict my argument to the claim that vagueness is necessary in language, while precision is useful but not necessarily achievable.

We must consider the source of vagueness if we wish to decide whether it is an eliminable affliction of language or an unavoidable element of language, a necessary mixed blessing. In the literature it is generally claimed that the source is 'empirical predicates'. Another source we might consider is proposed by Margalit²⁵ the adaptability of natural language to all purposes. This is a distinct argument, and in my opinion an equally compelling one. Such examples as the color sorites are what convince most philosophers that, regrettably, vagueness is an unavoidable nuisance in language. The argument is that we must have terms whose application can be decided by just looking; our lookings have a limited power of discrimination; many things vary more finely than we can tell by looking. These three together create the possibility of the 'practical sorites' such as the color sorites²⁶. If a language is to be used by finite creatures, there will be discriminations too fine for them and if there are to be terms for the items entering into these comparisons, the color sorites situation will obtain. And why must there be terms for such items? Well it seems obvious that creatures like us need them in some evolutionary sense to deal with the environment, and that we enter language as individuals at least by learning such terms. Arguments related to this one are given by Dummett and Wright to conclude regrettably that terms whose application is to be decided merely by looking, so terms whose use must be learned through processes involving ostension, will necessarily be vague. Speech seems to be *based* on vague terms.

Could they all be removed by precisification? It seems to be a presumption of many writers that vagueness is eliminable from language. Frege and others have bent their efforts toward a logically perfect language from which vagueness would be

²⁵ Margalit [1976: 215]:

²⁶ See Dummett [1975b]

entirely absent; Russell²⁷ tells his audience that his own notation is just such a one, calling it a language. Haack²⁸ finds no objection in principle to "Carnap's program" of the entire elimination of vagueness from language through precisification²⁹. Fine supposes in the course of his discussion that there might be a finite basis of vague terms that could be "precisified out". Even the most ardent theoretician must grant the need for observation language as part of scientific language. Now although observation reports may be couched in the new precise language after precisification, the setting up and conduct of the observations, the manipulation and evaluation of the results, and the communication of the findings will involve the use of considerable parts of the vague, not-precisified language. This is obvious on any glance at scientific literature. The point is simply that the conduct of science requires scientists. Russell is oversimple, it seems to me, in claiming³⁰ that 'a metre' is a vague term because the specifications of the standard metre are themselves vague. It is perfectly true that they are; nevertheless when we make a judgment that my height is 1.95m, this is not a vague claim. The vagueness has been *shifted* to the apparatus for the control of metre and related terms and removed from the scene of their application. This is another example of the local/global point which I have made a number of times: it is always possible to adjust our sign-systems so that particular signs or texts play a particular kind of role, but there will generally be compensating adjustments needed elsewhere which prevent the complete generalisation of the first alteration.

Our language begins in vagueness and there is no reason to think that it can be eliminated entirely, or even in great measure. Therefore the crucial question is that of the origin of precision, not that of vagueness. This leads us to consider the logic of measurement. But first I shall briefly consider the idea of logics of vagueness - though my conclusion is that it is a red herring.

²⁷ Russell [1923: 84]

²⁸ Haack [1978: 164]

²⁹ Carnap [1950]

³⁰ Russell [1923: 86]

§126 The logic of vagueness a red herring

There are two programs in the literature for the taming of vagueness, which I will describe in a moment. They are both mistaken in thinking that it needs taming, or can be; and they are wrong in presenting themselves as alternatives. Both have something to contribute to our understanding of idealisation, though neither by itself will help us much; rather, each alone is likely to confirm old illusions. They both suffer from the *replacement* conception of semiotic work; as always, my view is that we must see it as *extending* our semiotic resources. On the one hand there is the attempt by Fine (and independently by Sanford) to use the supervaluation approach to make more definite the solution proposed by Dummett to the sorites paradox³¹. The other main approach has been through the development of fuzzy mathematics by Zadeh and his followers. The Fine approach has the advantage that it is based on the concept of *sharpening*, which is *prima facie* the process which underlies concept-formation in science. On the other hand, it has the disadvantage that it declares sorites invalid. The fuzzy approach, which is based on the idea of giving sentences degrees of truth between 0 and 1, has the advantage that it sits well with one's intuitions about the acceptability of the statements and inferences in a sorites, of its gradual diminution. On the other hand it leaves language as it is and so it will not contribute to the question of the origin of precision.

If our "problem" was to resolve the puzzle posed by sorites, I would adopt fuzzy logic, along the lines of Priest³². In fact, though, I think that such paradoxes are only interesting because their consideration forces us to consider vagueness and its role in language more closely. Consequently it is necessary to give more consideration to the Fine approach. The main weakness of the Fuzzy program is that it neglects the enormous profits that have derived from precisifications in the

³¹ In Dummett [1975b]

³² Priest [1979]

past. Correspondingly the main strength of the Fine logic is its recognition of the possibility of precisifications. How can we combine the correct insight of the fuzzy view of the logic of vague terms with the Fine recognition of the value of precisification? My suggestion is that we can do so if we renounce any attempt at *global* reform of language, as well as any temptation to regard fuzzy logic as making bivalent logic redundant. Vagueness is not eliminable from language, so that the fuzzy logic is necessary. But we shall be often well advised to seek precisifications, so that the bivalent logic has its place too.

The Dummett/Fine idea is this: if an empirical predicate is replaced by a precise one, sorites no longer arise. Various precisifications are possible and we can repudiate sorites by relativising such inferences to the set of all possible sharpenings of the term on which it is based. A vague argument is valid if the conclusion comes out true on any sharpening of the premises to truths. But just what is a sharpening? Fine uses the term 'precisification', deriving from Carnap on concept formation in science. Fine calls it a "conservative meaning change" and "the elimination of truth value gaps". So declaring that the small numbers stop at 100 is an acceptable sharpening for him. But *is it* an acceptable precisification to just declare that 'small number' means ≤ 100 ? From the purely classical logical point of view, where the aim of sharpening is only to eliminate truthvalue gaps, perhaps. But what Carnap³³ means by precisification is the replacement for scientific purposes of a classificatory concept ('small') by a comparative one ("smaller than") or, preferably, a quantitative exact counterpart ("as small as n"). Since Fine's quantifications are over *complete* precisifications, one must suppose that Carnap's less attractive comparative explications won't do. But will the kind of arbitrary line-drawings which Dummett and Fine contemplate do either? Carnap³⁴ is quite explicit that the possibility of the more desirable quantitative replacement depends on our discovering a suitable *technique* for measuring it. Notice also that

³³ Carnap [1950:1-18]

³⁴ Carnap [1950:14]

he finds the actual explication of 'fish' used by scientists to be preferable to others that are possible by virtue of permitting "more general statements". So not every logically possible sharpening is a possible improvement to language from the point of view of the reforming scientist. Why should pointless possibilities be included in the set over which the quantifications are made in settling the validity of sorites? One may also doubt, if one is impressed by the concept of open texture, whether there are any complete precisifications at all. As Gaines³⁵ says "The ultimate precisification is to treat every event in the world as different from every other - which it is!" The fact that we think of length as measured on a scale including all real numbers is actually an idealization of the state of measuring technology at any given time; no precisification that has been carried out is complete in Fine's sense. Precisifications are not done in the way Fine and Dummett suggest! So it seems to me that contrary to the impression one gains from the literature, the Fine/Dummett idea of sharpening is not the same as that which Carnap, Hempel and others have used to analyse the actual course of meaning change in science. This reinforces Sanford's question: not only have all these possible sharpenings not actually been made, no-one in their right mind would make them either. Why should possible but unactualised sharpenings be relevant to the validity of sorites?

Thus the "logic" of vagueness is something of a red herring for us, and we shall do better to look directly at what is involved in measurement - real sharpenings. I shall argue that real sharpenings are bound up with a technique and social practices, and therefore have pragmatic presuppositions.

§127 Kyburg on measurement

There is no good reason to suppose that the scientific part of language could do duty for the whole of our real language. It is commonly suggested that a linguistic innovation such as precisification through measurement of 'red' or long' is a replacement - and with something better. I have argued already that this is quite a

³⁵ Gaines [1976: 629]

misleading idea. Such innovations *add* to the language without any corresponding *subtraction*: 'red' is still a common term despite our sophisticated talk about the wavelength of light; 'long' is still a very common term, often used about texts for example, long after the precisification of the concept of length occurred. When precisification occurs it occurs through the *extension* of the language; the original vague terms remain. But an extension cannot occur without concomitant mathematising action. Instruments must be created for the assignment of quantities, and used. And that won't be done without some impetus: it has to be worth someone's while to do the work involved. I will now show how such pragmatic requirements are disguised in even the best attempts to describe the logic of measurement.

Kyburg³⁶ has provided us with the most convincing account to date of what resources a language must have for quantitative measurement to be possible. By a detailed argument Kyburg builds up the resources which a language needs for him to be able to prove suitable theorems along the lines of Krantz et al, to show that a term can be used for extensive measurement. He begins with predicates for 'rigid body', 'collinear juxtaposition' and 'longer than'. (Notice that we start with a comparative, not the plain vague "long".) These provide the basic sentences which are of interest, of the forms 'b1 is longer than b2', 'b3 is a RB' and 'b4 is b5•b6'. An axiom of distinctness and an assumption of a 1:1 naming relation complete the first level of his eventual measurement language, L1. So far no interesting inferences are possible. Adding axioms for asymmetry and transitivity for 'longer than' admits the possibility of erroneous sentences 'RBb' so that consistency requires some rejection rule; Kyburg proposes that of minimising the number of rejections. Thus the second level language L2. Inference to sentences not personally observed is now

³⁶ Kyburg [1984]. There are at least two advances in this work over the best of the previous literature, such as Krantz et al [1971], Ellis [1968] and Campbell [1957]. First, Kyburg recognises the need for two terms for comparing lengths, one precise and one not, in the measurement process; thus he advances beyond replacement ideas to extension ones. Second, he builds the statistics of error into the logic of measurement from the beginning. (I shan't use this second feature of his account.)

possible. At the next stage it is allowed that *any* observation may be in error and some axioms about juxtaposition are added which require a more sophisticated rule for rejection, that of making the rejection rates for sentences of different kinds as equal as possible. Many more sentences can be accepted now, but none are incorrigible. This is language L3; to get to L4 we must ensure that 'not longer than' is transitive and to do so requires introducing *new* predicates 'is truly longer than' and 'is truly equal in length to'. These are the exact terms which supplement the inexact terms 'longer than' and 'is equal in length to' originally in the language; suitable axioms are added for them and equivalence classes of objects under the second are formed to provide us with lengths. We also need to acquire a standard unit b_s and assume a number of axioms, particularly one of integral subdivision for rigid bodies. Adding suitable operators to permit combination of lengths, Kyburg shows that we now have *extensive measurement*, in the sense that the calculus of Krantz applies to 'is truly longer than' assuming Kyburg's axioms, but that there is no cognitive gain because there is no connection between the two concepts of 'longer than'.

To effect one, and arrive at true measurement language L5, the relation 'x is indistinguishable in length from y' is used. For each measurable object, there is a class of other objects which are indistinguishable from it, but this is not made an equivalence class (by using the ancestral of indistinguishability) for the reason that to make it so would introduce sorites paradoxes³⁷! The measurement relation is defined in terms of the indistinguishability relation and the standard unit in an intuitively obvious way: it assigns that rational number m/n to a rigid body when it is indistinguishable from m juxtaposed copies of an n th submultiple of the standard unit. It does *not* assert that the body is in the equivalence class $m/n * [b_s]$. Error is defined in the obvious way (difference between actual length and observed length) and Kyburg goes on to argue that his apparatus will now permit us to assign confidence intervals to the values of those quantities in which we are interested, in

³⁷ Kyburg [1984: 71]

the way which is actually found in scientific practice. The details of all this are extremely complex and I have tried to give only enough to indicate the nature of the discussion of measurement in the literature, and to motivate some points I wish to raise about it all.

First, although Kyburg motivates the introductions of the various bits of linguistic apparatus by promising cognitive gains, it is not made clear why they are *gains*. Second, the apparatus introduced is not all linguistic: the "strong existential assumptions", as he calls them, are a way of getting *measurement technology* into the process without considering what it presupposes. That is, axioms are assumed which can only be true if certain technology is in use. Third, we must be in a position to actually *effect* the inferences from the observations which are the point of these manoeuvres, but these are hardly mere syllogisms in Barbara. They require considerable calculations. Fourth, it is necessary to notice what Kyburg's account is capable of explaining. As he has it, the key connection between the calculus, which the exact concept obeys, and the inexact concept, which is used in making reports of empirical judgments, is the concept of error. Using it, defined through his formalism, he can show that present scientific practice in predicting confidence intervals for real lengths is rationally justified. Or at least, that is his claim. But even if it is correct, this cannot show us the essence of application, for lengths were measured and mathematics used on the measurements long before science. Theory of errors is actually only a few centuries old.

I shall discuss these points in the next section.

§128 Measurement's pragmatic presuppositions

In the last section I pointed out presuppositions in Kyburg's account of the logic of measurement, presuppositions about the pragmatics of measurement. Here I enlarge on what they involve.

First, cognitive gains. It is apparent that precisification requires considerable work, not all cognitive. I think precision can only arise where there

is a need. When is a sharpening a useful one? What kind of gain may ensue? Kyburg concentrates on, in essence, prediction, but this is to read into all of mathematics the later developments of science. The first gains from measurement were quite different. To introduce the integers into talk of baldness would destroy the point of such talk. A precisification of baldness would actually serve no general human purpose; but if it did, then we would seek a measure of hairtness-of-head³⁸. There *was* historically good reason to sharpen the concept of heap to that of the bushell, or of the pace to the yard and so on, since the organisation of complex social events in early civil society {war, government, property in land, taxes etc} depended on being able to deal with many judgments of distance or amount, treating them in a way which made their *author* irrelevant. The *authority* for them was externalised. The reason we cannot make a sorites based on 'this length is 1.95m' is that what is judged to match a standard does *not* thereby become a standard too³⁹; so there is no scope for chains of "metres" of gradually decreasing length, on the model of a colour sorites. The adoption of a standard is a change to our practices which bites deep; with it we can refuse to be bullied into sorites because we can appeal to an external authority - we are no longer compelled to "make it all up ourselves" from how it looks. There is a valuable point to the elimination of practical sorites, for example clipping coins. But granting that the idealizing Rule of Bivalence is a very powerful tool, and supposing we understood it, that does not make it the be all and end all of linguistic labor. Poets don't want a logically perfect language, and nor do people concerned to gloss over the conflicts of interest of real life, that is, civilized people and their proxies in public life. Popper has urged that precision, like certainty, is an aim which should only be pursued ad hoc, not for its own sake. Precision may well

³⁸ This would not be hard. Much "empirical" social "science" consists in doing that, to no real purpose because there is no real consequent mathematisation of the world. Partial exceptions are IQ and the opinion poll.

³⁹ This is the real point of Kyburg's not introducing equivalence classes of indistinguishables.

be obtained at the expense of clarity. For as Gaines says⁴⁰: "it is dangerous to assume certain forms of precision when they do not exist in the world".

The second presupposition of measurement is the establishment of a system of standards. What does establishing such a system require? It involves considerable praxis. We need to establish a paradigm length, and a way to make copies of it, and a means to count copies, and make and count submultiples. We must have a technique for matching copies to parts of lengths. All this requires us to have a technique for marking copies of submultiples of the standard, and techniques for juxtaposing and counting them: rulers, and mathematics in the form of simple geometry and arithmetic⁴¹. Despite Kyburg's sophisticated idea that the link between exact and inexact terms is the notion of error, the actual link is made by *using rulers*: As he says⁴²

"When we measure a rod with a meter-stick, we have k centimeters, all handily juxtaposed in a rigid body, and each of those centimeters is a handy sequence of juxtaposed millimeters. The observation is not as complicated to make as it is to analyze."

But there is a confusion here: the act of looking at the juxtaposed ruler and rod is in a sense simple - but really it is just as complex as his analysis (else the analysis must be wrong !) To make a measurement presupposes making the ruler and its user too: and these are complex processes indeed. Measuring depends on being a reader; rulers are made only in mathematised societies.

There must be further social practices: a unit is determined by reference to the paradigm; and it must be maintained by some external legislative power. The standard metre is conventional but not in the "gentleman's agreement" mode which 'convention' often evokes, and which may once have governed the use of words. (It does not anymore, for we now have similar external enforced authorities in dictionaries, schools etc.) Exactness holds; but it holds because we collectively

⁴⁰ Gaines [1976]

⁴¹ And what in turn are the requirements for that? I have argued already that it presupposes the master technology of writing.

⁴² Kyburg [1984: 72]

make it hold. There is a social pragmatics of standards. When we create a realm of exact lengths, we commit ourselves to certain new patterns of behaviour. This committing ourselves, though, is not the free and equal consent of Kantian wills to a social contract. On the contrary, laws backed by force are involved. The second presupposition then is a certain technique and the social practices which produce and maintain it. So, precisification is good if it works - if there really are indiscriminable differences and we can find a way to create a measurement technique. Otherwise we are better off, or may simply be limited to, using fuzzy logic. But precisification depends on certain technical, material possibilities, not only the logical ones usually discussed in the literature on measurement; or rather, logical ones disguise pragmatic ones. For length they existed in antiquity; for intelligence, perhaps not yet; for wisdom, certainly not.

The third kind of supposition of measurement is related to the need for some gain. The gain must be realised if it is to remain a motivation. For the rationalisation of large numbers of observations, made by different individuals at different times and in various circumstances, it is not enough to create a standard and technique for making rulers, nor yet to train people in their use. It is also necessary that the results should be processable and processed. This entails a method of recording and archiving them, and relevant techniques for carrying out the calculations which produce the gains we are after. I have pointed out in Chapter 5 that the earliest uses of writing include just such recording of measures and calculations based upon them for accounting purposes. The earliest mathematics was certainly developed to support such activities. Each advance in measurement brings concomitant needs in archiving, processing and training for the users of measurement.

Empirical predicates must be vague, mathematical ones precise. How are they brought into connection? By willed juxtaposition. Juxtaposition, in that exact concepts do not replace inexact ones but extend them through textual accumulation. Willed, in that exactness is a demand or convention, not a fact of

nature. There will not be measurement of length without these pragmatic desiderata, for people simply will not be engaging in the practices of measurement; the idea that the terms could be "in the language" without being used in practices of these kinds makes no sense. The danger of "logical" analysis of concepts is that it disguises such facts in axioms. Kyburg's analysis of the logic of measurement may be spot on, but it cannot help us with the real problem of application because it ignores the pragmatics: how is it possible for us to make the linguistic changes which he shows us in sequence? It is possible only because we embody the earlier stages in instruments used in basing the later ones. Foundations are needed in the material world, not in some crystalline logical sphere!

The upshot of this discussion is two fold. First, I have argued that underneath the abstract formulation of the logic of measurement we may discern a necessary foundation of material social practices, skills and artefacts, part of the context of mathematical work which I have discussed above. Second, we have as yet no explanation of the normative element in applied mathematics. To formulate a calculus which describes extensive measurement is not to explain why conclusions arrived at with it are binding, although calling some of it axioms certainly encourages the seeming ineluctability of logic. Third, we see from the central situation of measuring as described by Kyburg that if we *don't* neglect the context, the ruler and its educated users, we might view the key act in the application to be the counting of millimeters. In the following sections I will examine Wittgenstein's account of how mathematics imposes normative force, by considering only the simplest applications of arithmetic. His discussion, and that of his critics, tends to ignore the contextual factors which I have pointed out, but we shall bear them in mind throughout.

§129 Wittgenstein's conventionalism: following rules

Wittgenstein's idea about application, of mathematics or of logic, is a conventionalist one. Mathematical sentences are not descriptive of the world, but

regulative of our descriptions of it. Most conventionalisms turn this into a different kind of description, description of the logic which governs our applications. This is a subtle and pervasive error which Wittgenstein strenuously combats. The conventionalist picture is that the rules which mathematics consists of are arbitrary adoptions of ours which guide our applications of mathematical concepts. The application problem is fundamentally this: if a worker is paid \$8.50 for each hour worked, and works for 14 hours, then the worker must be paid \$119 altogether. 'Must be', because if you multiply 8.5 by 14 you *must* get 119 - if you do it correctly. Yet we all know that mistakes occur, perhaps the worker actually gets paid \$117. Why do we call it a mistake? Because the correct answer is \$119. We know that because $8.5 \times 14 = 119$. Now the question is, how does this last fact bear on our doings? How does 'must' come into it? The regulative answer is that ' $8.5 \times 14 = 119$ ' acts as rule which constrains what we may do with the signs '8.5' '*' and '14'. Something like this: in a certain situation, '8.5' and '14' and '*' are correctly used, and the rule is that $8.5 \times 14 = 119$, so '119' is correctly used. This is meant analogously to, or as a case of: p and $p \rightarrow q$ are acceptable, and we have a rule of modus ponens ($p \& p \rightarrow q \Rightarrow q$), so q .

Lewis Carroll's objection holds in both cases: why must we conclude Z , from X and Y and a rule that $X, Y \rightarrow Z$? Do we not need a rule that $(p \& q \& (p \& p \rightarrow q) \Rightarrow q)$? And so on. This standard objection, given also by Quine⁴³, is that explicit adoptions of conventions cannot be numerous enough for all the infinitely many implications of logic. These must therefore be *drawn*; but to draw them requires the use of logical rules which must precede them.

Wittgenstein's novelty in conventionalism was not intended to meet this stock objection, though it does. Rather, it was aimed at the incomprehensible picture of being guided by rules which informs the debate about the matter. It is common at present to discuss, at impenetrable length, Wittgenstein's "rule-

⁴³ Quine [1936]

following considerations"⁴⁴. But the debate seems to me to centre around the wrong point. It is not really the point whether Wittgenstein was proposing a sceptical paradox, as Kripke claims and most others deny; the point is that he was proposing an alternative viewpoint on how rules transmit their normative force in actual praxis, alternative to the *magical* ideas which Wittgenstein considers in the parts of his *Investigations*, and other writings, which precede these problematic passages.

Returning to the multiplicative case, *how* does the rule oblige us? Wittgenstein argues that the usual answers to this question are all unsatisfactory, whether we call on intuition, on the rule guiding us like words when we read, or as something we explicitly consult every time - none of these actually matches what happens and how we talk about it. He is forced to consider the idea of following a rule down to its simplest case, that of continuing a series, and even the simplest kind of series 1,2,3... and the simplest series of all 1,1,1,.... How does the simple rule 'go on in the same way' oblige us to write down '1' every time? For what does 'the same way' mean?

If we examine how the use of such a phrase is learned, we find that only partial examples and instructions are given, people have to 'catch on'; if they don't, we say that they don't yet understand the words. But the evidence of catching on is not some black and white litmus test, but *agreement in praxis* with what we all do already. The point of the rule-following discussions is this: there is *no transcendental backing* for rules, no ghostly commandments lying behind them acting on us when we have grasped them. It is true that after 1,4,9... the next number is 16 if squaring is meant; but this shows only that if some other number is produced then squaring wasn't meant, either because something else was meant or because the person is muddled about squaring. In the end, "meaning squaring" comes down to conformity in action with the general practice of squaring. Given a rule, either by example (1,4,9,16...), by formula (x^2) or by some more complicated

⁴⁴ See Wright [1980: passim], Kripke [1982], Baker and Hacker [1984c], McGinn [1984]. There is a good deal more besides.

method, I show that I understand the rule if I can go on applying it, that is in this case, if I can continue the series. The rule determines how I go on, in that there is only one correct continuation, but we must understand what this means. It does not mean that I am somehow compelled to write down 25 next, or that 25 is written down faintly somehow for me to trace over, or that the look of the formula somehow guides me in writing down 25. Wittgenstein considers suggestions such as these and shows in detail how they are untrue to the way we talk about the situation⁴⁵. It means that if I do something else I am not following the rule, that is I am not using the concept of squaring as we presently have it. My deviation might be so advantageous that we change our concepts, and such things happen. That would not show that we all presently have the wrong concept of squaring however, that does not make sense: the concept is *constituted* by our agreement in such continuations, only occasional aberrations, accepted as such, are possible. If there were general disagreement, we simply should not have a workable concept here. The idea that the concept of squaring exists somehow whether we have got it or not, Wittgenstein regards as a bit of mythology. It arises from the attempt to treat following a rule as only the sum of various individuals' conforming actions, whereas it is actually an essentially social thing, a *practice*. The misleading ideas are all attempts to identify aspects of *me* as that which constitute my compliance, when it is not any such thing that can do so, only features of *us*.

It is characteristic of mathematical concepts that there is a peculiar uniformity of agreement, different to the kind of agreement which holds for other concepts we have. I believe that it was by seeing why this is so that Wittgenstein must have shifted to his family resemblance account of concepts like intention and Moses and number (which is not really a concept of mathematics!), in order to extend his social, pragmatic account of concepts to non-mathematical ones, where unanimity is not the norm⁴⁶.

⁴⁵ Wittgenstein [1953: §139ff]; also [1967: §296ff]

⁴⁶ Ian Hacking has made a similar conjecture in his [1985].

I will consider in the rest of this chapter several ways in which Wittgenstein's point of view has been challenged, showing how the point of view on mathematics developed in Part B above can fill it out. But first I will bring it to bear on the distinction between mathematical and non-mathematical sentences.

§130 Mathematical versus non-mathematical sentences

In order to make his argument for a conventional interpretation of necessity based on his notion of rulefollowing, Wittgenstein draws a very sharp distinction between mathematical and nonmathematical sentences, or rather uses of sentences. This is required for his differentiation of roles for them to make sense. Wittgenstein has a criterion for the difference: the difference between a calculation and an experiment. He repeatedly argues that a calculation or a proof is not an experiment, and that this difference is absolutely fundamental for philosophy of mathematics. He argues that a calculation is not an experiment because an experiment can turn out in various ways, whereas a calculation, by definition, has a unique right answer. Mathematics is normative. So $25^2 = 625$ is being taken as pure mathematics if we brook no possibility of anything conflicting with it. It may be taken as a prediction if I want to see if you can square 25, yet your doing so is an experiment only insofar as we want to see *if you can do it*, not to see what 25^2 might be. Notice that the same sentence can equally well report an experiment, a calculation or a prediction: it is only the *context* which can determine for a particular use of a sentence, which it is.

Wittgenstein's favourite parallel for the role of the mathematical sentence is the way a ruler is used in measurement. The standard metre is not measured, and it cannot be questioned whether it is a metre long; a ruler is held up against reality to measure it. Similarly, a rule, say $2+3=5$, is held up against counting and adding. It would be easy to get confused here and find a disanalogy, in that there are various acceptable results of measurement, only one of calculation. There is a problem with Wittgenstein's parallel, but it is this: measuring is a more sophisticated stage in

counting. It cannot really help us to understand addition to explain how it's done by appeal to something more complex which relies on addition. Still, Wittgenstein's point is perhaps only that the ruler is taken for granted in measurement; similarly, he suggests, the mathematical rule is taken for granted in its use.

While I agree with Wittgenstein's contextual approach to the determination of the role and meaning of utterances, I think that the sharpness of the distinction he draws between rule and application is exaggerated. Many utterances take some of the characteristics of both ends of a spectrum. How else can we understand the process of learning mathematics but as the gradual, not sudden, changing of the status of certain sentences from observation to rule? On his own account of the invention of mathematics, a new item is established as a rule by use: it cannot therefore be functioning as a rule in its early uses. The gulf between the logical and the empirical is an abyss into which many philosophies of the twentieth century fall, because they cannot find any third term to act as a bridge between them. Wittgenstein sometimes seems to be still too far under the sway of Early Russell in this regard. But elsewhere he is clear that these are derivative categories - really, what is given is the form of life. What he lacks is a vocabulary for talking about it which is not dominated by the picture he wishes to oppose.

A better way to draw the distinction which is needed is to distinguish among the signs used in mathematics, as I have done in the preceding chapters. I propose that we identify pure N-signs with Wittgenstein's mathematical sentences, and W-signs which involve N-signs with his non-mathematical sentences. I have argued in Chapters 5 and 7 above that these different kinds of signs have different roles in text, and I suggest that this is a way of making more specific the contrast between rule and use. Moreover, we can get away from the Carrollian problem if we recognise that signs are only signs and what counts is the way they are used: the difference between rule and use is a difference in praxis: what we do with N-signs is *different* from what we do with W-signs. The W-signs are for making descriptions, the N-signs are used for showing designs. The continuum of analyticity in W-signs

corresponds to the extent to which terms in them are being taken as equivalent to N-signs. Thus 'one and one is two' is more or less analytic according as 'one' and 'and' are identified with '1' and '+', the signs used in the calculus, Peano arithmetic.

§131 The programmed formula challenge

It is tempting to respond to Wittgenstein's discussion of rules and concepts by saying "bosh, 25^2 just is 625". But this is not an argument. An argument which seems powerful is this one: Wittgenstein is happy to agree that any sequence will do, so let us consider the sequence of numeral names in base 10. Now it is an easy exercise to write a computer program to print out this sequence, with no limitations except practical ones on how far it goes. Doesn't this show that there is no element of decision in the continuation of the sequence? The program can produce numerals far longer than any ever written down before, yet we are all quite sure that we know what it will do and that we shall agree with it, unless some mechanical fault occurs. Aren't we constrained in just the way the computer is, to follow the rule, embodied in its or our program? Of course, we are not: does that refute Wittgenstein?

Consider multiplying two six digit numbers, say 123456 and 234567, with a hand-calculator: there will be a rounding error, which we know how to adjust for if we know how to use a calculator. The machine actually functions as it is designed to do, but the result, because of the limitations of its design, is not in fact the correct answer. On one calculator I own, I get an error message; on another, the answer 2.8958703 E10. Actually the answer is 28958703552. What the calculator actually calculates is not the function $*$ but a related one which truncates to 10 significant figures (and will only give an answer in the range 10^{-99} to 10^{99}). The *practice* of giving answers correct to a certain number of significant figures arises out of circumstances like these, where the correct answer is too hard to come by or uselessly accurate. This is a change in the concept of $*$ which can be ignored for small pencil and paper calculations and for some theoretical purposes, but which is of great importance for others. Perhaps we might need similarly to expand our

concept of the sequence of numerals if we really needed to produce tokens of big ones for some purpose. But what reason is there to think that the magic formula "in principle" will circumvent such limitations for any actually possible machine?

In any case, the implied parallel is misleading. The proposed program does not show that there is no element of decision in such calculations, because although I don't usually make decisions when I calculate, I just do it in as automatic a mode as the machine, I can and must make decisions in monitoring my output for errors. Where the possibility of a decision arises is in what we do about results. We must decide if the machine is functioning properly, we must decide what to do with output which seems suspect because the machine is running up against physical limitations of time or wear or space to print the answers, particularly if inconsistencies begin to creep in. Eventually the output numerals will be too long for us to survey and we will be unable to check them at all: do we still believe that the results are correct unless the machine crashes? Producing a numeral using a machine does not change anything over using an abacus or one's fingers. This line of argument against Wittgenstein does not work because it begs the question. It assumes that there is in some platonic sense a sequence of numerals which the machine traces over. If we look at the realities of such a situation we find that exactly the same kind of considerations which led to the argument come into play.

§132 Alternative mathematics: is it comprehensible?

Some writers have tried to refute Wittgenstein's conventionalism about mathematical concepts by detailed consideration of the "alternative mathematics" adumbrated in some of Wittgenstein's examples. The classical urging of Wittgenstein's particular kind of conventionalism claim is by Gasking, and Castaneda, Lehman, Wright and Craig have continued the discussion⁴⁷. The argument is this: Wittgenstein and Gasking claim that we could adopt $25^2 = 624$, or $5+7 = 13$, with concomitant, perhaps extensive, adjustments to related concepts and

⁴⁷ Castaneda [1959], Lehman [1970], Wright [1980], Craig [1975], [1985].

practices, and get on with our business, perhaps with some inconvenience. (Our business would have changed a little, of course - or perhaps a lot.) Gasking extracted this argument from his inchoate audition of Wittgenstein⁴⁸ and presented it relatively clearly; the later literature comments on it by more detailed consideration of what exactly the concomitant adjustments would have to be. The essential question is whether stories such as Wittgenstein's and Gasking's can really be filled out without becoming clearly nonsensical. Castaneda, Lehman, Craig, and in some paragraphs Wright, deny this. In this section I will discuss Gasking, Castaneda and Lehman; in the next, Craig and Wright⁴⁹.

Let us call a *Gasking arithmetic*, a calculus which differs from standard arithmetic, by including at least one proposition which would be denied therein, such as⁵⁰ $5+7=13$. Gasking argues that by suitable changes to our practices of counting or measuring or our physics, the differences from Peano arithmetic could be compensated for and all our usual jobs still get done. The one he considers is tiling a floor. Thus for example we could use $6*4=12$ but measure in Ellis' manner⁵¹ and cover the floor, or stick with $6*4=24$ and count as usual but change our physics to account for it if only 12 tiles fit a 6 by 4 room.

Castaneda disputes the tenability of these claims by focussing on what would be involved in using a Gasking arithmetic. He explicates this notion by identifying standard arithmetic (SA) with the five Peano postulates and five principles of counting; then a Gasking arithmetic is one which differs from standard arithmetic in one of ten basic ways, namely by denying one of these ten postulates or principles. He argues that in each possible case this could only be done if the Gasking arithmetic were *trivially different* (employing variant or extra,

⁴⁸ Gasking [1940]; cf his remarks in Fann [1967]

⁴⁹ I do not intend to discuss the recent debate around Kripke's sceptical interpretation of Wittgenstein's rule-following discussions, because I think it is clearly a *misinterpretation*, and that this has been well shown to be so by Baker and Hacker [1984c] and McGinn [1984] independently.

⁵⁰ This is the most favoured example in recent writing, it derives from Kant; Gasking used multiplicative examples. It makes no difference.

⁵¹ A bizarre method is used in Ellis [1968] as well as in Gasking, to make similar points.

useless numerals) or *trivially changed* (different basic operations, e.g containing 2^* rather than $*$), an *extension* of standard arithmetic, or an *inapplicable* one (we cannot count past 8, or face contradictions or ambiguities).

Lehman corrects some slips in Castaneda's argument, and argues that in fact there *are* Gasking Peanian arithmetics⁵², that can be used for counting, but they cannot be used as shortcuts for counting; and there are non-Peanian Gasking arithmetics which may be perfectly useful - as Castaneda in effect conceded when he admits that a Gasking arithmetic might simply extend standard arithmetic. But Lehman thinks that attempts to use the theorems of Peano arithmetic with deviant counting principles will lead to contradictions. We notice that both these writers simply take logic, and in particular the standard attitude to contradictions, for granted. Well, Gasking didn't mention revising logic to make compensations, but there are plenty of writers nowadays who would. Wittgenstein had no great fear of contradictions.

The upshot of the debate to this point is certainly not a defeat for conventionalism; and this is not surprising. Any mathematician would have been happy to point out the existence and use of a great variety of nonPeanian arithmetics; the most obvious nonextension example is clock arithmetic. Perhaps it took a tortuous dialectic to recognise such possibilities because such arithmetics as extend or factor SA typically deviate in a Peano postulate *and* a counting principle. In the case of clock arithmetic we have 1 following 12 *and* we happily correlate the same numeral to several different hours at once. One might also point out that the process of extension of the number system comes to a "natural end" at the complex numbers precisely because there are several equally compelling and incompatible ways to go. The situation is really just as in Geometry, where there are well-known alternative mathematics. If it is argued that these are not *really* alternatives because they / part of *actual* mathematics, then I think this betrays a confusion. No-one, least of all Wittgenstein, would claim that we can adopt a

⁵² I.e. that differ in respect of a counting principle from standard arithmetic.

deviant concept of addition and use it *simultaneously* with our present one. It is a peculiarity of mathematics, in fact the one which we noted was a problem for Kitcher's account, that "refuted" theories often turn into alternatives in mathematics - it happened with geometry and with set theory. There are lots of other examples in abstract algebra, and everywhere else in mathematics where systems have been investigated with an eye more to formal possibilities than direct application. Let us be quite clear: in a specific act of application I can use only one geometry, and so only one concept of distance; but "there are" several which I might use, and I may have a hard empirically informed choice to make - alternative mathematics is a reality!

§133 Craig's challenge

Dummett is the link between Wittgenstein and the Wright/Craig discussion of the same issue. In his review⁵³ of the first edition of Wittgenstein's *Remarks*, Dummett poses essentially the problem which they discuss: could a deviant community really hold on to $5+7=13$ and always put the blame on miscounting, no matter what, for any clash where three counts give 5, 7 and 12? Craig develops this question with great care into an argument against linguistic conventionalism. Wright's chapter 22 of [1980] is devoted to an attempt to protect Wittgenstein's conventionalism, as interpreted by him, from this argument. To do this he introduces a figure called the Cautious Man, who is a kind of parallel for necessity of the morally-blind person discussed in ethics. Craig dubs the position "minimal noncognitivism" in recognition of this parallel. I shall here discuss Craig's argument, which I believe has some flaws. However these are not those to which Wright points, and Wright's escape hole for conventionalism (to which Craig half agrees in his later article) is, I believe, a position bearing little relation to Wittgenstein's. I shall discuss Wright's Wittgenstein in the next chapter.

⁵³ Dummett [1959]

Craig's main argument is this: it is *unimaginable* that circumstances should afford me the best possible evidence of 7 white pawns and 5 black pawns but 13 pawns, so it is impossible, so any doctrine which implies its imaginability must be wrong. If I count and get 7, 5 and 13 as stated, in optimum circumstances, then I *must* have made a mistake: there are only a few possibilities, which can be listed. Otherwise it would be possible to repeat the case with $0+0=1$ and this can be ruled out on grounds of simple inconsistency. Thus Craig. This is an argument against Wittgenstein's idea that alternative mathematics is possible rather than against the rule-following argument directly. If it were correct, it would rule out alternative mathematics and we should badly need an explanation for the possibility of nonEuclidean geometries, clock arithmetic and set theories with and without the axiom of choice.

Craig's argument depends on the concept of having the best possible perceptual evidence. This concept is somewhat obscure. One of Husserl's fundamental insights, I believe, is that a physical object cannot be exhaustively surveyed: there is always a further perspective which can be taken on a physical object, always further perceptual information which cannot be added to our account of it. This is because it and we are located in threedimensional space, and we can always change our relative positions; the light can be changed in indefinitely many ways; there is an indefinite range of instruments which we can invent and use to mediate observations, and so on⁵⁴. Craig tries to forestall this kind of argument by the usual device: "in principle". Once again, I reject this ploy. In the present context, I believe we can demand that Craig explain why the notion of 'best' makes any sense here.

⁵⁴ A character is not so indefinitely determinable; once I have, with Eeyore, granted that that is an A, there is no more to be said about it as a character: it is defined by its role in the system, by its relations with the other possibilities which it might have been - an H or an R, etc. As a physical object, a token is of course indefinitely determinable, it is not even two dimensional as we sometimes think - though the third dimension generally plays no role in its constitution as a character. I will take this up in the next chapter.

My second objection to the Craig argument is this. It is perfectly possible to describe circumstances in which $0+0=1$ makes sense. If we are forced to give integral answers, and you and I both have 40 cents, then we each have 0 dollars; but together we have 1 dollar. Craig might reply to this in several ways. First, it does not relate to his example of the chessmen. Now Craig changed his example from counting *children* in his first article to *pawns* in his second, *without comment*. Was this because there is less scope for argument about fuzzy cases with pawns (hermaphrodites, babies, dying people etc - how do we count them?) ? If so, then we need some justification for his assumption that these are "paradigm" cases for the applicability of counting and are decisive for its nature. The historically and personally paradigm cases of counting are not so clear cut - despite his attempt to distinguish genetic considerations from questions of the nature of necessity, I don't think we should ignore them entirely; people actually learn to count with vague predicates⁵⁵. What the 40c case shows is that *of course* integral arithmetic is inadequate in many cases - that is why we have developed that of the rationals, etc.

I have a third objection to Craig's argument. He assumes that his deviant community have the same concept of counting that we do, they just have this one quirk. This is similar to a mistake that Dummett makes in his fantasy⁵⁶ of the community that counts but does not add. Actually, to count as we do *presupposes* addition, because a numeral system adequate for N must employ additive principles. If we examine the counting concepts of communities with numerals only for 1, 2 and many, we shall find a different story. '2' has to mean the same as '1+1' in even the most primitive counting system doesn't it? Could we have $1+1 = \text{many}$, despite '1', '2', 'many' being the numerals? And won't $1+2=3$ be part of the introduction of 3, the advance beyond 1,2, many ? It is not fair to Wittgenstein's position to suppose that the deviant community should simply adopt $5+7=13$ arbitrarily. Rather, they would do so as the conclusion of some consistent

⁵⁵ Notice also that pawns are mathematised objects, *made* to be indistinguishable except for their positions - ideal counters!

⁵⁶ Dummett [1959]

experience which they describe that way. This means that they would have a *technique* to pose against Craig's, perhaps using not pawns but prawns. This is to imply of course that they must have *further* conceptual differences from Craig. These practices might make sense with $5+7=13$ and not with $5+7=12$, and be so much more important that the difficulties were just shrugged off. Anomalies are. For some Christians, $1+1+1=1$ in discourse about godly persons.

I conclude that Craig's argument does not refute Wittgenstein's notion of the application of simple mathematics. Now let me relate it to less simple cases.

§134 The spectrum of application

There is a wide variety of text acts which might be called application of mathematics. It is a mistake to take Newton's *Principia* as our only paradigm, and equally it is a mistake to take ' $5+7=12$ ' as one, even more so 'one apple and one apple is two apples'. I have had in view several examples in this chapter (trying to avoid a one-sided diet) and I now wish to consider the comparison of different kinds of application. Different applications involve different amounts of mathematisation, that is to say they presuppose and entrain different parts of mathematical praxis. Application is matter of degree; there is less or more sedimentation to reactivate⁵⁷, less or more prevalence of the relevant practices of calculation, less or more use of the archive, less or more contact with the processes enforcing standardisation, and so on.

The sentence 'one apple and one apple is two apples' is mostly used as part of the mutual definition of calculus and terms through which arithmetic is learned, of which the learner has no grasp without the constant reinforcement of such counting perceptions (cf small children playing Ludo). To describe this as an application of ' $1+1=2$ ' is forced, and a symptom of a misplaced desire for too uniform an account of the use of such signs. On the other hand 'eighty-one apples plus sixty-one apples is one hundred and forty-two apples' generally would be an

⁵⁷ This Husserlian phrase will be explained in the next chapter.

application of ' $81+61=142$ '. Someone *might* apply ' $1+1=2$ ' in this way, but the circumstances would be very unusual. In the more usual case there is still a gradation from the mental arithmetic of giving change, in which the sense of application is that words are simply being used in their root senses, to more complex examples where the change-giver must get out the calculator or the back of an envelope. In these case we *see* mathematics being applied (to envelopes). We don't however see any concepts being *replaced* with exact ones, what we see is someone *extending* their (spoken) text with some notation.

The simple arithmetic of the reals may be applied by the use of relatively simple instruments such as rulers, but more complex mathematics is applied through more complex arrangements of mathematised artefacts. Consider what is involved in the application of Quantum Mechanics in making lasers. A proper description of this would be a huge task, for there is an enormous ramification of applications of mathematics in the activities contributing thereto. The extent of this material mediation between theory and application is not generally remarked, yet it grows ever greater as the mathematisation of the world and the extension of theory continue. Since the construction of the contributory artefacts in such mediations will themselves in general depend on the prior use of other instruments created with earlier theory, any application of moderately developed mathematics is a relationship of great complexity, completely distorted by concentrating on the alleged "logical forms" of a pair of sentences. 'Application', like all the others, is a family resemblance term and any simple formula expressing its essence is bound to be wrong.

§135 Making mathematics and mathematical praxis

In the discussion in this chapter, I have illustrated three ways in which my account of mathematical praxis throws light on the making of mathematics. First, the activities involved in creating and applying mathematics are predominantly text acts in which the specific nature of mathematical sign-system⁵ play essential

roles. Second, attention to the actual context of mathematics, the Archive, Readership and mathematised world, preclude any simple essentialist idea of application. Third, the normativity of mathematics arises through social praxis underlying the reproduction of the mathematical context. Let me make this more explicit.

It is not difficult to see from the foregoing discussion and examples that invention, measurement and application depend on text acts making use of all the codes in the mathematical sign-system. Words are needed to state the reasonings which lead from the acts of measurement and their reporting to their transmutation into cognitive gains. This is rather obvious, for "applied mathematics" always has a higher ratio of words to notation than its purer counterpart, if for no other reason than that the object of application has to be mentioned in the text too. In an extreme case like the Körner apple-example the contrast appears to be complete between purely notational pure mathematics and purely verbal applied, but this is somewhat of an exaggeration since the two signs in question would hardly be used without some context which would lessen this opposition. Better examples can be found in any text of econometrics or hydraulic engineering, or in exhibits ##4, 6B, and 8. They confirm the need for both words and notation.

Notation is required in order for the calculating practices to be carried out which are involved. In pure mathematics of course, derivations are at least as important as calculations proper, but in applications the proofs are often hived off to the pure twin of the applied discipline. Notation is still needed however, because measurements must be recorded in a form suitable for combination and they must later be combined. Most importantly, Notation is necessary for the increasingly explicit contrast of rules and uses.

Diagrams are needed in order to create the mathematised instruments which are needed for the amassing of measurements. Many of the diagrams in Euclid should be viewed as working drawings for the construction of rulers -

propositions which show how we can place a given segment at a given point or divide a segment into a number of equal parts may be in Euclid part of a "purely" theoretical development, but the point of the development lies outside it in embodiments of the structures with which it deals. The restriction to ruler and compass in classical geometrical construction problems is neither a self-imposed but meaningless task of the kind pure mathematicians have always liked, nor simply an arbitrary religious diktat as Seidenberg claims to deduce from the Vedas⁵⁸, but surely in origin a symptom of the perennial problem of mathematisation: how to create ever more complex measurement artefacts from those we already have.

Paragraphy is essential for the archival binding together of the various texts and textual practices involved in this complex process. Methods are necessary for referring records to one another, for their storage and location at need; for executing the accounting or astronomical algorithms on them; for promulgating the results. If the new Pharaoh should want to compare this year's taxation with that of his predecessor's first year, the documents must be producible and comparable.

Make that President, not Pharaoh (or investigator), and similar but more detailed remarks hold true.

Only the simplest mathematical application, counting, can be seen as effected without text acts, and if counting were all we did we would not have mathematics⁵⁹. It may seem that text does not enter into the debate about Craig's challenge. But my point is that the context in which such a scenario makes sense must be one which is textual; I do not deny that we also act non-textually, though we generally write down the results of observation and calculation. But these acts have their being in a social context which is a textual one. It is for this reason that Wittgenstein's restriction of his examples to the simplest ones is rather misleading,

⁵⁸ Cf Seidenberg [1962a,b]

⁵⁹ Zaslavsky's *Africa Counts* [1973] demonstrates this rather graphically: the majority of African societies she considers have only counting, and I would say, not mathematics. Of course, there is no necessary and sufficient condition for mathematics.

because it encourages us to ignore the massive social reality of mathematics. The criticism of him by Kreisel and others⁶⁰ for so restricting his examples is less well-founded insofar as it is based on his omitting the latest word in mathematical logic. (A satisfactory philosophy of mathematics must be applicable in 1800, and to the mathematics of that time; but of course these critics identify philosophy of mathematics with foundations.) The concepts of mathematised world and readership help us see that invention and application are each to be found in a spectrum no part of which should be looked at to the exclusion of others. The way to keep all the various kinds of invention and application in view together is to think of them as elements in mathematical praxis, supported by the mathematical context. Application with measurement and without it are interdependent.

The normativity of mathematics is found in social praxis, not in some transcendental logical heaven from where the laws of logic rule our actions at a distance. Wittgenstein's conception of application does not go into the ways in which social practices underlie the way N-signs are used as rules, which I have illustrated in my discussion of measurement. But the fundamental point that mathematical sentences are not descriptive, explicated by his social characterisation of concepts and their use, is true to actual mathematical praxis and confirmed by greater attention to it.

These themes will be discussed further in the next chapter, where I turn to the cognitive peculiarities of mathematics considered in the light of my conception of mathematical praxis.

⁶⁰ Kreisel [1958]. Bernays [1959] has similar remarks.

Chapter 12

Mathematical knowledge

***'Some proofs merely command assent; others woo and charm the intellect.'
- Lord Rayleigh***

§136 Problems remaining

Having considered the creation and application problems in Chapter 11, and shown how my conception of mathematical text, context and praxis can illuminate those family resemblance terms, in this concluding chapter I will consider the remaining problems identified in Chapter 1. The Chapter falls into two parts dialectically. I deal first, relatively simply, with the object, access and infinity problems. I regard the problem of objects as largely a pseudo-problem, deriving from our projecting properties of text. The problem of access in consequence is less a philosophical problem than an empirical one about the use of texts, and the problem of infinity, as I have treated it, as a case of a pseudo-problem. I do not pretend that there is not a great deal more than I shall say, that could be said about each of these problems from the point of view developed here. I shall indicate a few directions for further work at the end. What follows will suffice to indicate how my approach can throw light on these problems.

The major part of the chapter is directed at giving a positive account of the cognitive peculiarity of mathematics by making clear the role in mathematics of notation, using the accounts of text, context and praxis of Part II. Once again I shall be trying to show how the important terms - knowledge, proof, truth and so on - are family resemblance terms which have been *extended* as mathematics has developed. The claim which I shall argue for may be summed up thus: mathematical knowledge is peculiar in that it includes *evidently justified novel textual truths*. The argument consists of three strands, explaining three ways in which notation contributes to mathematics. I argue that the notion of truth which we now have has been *extended* by our history as a textual culture, and that mathematical truths are *textual truths*. This part of the argument draws on my discussion of writing and text in general. Mathematical truths are then distinguished from other textual truths by their bivalence, which I attribute to the use of N-signs in their statement. This is the first contribution of notation to mathematics. Next I argue that mathematical truths are *evidently justified* through their proofs. I argue that proofs have this

power because they are surveyable, and show that they are surveyable because they are text acts effected in the mathematical signsystem. Notation's role in this system is its second contribution to mathematics. In the course of this discussion I defend Hilbert's epistemology of mathematics and explicate Wittgenstein's claim that a proof is a design. Finally I return to the notion of definition and argue that mathematical definitions are distinguished from other text acts of definition by the involvement of notation. They are the means whereby novelty is introduced into mathematical truths. This is the third contribution of notation to mathematics.

I conclude with some general remarks about continuation of this work.

§137 Objects

Let us begin with the problem about the peculiar objects with which mathematics appears to be concerned. To what *does* the theory of real numbers ontologically commit its user, if anything? To a set of items which satisfy 13 axioms, whose content may be stated in terms of the operations of addition and multiplication and the relation 'greater than' - **if** it is recognised that these operations and relation as far as the theory goes are characterised by the 13 axioms, and whatever follows from them. Thus, real numbers are closed, commutative and associative under addition and multiplication, there is an additive identity 0 and a multiplicative identity 1, an additive inverse for each of them and a multiplicative one for each except 0, and multiplication distributes over addition; each of them is positive, negative or zero, and the sum and product of two positives is positive. *That's it!* There's nothing in all this, and nothing that can be deduced from it, as to whether real numbers are in spacetime, or are involved in causal relations, or are eternal or whatever.

By regarding mathematics as a text effect we can account for the puzzling "properties" of mathematical "objects" and our knowledge of them. The properties of numbers are of two kinds, the kind which are treated of in mathematics, and the

kind which are treated of in the philosophy of mathematics. The former are not puzzling at all - is anyone puzzled how numbers can be prime? (If one is, study number theory.) The latter are puzzling because they are not really properties of numbers at all but are obtained by a kind of projection against which Wittgenstein warned - the projection of the form of representation onto what is represented. To be more specific, my claim is this: the puzzling properties of mathematical objects are metaphorical uses of the properties of mathematical signs. And mathematical signs are written. That is why the writtenness of mathematical signs would be essential to the philosophical problems even if it were not to mathematics itself. The timelessness of mathematical objects is a metaphor for the relative timelessness of writing, their causal inertness for the relative causal inertness of signs, their iterability for that of signs. The key to accepting this point of view is this: the puzzling "properties" of numbers and such are never relevant to their use. The philosophical interpretations of mathematics make no difference to its efficacy or applicability (though they do fuel a certain amount of work in mathematics). How do we reach the state of puzzlement, indeed? How do we get the idea that numbers are timeless etc? By assuming that they are "things" like sticks and stones yet somehow mysteriously different. Then we look at the properties that physical things have and find numbers lacking.

Mathematics is not descriptive of the world, but prescriptive of our practice in it. But there is a pressure to take it as descriptive. Wittgenstein proposed the idea (essentially Kantian, via Brouwer) that we have a strong tendency to project the form of our representation onto the subjects of our discussions. The problems about objects and of our access to them are pseudo-problems in that they are given seeming urgency by this projective illusion. A similar suggestion has been made by Meyer in discussing the philosophy of logic of Curry¹.

The properties of our system of signification, that is properties of text, provide the basis for properties attributed to mathematical objects. This happens

¹ Meyer [1987]

because Notation enables us to focus on structural properties of our signs. One rather clear evidence of this is the Henkin method of providing models which uses the signs themselves.

One of the most striking features of a mathematical system, when cast in the fashionable style of the twentieth century as a formal system, is its explicit prescriptionism. We *may* write certain signs, we *may* draw certain inferences; others are proscribed. It is hard to understand how anyone can find it a mystery what is the source of logical compulsion in mathematics when its modern form makes explicit at the outset the details of the form which this compulsion is to take. Perhaps it is not recognised because there are also implicit prescriptions at work in the constitution of the formal system which are never acknowledged. I mean those which govern the recognition of tokens of its alphabet, which I have already discussed. Without returning to the debate about the role of resemblance in this process, it is undeniable that recognising an 'x' is conventional. And conventions are for the most part maintained by compulsion: children are *taught* to recognise 'x's, not involved in negotiations about what we will call one, despite that some modern educators might dispute this. What they have been made aware of is that, being conventional, our systems can be changed. This doesn't mean however that they can be changed at will.

The importance of these unstated conventions of identity is that they form the basis for the explicit rules for identity which mathematical calculi erect. And it is their very conventionality which underlies their mysterious ineluctability. It is not that $5+7$ simply is 12 as anyone can see, rather it must be twelve because x must be x. And why must x be x? Why, because that's what x is for: the whole existence of letters consists in their systematic identity maintained as an arbitrary network of differences a la Saussure². What could x be if not x? It could only be y or z or some other character, or nothing at all. There are only a small, finite, number of possibilities. What I am saying then is that "everything is what it is and not another

² Saussure [1959]

thing", but for certain purposes we say otherwise. But why do we say it so vehemently about the signs used in mathematics? First, because mathematics occurs in *written text*. Second, because mathematics relies on *notation*.

Being written, the signs of mathematics are relatively fixed compared with speech. Identity is a worthwhile issue to raise about two written signs which remain for lengthy inspection; but it is pointless for two spoken signs which are not only both gone by the time the question arises, but generally could not have been attended to together anyway. This perdurance of the written text has been the source of profound changes in cognitive arsenals (and thence of real ones), some of which are described in the works of Goody and Ong to which I have already referred. The particular point which is relevant for present purposes is that the concept of identification, which in oral culture clearly refers to the willed *taking as the same* of what is previously different, is transformed in literate culture to a concept whose centre is the *discovery* of the hidden sameness in what are apparently or in only superficial ways different. Many discussion of identity seem to deny this duplicity in the concept of identification, which leads to pointless imbroglios over the ship of Theseus and the like.

I have argued that Notation has a number of roles, the most characteristic of which is the provision of textual counters for surveyable calculi governed by explicit rules. One of its effects is objectification. Because it is the effect of Notation to create a radical distinction, a felt distance between talk and what is talked of, we have the problem of objects. Thus the idea of a mathematical realm is not so much a "projection of language" as Meyer has it, as an *effect of text*. Yet the objectification which really is mediated by text is no reason for the transfer to the 'objects' so made available³ the kinds of properties which other objects have, specifically the objects of perception. General features of notation explain features of the form of objectification it supports. Of course, the mathematical properties of

³ I think it is perfectly alright to talk about mathematical objects, as long as they are not thought to have "queer" properties. They just have different kinds of properties to those of chairs; properties studied in mathematics.

mathematical objects are determined by the structural features of the notation as given in logico-formalist style; but the 'worrying' features can likewise be seen to mimic features of notation which that ignores. Notation is written - mathematical objects are perduring, inactive, contemplatable: just like text. Notation is linear - proofs are too. Notation is surveyable - mathematical objects have no hidden properties. Notation is jugglable and compositional - mathematical objects can be 'grasped', 'concatenated', 'constructed'. And so on.

Both the other signsystems of mathematics contribute to the impression of objectivity⁴. Diagrams arise from pictures and pictures provide prima facie evidence of objects - pictures are primordially pictures of real objects. The philosophical puzzle of the kind of being which unicorns have attests to the vitality of this source of objectification. Diagrams show things. Now in mathematics as elsewhere they are used to show things which could not be physical objects, but this is not the peculiarity of their contribution in mathematical text to objectification. Rather, it is this: the way we treat diagrams in mathematics is different in that we idealise them - we ignore the thickness of the lines, etc. We take them as if they were infinitely thin, as we say. Now what should not be inferred from this, but is, is that in "pure" mathematics the diagrams concern objects with these impossible properties. Really, insofar as diagrams concern objects in the usual sense they concern just the usual objects - the ones to which there is such a problem of explaining how mathematics can be applied! There are no physical ideal triangles with completely straight edges with no fuzziness. When we talk about such things we are talking pure mathematics. The error is to mistake new kinds of talk about old objects for old kinds of talk about new kinds of object.

Similarly there is a great contribution from Paragraphy to the sense that mathematical objects are just *there*, which underlies the natural platonism of the working mathematician. The whole role of Paragraphy is to knit together the

⁴ By this turn of phrase I do not mean to imply that there is anything phony about the objectivity of mathematics. On the contrary, it is the paradigm of objectivity. That objectivity is misunderstood does not derogate from its reality at all.

objectifications in disparate text, to create the consciousness that we are all talking about the same numbers, never mind how differently expressed. The paragraphic system attempts to create one unified virtual text out of all the various actual texts of which mathematics consists; where this is threatened by barriers of language and the distances of time, paragraphic bridges are created - translations, glossaries, expositions of Euclid in modern dress, etc. *What more* can be asked of an object, say π , than that it is so multiply connected to so many other objects (circles, other numbers, functions etc etc), is recognised as the same over millenia and all cultures and by all kinds of people - blind, deaf, black, white, or nonagenarian, in a multitude of different contexts?

The mathematical context contributes otherwise too. The mathematised world provides vastly multiplied opportunities for the idealisations which support the idea of the mathematical realm, because living in such a world, one is constantly meeting objects which can be taken as, speaking platonically, inferior copies of the Forms. There just **are** lots of (approximately) parallel lines, circles, sets of ten objects etc in the world now, each of which gives one the basis for abstracting (whatever exactly that involves - we're not doing psychology here) to corresponding mathematical objects. You won't believe in dodecahedrons if you've never seen any kind of representation of one; but it's easy when you have one on your desk with a calendar on it, one month per face. There are many examples in the history of mathematics of the reality of various kinds of objects being denied, argued about and gradually accepted as they become familiar (negative numbers, complex numbers, nonEuclidean triangles...). They become familiar precisely insofar as they become embodied in the context.

The mathematical readership contributes to the objectification of the subject-matter of mathematics through its communal nature. You cannot be a mathematical worker, i.e. reader and writer, unless you are aware that you are just one of a (large) community of such workers. You know that there are hordes of other people "working on mathematics". They can't be working on nothing, can

they? The alternative to the idea that these people are all dealing with definite objects appears to be that they are just making meaningless scratches. Which is false, as one knows from personal experience! So one's subjective impression that one is dealing with independent objects is supported by the knowledge that other people are also so impressed. Together with the unity of opinion which rules in mathematics this gives a powerful basis for the idea of its objectivity. Notice that I am not denying that objectivity, I am trying like Wittgenstein to understand what it is. But contrary to Wittgenstein I have argued that the material specifics of the communication system of mathematics must be seen as the basis of the social praxis which supports that objectivity. Actually the uniformity in mathematics is not as total as is often claimed. Lots of people produce varying answers to problems and calculations. But the way it works is to rule out all but one of them. People who keep producing different answers are deemed innumerate. Notice that just these people are the ones who will say that they don't *understand* the signs of the mathematical text, who think of notation as gobbledygook.

Wittgenstein suggested that metaphysics was the 'shadow of grammar'. In this section I have argued that for mathematical objects it is better thought of as the shadow of text.

§138 The kind of thing known only through signs

Let us go on to the access problem. (Naturally, given what I have said about objects, I shall not be dealing at length with this problem.) If there were no people, the physical facts about the world would of course be mostly what they are anyway. The superreality of mathematics, its paradigmatic objectivity, seem to most writers to require that mathematical realms and the facts about them would be just as resistant to the extinction of all consciousness. But even if we grant that there are mathematical objects and that mathematical truths are objective, I do not think that that we must allow such an inference. On the contrary I suggest that a only a

clear understanding of what kind of objects mathematics treats of, and the nature of our "access" to them can dispel such platonic miasma.

Numbers are the kind of thing which are "available in presence and absence"⁵. But only within a tradition. Therefore, *through signs and only through signs*. So my view of the nature of mathematical objects has, as do others, a concomitant view of how we know them: we know them precisely through the use of mathematical signs. There is only the one reality, but our access to it is multifarious indeed. At bottom, it is through our senses; but it can be rather indirect, heavily mediated, constituted by concepts.

In his *Origin of Geometry*, Husserl discusses how it is possible for there to be traditions, such as the cumulative tradition of geometry, in which the sense of older signs is retained and built on in later ones. This is a problem which cannot be found in spoken language, for the present sense and use of a word like 'dog', though it certainly is continuous with and derives from earlier senses and uses, does not in any necessary way include any particular past uses. One of the sources of the instability which Wittgenstein's account of mathematical concept-formation seems to imply is the lack of an explanation why this process for the term 'number', has a cumulative force lacking for 'dog'. Husserl pointed to the answer, namely writing, though in the fragment in question he did little to develop it. In order to obviate the need for the clearly unmanageable 'reactivation' of all the geometrical constructions underlying advanced work before it can be continued, Husserl points out that, as he puts it, meanings are *sedimented* by their being written down. (And, I believe he would have agreed, by being embodied in the mathematised world - an idea which has some affinity with another of Husserl's later concepts, that of the lifeworld) This is the solution to the problem which Kitcher reports⁶ from Descartes, as making problematic the "storage" model of proof - how all the self-

⁵ Cf Miller JP [1982] This work explains some aspects of Husserl's lifelong project of explaining how we have access to these cognitively complex objects. One reason I say little about access is that I believe that to say much requires one to work over this Husserlian tradition - to which the work of Klein belongs.

⁶ Kitcher [1983: 43-5]

evidence necessary for the understanding of some complex deduction can be kept in mind at once. The answer is that it is not, it is put down on paper where it can be *surveyed* altogether⁷.

The other idea of Husserl's which has been used in the preceding discussion, as "sugarring" in Chapter 7, is that of "empty intuitions". This is a philosophical account of the basic idea of formalism, simply that signs may be used ignoring their interpretations. The value of Husserl's meditations on this and related phenomena is that we can offer the following reply to the common gibe that the formalist description of mathematics reduces it to a mere game. The answer is this: although in formal mode signs *are* taken without their customary interpretation, these interpretations are *not* thereby discarded. On the contrary, they are sedimented into the signs and are available for reactivation at will. (Thereby making the "mystery" of application equally vapid.) My contribution to this idea is to point out that it is the specifics of mathematical signification which make this sedimentation possible - the interplay of word, notation, diagram and paragraphy. Without this detail it is impossible to see why what Husserl says about geometry should not apply to history or even mere written gossip (granting that these are not identical).

So my answer to the access problem is simply this: we have access to mathematical objects as we do other abstract objects, through text. We have access to the specific kinds of mathematical objects we know about through the particular kinds of text which mathematics consists of. This access is of course perceptual but hardly only that. On the contrary, the difference between the baby or illiterate

⁷ The introducer of Husserl fragment, Derrida, evidently took to heart the salience of writing pointed out by Husserl, for in his later works he has presented a most thoroughgoing critique of "logocentrism" as he calls it, the view that writing is mere transcription. He finds errors associated with this idea to be at the root of the whole tradition (!) of Western metaphysics and continues the attempt at its "destruction" begun by Husserl's greatest disciple, Heidegger. Without wishing to get involved in these larger questions, I merely mention them here because oddly enough Derrida has almost nothing to say about mathematics, or even science, in all his voluminous writing. This is despite his continuing to develop ideas clearly deriving from Husserl, for whom, as Derrida himself says "The mathematical object seems to be the privileged example and most permanent thread guiding [his] reflections" (Derrida [1978: 27]) So all my preceding analysis, though provoked by Husserl, Derrida and the associated tradition, hardly draws on it.

looking at a text, and my reading it, is all the difference in the world: I thereby have access to certain mathematical objects which they do not. The idea that I also have introspectible access to them without a text is, though true, misleading. For I only have access to them without a text *after* I have acquired, at some pain, access to them through text. If all texts and writing implements were destroyed so that I never saw a text again, my access would gradually fade and disappear - and so would the mathematical realm.

It is probably obvious by now what is my view on the debate about mathematical intuition as a channel of access to mathematical objects. Yes, there is such a faculty, but there is nothing mysterious about it, it is simply one's ordinary intuition used on mathematical text - *assuming one has been educated into the readership*. When one has been so educated of course one's intuition is not so ordinary as is that of the mathematically illiterate. There is no need to posit extra faculties, only what is obviously the case, that our uncontroversial faculties can be and are *extended* by our cognitive activities.

§139 Infinity as a text effect

As Benacerraf and Putnam remark, it is not easy to say exactly what really is the problem about the infinite in mathematics. At one time the completability of infinite processes (Zeno), another the legitimacy of appeals to the infinitely small (Weierstrass), most recently perhaps the paradoxical consequences of Cantor's theory of infinite numbers. But these all seem to be worries about whether the infinite will mess things up. Formalists and Intuitionists are at one in thinking it advisable to do without it. Another strand of problems is the question, or rather group of questions, of whether infinite concepts, structures etc can have any application to the apparently finite physical world, or be known by apparently finite minds. Discussions such as those of Waismann and Rucker⁸ seem to need to mix together a whole lot of these issues. The article by Thomson in the *Encyclopedia*

⁸ Waismann [1982], Rucker [1982]

of *Philosophy* tries to systematise them along the lines I have indicated here, but is marred by the assumption that the creation of modern mathematics has settled the philosophically important issues.

One specific question which it is sensible to consider at this point in the present investigation is this: are there infinite objects, as some mathematical objects would seem to be, and how can we have access to them? I will restrict the discussion to the set of integers Z for simplicity. How is it possible for us to be able to use all the integers? There are at least three different aspects to the answer to this question. First, we have names and other signs for the set itself; second, we have the apparatus of quantification and set theory for making statements and proving them, about all integers; third, we have a system for generating the signs for all of them. The example of the real numbers shows that these devices are of independent importance, since the most primitive system, of having individual names, fails to work there. (Most real numbers have infinitely long decimal expansions and so no finite "name". There are a few exceptions like π , but only a few.) This is one reason why the geometrical model of the line is so important for the use of the reals, whereas no such model is needed for the integers, since their names will serve to bear the sedimented meanings to new uses. This answer to the question may seem to be an example of a tendency which I earlier joined Wittgenstein in scorning - that of trying to answer philosophical perplexity about mathematics with more mathematics. Actually this is not so. What I am saying is that in this particular case, the specifics of the textual channels of access to the objects in question consist of these particular devices. If we took another kind of mathematical object, say varieties of groups, then other notational specifics would necessarily need to be mentioned (not used!).

So what I have said is that Notation plays a role in giving access to these particular mathematical objects, infinite in number though they be. Of course Word is needed too, and Notation cannot work without Paragraphy and Diagram, so all the mathematical signsystems are involved. The identity of the integers across the

various number systems and in many disparate texts is maintained by Paragraphy; Diagram plays a number of roles, for example in the geometrical forms of numbers of the pythagoreans and in the sheer graphicity of the notation system for natural numbers which we use, which relies on position, linearity and a finite alphabet of signs which are recognised by their shape.

The mathematical context plays its role here too: though many possible notations for natural numbers are available, one must know some specific one to read about them; the archive is needed to provide the polyvalent interpretation of the signs in question in all their innumerable contacts with the mathematised world, which is a ready source of examples of what may be counted indefinitely, for example clocks numbering time (there is always a next moment), and odometers space, the money passing through the bank, the customers through the checkout, the very words one writes. In all these and many other cases the iterability of making a next one is present. The increasing prevalence of microprocessors which run on clocks continues to spread this denumeration of the world.

§140 Mathematical knowledge

I now turn to the problems of proof and truth. Following the plan outlined in §136 I shall now account for the cognitive peculiarity of mathematical knowledge in terms of the text acts of proof and definition. I distinguish knowing mathematics and mathematical knowledge. For example, if I am told by a reliable (authoritative!) source that the Riemann hypothesis has been proved, and I believe it, then I gain possession of a new mathematical fact, I know that all the zeros of the zeta-function are on the line $\text{Re}(z) = -0.5$. But my knowing it is not mathematical knowledge in the strictest sense, for I could not explain to you why it is true. Quite a lot of the "mathematical knowledge" which anyone has is grounded in the real mathematical knowledge of others in this way. There is nothing wrong with this situation, but it must be recognised. The extreme case is easy to recognise, someone whose mathematical "knowledge" is entirely secondhand simply doesn't know any

mathematics really; and many people's membership of the mathematical community ends at the point where their knowledge of algebra or calculus or whatever takes on this form - they can only parrot it⁹. This extreme case is very common, in the sense that most people have these hazy areas at the limit of their command of mathematics. Everyone relies on the knowledge of others to some extent. Where one has true beliefs in mathematics, the justification which makes them knowledge for oneself is not the existence of a proof but one's own grasp of a proof. Even if Fermat had a proof, even if one is written down on some long-lost document of his, our mathematical knowledge does not at present include his "last theorem" because no one knows a proof of it.

Mathematical knowledge is knowledge gained through certain kinds of sign, it must be got through the process of proof-following (or making - the difference is one of degree not kind); it is not enough that proofs are possible or even exist. The process in question is one of textual activity. Just as the development of mathematical knowledge is a textual process in which the archive is used in its own extension, as I argued in the last chapter, so is that of its reproduction, which I examine in this. So I have adopted the traditional view that proof is the distinguishing feature of mathematical knowledge and I shall be arguing that proof must be understood in the textual framework I have developed in Part II, and that the epistemological problems of philosophy of mathematics - truth, certainty, necessity and so on - are equally amenable to illumination from this direction.

§141 Mathematical knowledge as text effects of proof

I do not believe that mathematics is more *certain* than other knowledge, since I take certainty to refer to our subjective feelings about our knowledge. Much religious or racial bigotry achieves a lamentable *ne plus ultra* of subjective assent. What is distinctive about mathematical knowledge is not the degree but the kind of certainty. Its certainty is peculiarly well *justified*. Indeed mathematics provides a

⁹ The test is whether they can use it! For proofs can be parroted too.

dominant paradigm of justification and this is its distinction. Mathematical truths, theorems, are justified by proofs. I shall use the term 'evidently justified'. But what are truth and proof? I say that they are, in mathematics, text effects. When we grasp a theorem it is because we understand a proof; understanding a proof is gained through work on text; the kind of understanding is determined by the kind of text; and mathematical text has specific features which I have analysed in Part II: it is written in a fourfold system of codes, <W,D,N,P>, in a context of Archive, Readership and Mathematised world, <A,R,M>.

Proof and truth are text effects because these are family resemblance terms in which one branch of the family has become dominant. The concepts of truth and proof in mathematics are textual extensions which have largely supplanted the older conceptions from which they developed. In the case of proof this dominance of the textual sense is virtually complete - the extinction of other kinds of proof than mathematical is almost total. We can see them lingering on a little in "metaphorical" uses of the term, as in 'proof whisky' and 'artist's proof'. But one doesn't take these seriously does one? Perhaps we should. In the case of truth the issue is rather more clouded. I believe that the mathematical variety of truth has gained a hegemony on the thinking of philosophers, who have spread a confused notion of it far and wide. The theory of truth called referential semantics has found favour because it is a precise theory. I have indicated that it captures part of the way mathematical signs signify. But it does so at the cost of ignoring the complexity of the semiotic situation in which mathematics is used. Mathematical truth is a new kind of truth, a species of a new genus introduced by writing. Mathematical truth is akin to truth in fiction and art (which is perhaps why Plato was so hostile to the poet and artist), as varieties of *textual truth*¹⁰. It is a mistake to seize on the notion of correspondence, which does indeed illuminate some aspects of mathematical truth, and try to generalise it. The same applies to any notion of truth: truth is very

¹⁰ Ironically, most philosophers treat as a problem the explanation of metaphor while relying on a metaphorical use of truth as their touchstone of literality (itself already a metaphor!).

much a family resemblance concept¹¹. Truth in mathematics is connected to the idea of *correctness*, and this is system-relative. (I pointed out in Chapter 11 how the notion of correction connects the learning and application of mathematics, scientific theorising and the establishment of social practices.) This point has been erected into a principle by the formal school; so it is doubly ironic to find the formal notion of truth of Tarski, which is clearly an entirely relative one¹², lauded as some kind of vindication of the absolute notion of correspondence! The mark of mathematical truth is the rule of bivalence. The attempt to cut off the purely textual truth of propositions from the other dimensions of truth is quite unsound.¹³

There is an idea that there is some kind of priority dispute in the philosophy of mathematics - does truth come before proof or vice versa? I suggest that this is clarified if we concentrate on text effects. Proofs come before truths - that is their role; but also, truths come before proofs - proofs call on truths achieved already. It is a false dichotomy. Neither notion could be explained prior to the other since neither can occur without the other - in mathematics. The peculiarity of proof is its manner of founding statements in the text - but the priority of proof and truth is entirely relative to the textual boundaries we look at. Textual truths are capable of entering into this kind of accumulation just because they remain there in the text to be called on later. They are distinguished not by the nature of what they say but by the kind of thing we can do with them.

Axioms, definitions and theorems are best understood as *textual resources* made by proofs. I argued in Chapter 5 above that defining has become through writing a text act. A definition is a characteristically textual item. Its availability for use is a textual availability - you check its exact statement; you confirm the

¹¹ The "stand-off" among the competing accounts of truth - correspondence, coherence, redundancy etc - is easily accounted for on family resemblance semiotics: these labels indicate different related uses of 'truth'. The search for a common essence is delusory. It is one of Wittgenstein's insights that quite a number of philosophical debates have this character.

¹² Cf Jennings RC [1986]. There is a literature about this but I don't intend to go into it here.

sense of the terms involved and you justify its employment under challenge, all by reference to the textual corpus in which it is located. I will explain shortly how mathematical definitions are special.

What makes something an axiom? The old view was that certain propositions were intrinsically axioms; that they were known by their indubitability and underivability. Modern axiomatic investigations have refuted these views and left many writers at a loss for any account of what makes something an axiom. My view is rather simple: certain statements are axioms because they are *labelled* as axioms. The discussions of labelling throughout the earlier chapters may ward off a tendency to think this a trivial account. Thus a sign is an axiom sign as a result of a text act, and we can understand easily why the "same" proposition can be an axiom in one system and not in another. (in the platonic view, of course, there is such a problem in accounting for the special status of axioms among the theses of a system that some writers are driven to give them up by adopting if-thenism, for example Russell.) A theorem is equally a textual resource - all theorems are produced with a view to application and use; every problem's solution suggests a myriad of further problems.

In the next few sections I will show how the framework of Part II can explicate the ideas that a proof is a design, and must be surveyable. This conception of proof will be then used to complete the refutation of Craig's argument against Wittgenstein's conception of necessity and show how necessity is produced, and to vindicate Hilbert's notion of proof against recent criticisms. But first I shall fill out the idea of textual truth, in particular mathematical truth.

§142 A proof achieves bivalent textual truths by deploying notation.

Examples like the Hansen paper which I analysed in Chapter 7 make it clear what is the real basis of the object-language/meta-language distinction. It is the role of the Notation system in a mathematical text whose overall articulation is governed by the written Word. N-signs are *contrasted* with words; their role is to be

displayed, as visual counters. Recalling also my point that notation retains the iconicity of the diagram system, I thus interpret Wittgenstein to be getting at something like my point of view when he insists that a proof is a geometrical design¹⁴. I will expand on this in the next few sections.

Being written, the assertions in a proof can aspire to textual truth; concerning notationalised objects, they are subject to the rule of bivalence¹⁵. Using these two conditions we can distinguish the truths of mathematics from other written truths such as those of history or law, and from other written nontruths such as those of knitting and music. The history of truth has only one event about which we know much: the introduction of writing which shifted the location of focus from the statement (the saying) to the sentence (the said). The debate in the philosophical literature as to which of these is the truth-bearer is, like many such debates, a confusion created by ignoring the temporal development of the concept in question. Prior to writing there were true friends and arrows, true planes and true words spoken in jest; truth was not limited to linguistic items and those to which it did apply were *sayings*. Truth and troth are cognate, and truth is etymologically related mostly to ideas of fidelity and loyalty as *human* rather than representational qualities. 'To plight one's troth' sounds like an anachronism now, but was importantly formative of the concept of truth. The effect of writing on this nexus of ideas was utterly central to the creation of our western world-view. The shift relies on a simple thing, the characteristic of the written that it remains beyond its utterance. The truth can become a property of the product rather than the process. We are so immersed in textualised culture that we take the suggestion that statements rather than sentences might be truthbearers as a trifle primitive - which in way it is. The novelty of written truth is its combination of two elements: the picture and the affirmation. I have described above in Chapter 5 how writing developed from several sign-systems including the picture; we know a little about

¹⁴ Wittgenstein [1956:I, §§24ff; II, 38-40;],

¹⁵ Mathematical texts *about* nonclassical logics are still conducted *in* classical logic, and therefore subject to the rule too.

the nature of affirmations for we are surrounded by them in everyday speech, but they are hard to see for the brighter light of the written forms in the texts with which we are now also surrounded. Some anthropological literature gives a little insight into the role of affirmations in a non-literate society, but the logic of the "savage mind" has not been much studied from the present point of view. Goody, in the works cited already, is one of the few writers alive to the issues.

There are theorems about chess; a text about how to play chess becomes mathematics when it demonstrates the impossibility of mating with two knights alone. The change turns on the purposes of the discourse: proffering this fact as an aid to play is not mathematics, but establishing its truth is. Most chess text is not mathematics despite being full of assertions using notation, because the generalities offered are not deduced. In music and knitting the difference is that the notation is used not to make assertions but to give directions for making things. There well could be theorems about knitting or music, though probably the best example of the way a notation can be theorised is given by the rise of computing "science". There is a vigorous research program devoted to the development of a theory about programs - program verification. The main aim is to be able to prove that programs are correct, i.e. do what is wanted. Many practitioners are derisory about this whole idea; but for illustrative purposes we need not enter into the debate, merely observe that the notations used for the creation of computer programs *can* be put under the rule of bivalence and treated theoretically, so that discourse which is certainly in its humbler examples continuous with knitting patterns approaches in these rarified regions the condition of mathematics¹⁶.

In our present culture the valorisation of written truths over spoken ones is pretty complete. Respect for what you read as true (or likely to be) is part of the dominant ideology. In the leading edge of culture this process has gone some way further: for most writers on the philosophy of mathematics it is a simple conceptual truth now that the primary meaning of truth is to be found not just in text, in

¹⁶ Mathematics does not have to be *valuable* to be mathematics.

standing sentences, but in the explicit model for such truths which has been worked out by mathematicians for the "sentences", the wffs of mathematics - the idea of truth-in-a-model, Tarski's "theory" of truth. The paucity of opposition to it is symptomatic of its assimilation into the thinking of philosophers¹⁷. When taken to task on this and similar points, formalistic philosophers often retreat to the claim that they are only trying to account for scientific discourse or mathematical discourse. But the logico-formalist hegemony has penetrated very far from these original loci, and in any case few have looked closely at real mathematical and scientific discourse in the manner which I have attempted here. An example of the excesses of the conceptual leading edge is Kripke's well-received paper on truth¹⁸. This actually achieves little advance in our understanding of truth, for it mainly suggests ways that "solutions" to the paradoxes may be had, as though such philosophers' conundrums are the real locus of the problem of truth, and it does it by talking of extending into the transfinite "obvious" formal tricks on hierarchies of artificial languages. The interesting pragmatic idea of the relative groundedness of sentences is not investigated for real cases. The invented examples about Nixon and company are soon left behind in favour of transfinite set theory - but there are and can be no examples of people using "languages" of the kind discussed. Such productions get great respect because of their technical cleverness - but they don't lead to any real understanding.

Legal truth is interesting to contrast with mathematical. Here there is also a pressure to bivalence, revealed by the phrase "the letter of the law", and the British adversary system and the fact that there are usually two possible verdicts. But legal statements are inherently subject to interpretation before judgment; those of mathematics afterward. The vagueness of legal statements is removed from mathematical ones by the use of notation. The greater the formality of discourse, the more it can be treated bivalently. Complete formalisation excludes the

¹⁷ But see Hodges [1986] for some argument against this prevalent identification.

¹⁸ Kripke [1975]

reinterpretation of terms which enables conjectures to elude the simple alternatives of truth and false, as Lakatos' discussion of Euler's conjecture shows in great detail.

Now I turn to the two main text acts which create textual truths, proof and definition. I take up the idea of a proof's needing to be surveyable, and the suggestion at the end of Chapter 7 that this is a power deriving from the threefold nature of notation as abbreviatory, formal and iconic. This part of the chapter is the second leg of the argument sketched in §136.

§143 Proofs have their power through being surveyable.

Proofs are sufficient justifications for new mathematical statements. Therefore they must be written. Why so? Because to justify, they must be surveyable; and to be surveyable they must be written. I will discuss the need for surveyability in this section, and its dependence on writing in the next.

Wittgenstein and Hilbert both emphasised the need for proofs to be surveyable, and I shall return to their views below. Tymoczko¹⁹ has recently argued that surveyability is a desideratum of proofs. He does this in the course of presenting the recent Appel/Haken proof of the four colour theorem²⁰ as a problematic case for the standard notions of proof, knowledge and certainty in mathematics. The problem is the presently ineliminable appeal to the results of very extensive computer calculations for a crucial lemma in their proof. This makes the "proof" - that is, either the proof as given by Appel and Haken, which appeals to the computer's results, or a completely written out proof including all the computer's working - not surveyable. Since in Tymoczko's view this makes mathematics empirical and uncertain, the acceptance of such proofs, he says, marks a change in the concepts of proof and mathematical knowledge. Tymoczko's argument has been strongly criticized²¹ but his concept of surveyability has not

¹⁹ Tymoczko [1979]

²⁰ An old conjecture, that any plane map can be colored with only four colors so that no countries with a common border are colored the same.

²¹ Detlefsen and Luker [1980], Teller [1980].

been directly analysed. Rather, his claims about the connections between surveyability, empirical content and proof have been contested. It is not really clear whether Tymoczko is saying that surveyability is necessary, so that the Appel/Haken proof is defective, or that it is no longer necessary, and the notion of proof has changed. He is suffering from essentialism about the meaning of 'proof'! Here I want to clarify the concept of surveyability, and only incidentally adjudicate the debate about the four colour theorem. I agree that surveyability is a demand on proofs, but this must be understood aright.

Tymoczko thinks a proof must be surveyable so that it will be free of empirical content which would make it uncertain. Here he is mistaken, as Detlefsen and Luker argue. They point out that nonsurveyability is not the only source of possible error in mathematics, all calculations are - and this possibility of error is good, it is the source of mathematics' empirical content. They even suggest that surveyability is itself a source of empiricity in mathematics²². Moreover computer calculations as in the Appel/Haken proof are not the only source of nonsurveyability - for example some proofs in group theory are just extremely long, like the classification theorem for the finite simple groups. But they deny that empirical content makes mathematics uncertain. Teller also argues that the possibility of error does not make mathematics empirical in content. (His further claim that since mathematics is not an essentially human activity, the significance of surveyability is only that proofs *we* can check must be surveyable, will not be discussed here; obviously, it is quite alien to my point of view.)

What does it mean to say that a proof must be surveyable, as Tymoczko does? If it is surveyable, it can be "looked over" "reviewed" and "verified" by a rational agent, "definitely checked by members of the mathematical community", "comprehended by the pure power of the intellect". Surveyability gives "a kind of certainty" and permits the "removal of appeal to authority". The mathematician "surveys the proof in its entirety and thereby comes to know the conclusion". After

²² I shall argue below that this is a misleading idea.

reading the famous story about Gauss and $1+2+\dots+100$, we "have surveyed the proof and are convinced". Finally, surveyability "provides for the democratization of mathematics by making proofs available to any competent mathematician".

There are several different aspects of proof mixed up in all this which need separate as well as combined investigation. The first concerns the written character of the *physical* proof; the second concerns a particular kind of *psychological* state needed for mathematical knowledge; the third concerns the *social* arrangements for recording and disseminating proofs. First, the idea that a proof may be susceptible to a certain kind of inspection - scanning, back and forth, integrative, minutely detailed. So, a proof is surveyable if it can be acted on by a mathematician, perceptually, as a physical object, in certain characteristic ways, to achieve certain specific cognitive ends. Here the written nature of proof is clearly crucial. Second, the need for the mathematician to be convinced through understanding or comprehension, and the fact that a proof may make this possible. The ends referred to by Tymoczko are characteristic features of mathematical knowledge: certainty, self-authentication and unanimity. These are gained by 'comprehension', 'definite checking' and 'verification', which are the psychological acts of surveying made possible by a proof's being surveyable. Thirdly, the openness of proofs in mathematics to all comers. The consequence of those cognitive achievements, made possible by a proof's surveyability, is the depersonalization of cognitive authority to which Tymoczko refers. This availability depends on the proof's reproducibility in surveyable form. So these are three reasons why proofs need to be surveyable. But what is it to be surveyable, and how is it to be assured? Tymoczko virtually identifies it with brevity, but this ignores other properties of notation and the mathematical sign-system. None of the disputants seem to even be aware that it is properties of proof *tokens* which are crucial. In the next section I point out how writing is the key factor.

§144 Surveyability comes from written expression

So, a proof must be surveyable so that it can be used for characteristic cognitive ends of mathematicians (human or not). It must be useful for convincing you of a result or explaining it to you (these are Steiner's uses for proofs) or for generating new knowledge or increasing understanding (Kitcher) if these are different. To do so it must have certain characteristics. Some discussions of surveying proofs suggest that all that is required is that the mathematician should be able to check each step. But that assumes that a proof consists of steps, whereas I agree with what I take to be the leading idea of Tymoczko's account, with which most mathematicians seem to concur, that the proof must enable a 'grasp' of the whole proof, must permit you to 'see' the essential organizing idea of it, see the objective wood through the trees of symbols²³. ND Goodman²⁴ writes of how doing mathematics is a visual process for him and only half apologises for using visual metaphors to explain his ideas about proofs. I am suggesting here that although 'grasp' is a simple metaphor, 'see' is not a metaphor at all.

What is necessary if a proof is to be surveyable?

It must be enduring. The lifetime of a proof must be long enough to permit survey, and to last between surveys so that various surveys can be compared.

It must have visual form. Only then can sight be used to comprehend it with scanning, back-and-forth checking etc. The other senses do not permit the processing of the amount of information involved in the *simultaneous presentation* of all the parts of a proof. All the parts must remain simultaneously present for repeated inspection.

It must be two-dimensional. It is necessary to combine linearity with compactness in order to represent a great deal of information in a small space which is simultaneously present to the attention. Three-dimensionality would provide an intensification of the economy but prevent the simultaneity.

²³ So sometimes a diagram will do for a proof, no steps at all.

²⁴ ND Goodman [1984]

It must be articulated. There must be parts of a proof which are arranged in a perspicuous relation to one another.

In short, *a proof must be written.*

No decent proof would be surveyable if spoken, the details of the beginning would be forgotten long before the end was reached. This is true enough for most philosophy papers, let alone mathematical arguments where precise details can make the difference between soundness and unsoundness. Actually, surveyability is a relative notion, and in verbal proof sketches it may be had by radically chunking the argument (give a name to that matrix!); this occurs in a sense with published proofs too in that they all have more rigorous longer versions by which they might be replaced. The point, though, is that there is a convention in the mathematical community at any given time about the acceptable level of chunking, which verbal proofs cannot match while they remain comprehensible.

So proofs must be written, but it is not true that any written down argument is a proof. Mathematical arguments have features other than surveyability which are explained by the specific features of the mathematical signsystem. To be justifications means to be deductions, and to justify new facts means to extend the meanings of mathematical terms. Deductions require W-signs, as I have argued in Chapter 7, contrary to the conventional wisdom of the logico-formalist hegemony. Extending the meanings of mathematical terms requires archivability, presumes an archive, and demands some level of formality. Formality requires N-signs. Archivability demands P-signs. Sufficiency of justification requires surveyability, which involves D-signs and/or their stylisation into N-signs. I shall argue for each of these claims shortly. I have argued throughout Part II that the W D N and P systems are relatively irreducible. All this together then will show that proofs must be made in the specific written mathematical sign system.

That proofs often require deductions, I take it requires no argument - for the mathematical concept of proof. Formality has been gradually forced on proofs by the need for them to be surveyable - Lakatos has shown how this works. In these

days the influence of logico-formalist hegemony is such that the *mark* of soundness in deduction is formality. This is a mistake, but it is nevertheless determinative of what mathematics is in our time. The logico-formalist hegemony has affected how some kinds of mathematics are presented - Hansen's paper is much more like the logico-formalism description of mathematics than any 19th-century text. The mistake of the logico-formalist hegemony is to ignore all the rest of mathematics which is not presented in this spare style and which grounds the "purer" texts. Proofs have always involved formal elements, from Euclid on. The use of axioms entrains formality. 'Calculations' must be present in proof, in a broad sense of calculation; and calculations require N-signs.

§145 A proof has a place in an archive, presupposing P-signs

Continuing to explore how surveyability is made possible, let us next consider the role of Paragraphy. The extension of meaning, rather than its simple change, requires the *sedimentation*²⁵ of the old meanings. This has implications for the laying down of proofs and for their uptake too. The semiotic point of proof is the dissemination of constructions, the production and reproduction of display texts "Look what I have made: you make it too!" (Recall my discussion of display as an important text act in Chapter 6.) Consequently the effects of proof are text effects: psychologically, to persuade us or to deepen our understanding of textual concepts and claims; and semiotically, to add to the archive.

Proofs are published in a form very different from "informal" versions first sketched to knowledgeable colleagues²⁶. Proof-sketches are in fact generally at least partly written, frequently appealing to diagrams and intuition or insight in ways that get expunged from the published versions. Why is it accepted as necessary

²⁵ I am adopting Husserl's term here, and I believe, his idea. But I am more concerned to be accurate to mathematics than to Husserl. I will discuss Husserl's idea briefly again below.

²⁶ There are some famous examples of this, Archimedes and Newton for two.

to make these modified proofs? Can the modifications readily be undone in verbalisations which don't also strip away whatever it is that they done for? The leading differences introduced into the published form of a proof are the addition of explicit links to the literature, suppression of motivation and of insight based on diagrams and other kinds of special examples, and stylised logical organisation. To the extent to which the links to the literature embed the language and assumptions of the proof in pre-existent formal mathematical discourse this process may be called formalisation. But in many cases the links are to relatively informal prior discourse, and in some others the new work itself contributes to the creation of norms of formality, for example Frege's *Begriffsschrift*. The apparatus of bibliography and citation is necessary for the reading mathematician to be able to juxtapose the new proof with the relevant old work and make suitable comparisons and contrasts. It is not in fact possible to listen effectively to two spoken proofs, even if it were to one (which I deny), so that comparisons in a spoken mathematics would be forced to be sequential²⁸. This would eliminate the power of survey and conspectus which sight has over written proofs and if it were to be reliable, make following new proofs and using old proofs uselessly slow, if possible at all in a short life. Even were this not so, (I shall discuss it further below), there needs to be some equivalent of the reference for the mathematician following a spoken proof to permit fetching a related one. A system of spoken names is hardly plausible, for the 200 000 new ones needed each year dwarf the invention of new names for chemicals, yet even that requires an enormous written basis for its control and use. How else can a huge hierarchically organised system of names be attached to its referents except by some system of labelling? - which means writing.

Decontextualisation is required in published proofs to make them safer and more universally accessible, no longer relying on unspoken assumptions shared by

²⁸ You cannot visually attend to two different things at once either, but written signs *remain* for your attention to return - speech does not. This simple contrast to which I keep returning is the crucial physical basis for such a large transformation of human culture!

giver and receiver of verbal proofs, or special non-generalisable aspects of examples. Naturally this process decreases their immediate accessibility to the outside reader so that hooks to previous knowledge are needed; hence the citational apparatus. But this only makes more nearly explicit what is implicit anyway in the spoken proof - that a proof does not stand alone but must be embedded in a shared background of language, logic and axioms (to put it in an excessively logistic way). Additions to the corpus generally acknowledge this by the explicit referential apparatus I have been discussing, but we should not be misled into thinking that it is not there implicitly in other parts of the mathematical discourse. Many tracts, particularly introductory ones on well-established fields make a gesture towards their presupposed knowledge at their outset, and many of them disclaim any specific such knowledge, merely 'mathematical maturity, and the interest and ability to follow a chain of reasoning'. This unenlightened shirking is rather common and many unwary readers discover that such tomes make less than perfect sense in the absence of specific prior experience too great and detailed to lay out in an elegant chapter zero. The conceit, that a clever enough mind that has had enough mathematics pass through it, no matter what, can assimilate any advanced theory simply by adopting the definitions as hypotheses and following the deductions, is a ludicrous one and far from the truth about minds as we know them.

Actually the sparse 'Landau style' has been seen as pernicious recently and attempts made to re-inject some context into contributions to discourse (For example, the Mathematical Intelligencer has editorialised against "mathematical telegrams"). But these are swimming against a semiotic Amazon of mathematical production, vast quantities of additions to the oceans of already discovered truths. The dynamic of mathematical organisation is created by the pressure to structure the ever vaster amounts of present information into accessible forms. The spare but complex logical organisation of mathematical text should be seen in the same way, as a device to pack a lot into a manageable space. Reading mathematical text is a very slow process compared with reading reported speech! (This in itself indicates

that written proofs are very far from being simply "equivalent" to reported speech.) While this kind of purity and minimal redundancy can be processed by the eye, speech requires redundancy and simple sequence of thought for comprehension.

Let us indulge in a little fantasy, and imagine the world arranged to provide the maximum support for the verbal mathematicians. We first suppose that they have an indefinitely improved information processing technology to provide rapid and easy search and retrieval of aural records of mathematical work from some kind of super record library. I say "Pythagoras theorem" to a little box on my desk and it quickly replies "Let..."; at any time I can make it backtrack and replay any part of the record, or open another file and play me some nonEuclidean theorem (though not at the same time) and so on. The constraints are that my inputs must be aural, as though they were the speech of another human mathematician, and provided one at a time. (But rapid intercutting is permitted.) Actually I think that any such technology is going to be dependent on writing for its design, construction and operation just as are our present computers, but let us pretend otherwise for the present. Could I learn mathematics, use it and make more of it with this tool in place of writing? In particular, could I access the accumulated riches of mathematics as I can through the written archive?

Suppose I wish to understand a certain proof of Moran's which uses the concept of a 'Bernoulli convolution', with which I am presently unfamiliar (this is a true story). How could I search the spoken mathematical archive to find out the things I need to know about this concept? In reality I might browse the shelves of the Barr-Smith library, knowing where books on that kind of topic will be; or I might have to go to the periodical literature, which I'd do through Mathematical Reviews. How could I get my black box to skim titles of books, contents or index pages of possibly relevant ones, or whole chapters of likely ones as I can with sight? Speeded up talk rapidly becomes incomprehensible in a way that rapidly scanned writing does not. So that is one kind of search not open to me. Well, a highly structured index system like Mathematical Reviews perhaps? This will take me to

the items that may be relevant, but when I get to the abstracts I have to either hear them in their entirety or we are back to skimming. If the index is perfect, my problem is solved, but if it is no better than existing ones, the same problem arises from both these search strategies, the impossibility of aural skimming. The question becomes then whether such a database could be created with such a 'superindex', without writing or anything equivalent to it. The nearest I can propose as an example of what might be possible would be derived from the 'theatres of memory' much developed before print made them unnecessary²⁹. How would a new item be entered into the database? By taking the index as given and classifying the new theorem according to the place of its leading concepts in the organisation of mathematics presupposed by the index. But then how will the index develop? The organisation of mathematics surely changes, indeed any genuinely new contribution must change it slightly at least. So the *index* must *itself* be evolutive. But to grapple with the very complex task of revising a classification requires one item displayed, the grid, and the other, the material, in the mind to be matched this way and that against it; or better, both on display for tentative juxtapositions. What would be precisely lacking to the verbal mathematician would be a clear space in the mind for this adjustment to go on because attention to speech precisely doesn't permit simultaneous explicit attention to two different parts of the archive, even the internalised part of it!

The arguments all come back to what is actually possible to us with the senses we have in real mental operations we can perform. This latest is the difficulty, writ large, of a mathematician trying to understand a new verbal proof by comparing it with older ones.

§146 A proof is a construction using the iconicity of the diagram system

Many proofs are entirely without diagrams and it may seem foolish to suggest that proofs rely on the diagram code of the mathematical sign-system too.

²⁹ Cf Yates [1966]

Nevertheless I make this claim. I have already given two arguments which support this thesis. First, I have argued in Chapter 8 that a given mathematical text is part of the archive, held there by paragraphic bonds in addition to the continuity of the spoken language which the mathematical sign-system extends. Any particular proof can still be relying on the diagram system, through its reliance on being part of the archive which *does* need the diagram system. Second, I have argued in Chapter 7 that the Notation system has internalised some features and functions of the diagram system as it has developed, particularly in this century. The diagram system lives on in texts like Hansen's in the iconic properties of the notation which it uses, notation which is held in common in most recent text by our paragraphic system. It is of course this iconicity which is articulated by the representational theory of truth for mathematics which the logico-formalist hegemony has enshrined in model-theoretic semantics. And it lives on in texts which do contain and use diagrams which form the context in which readers understand a text like Hansen's.

A proof is a construction in the sense that it consists in the building up of complex signs in novel ways rather than the introduction of new signs. Even nonconstructive proofs are constructions! First, in the course of such proofs items are constructed, even if not the ones proved to exist. Second, the articulation of the proof is a construction whose details determine its value or otherwise. For example, in Euclid's nonconstructive proof of the infinity of primes, it is shown that given some primes there is always another, by constructing a range of numbers among which a prime must occur. And the demonstration is a carefully sequenced argument which is built up from terms used in earlier parts of the text. In the Heine-Borel theorem, a number with the property in question is not constructed, but a sequence of sets is, whose intersection consists of such a number. Not all the sets are constructed in a simple sense, yet the sequence of them can be, and is, said to be constructed. Surprisingly many proofs of novel facts turn on a new idea for constructing a number, or a function, or whatever, with interesting properties.

Generally the point is not so much this item itself but what it shows about the classes to which it belongs. By constructing a transcendental number, Liouville showed that they exist - the "nonconstructive" Cantor proof equally relies on a construction, namely the "hypothetical" construction of another real number from any proposed enumeration of them.

In the next section I return to discussing why surveyability is important for proofs. Hilbert's philosophy of mathematics, formalism, is the nearest approach to an understanding of this among the classical positions. I will defend it against some objections which will enable me to return to the Craig challenge to Wittgenstein's account of necessity, and vindicate Wittgenstein and Hilbert together.

§147 Hilbert's Epistemology of Mathematics

Korner points out quite rightly that Hilbert starts from Kant philosophically. Firstly, from the surveyability of concrete items and the fact that they cannot be incompatible; they are the wrong kind of thing for incompatibility, only propositions can be incompatible. Secondly, from the conception in Kant of Ideas of Reason, which in Hilbert become "ideal statements". Hilbert's position was also influenced by the neoKantian Husserl with whom he corresponded. We must keep in view however that Hilbert also began from a specific mathematical situation. Certainly, he began from the problem of foundations for arithmetic made acute by Russell's paradox, but also from his own history of mathematical successes with particular characteristics. This is easily observed in the selection of problems and examples in his seminal 1900 speech on Mathematical problems³⁰. It is also at work, though less obviously, in his later formalist days. Three early successes worth bearing in mind are his existential solution of an algebraic problem that "should" have had a combinatorial solution, and his simplified proofs of the transcendence of π and e (impossibility proofs); his magisterial magnum opus on algebraic number fields, relying on Kummer's theory of ideal elements; and his

³⁰ Hilbert [1902]

Foundations of Geometry in which he brought axiomatic method to new levels of delicacy. Hilbert's was never a pessimistic philosophy like Brouwer's defeatist Intuitionism, which he detested. On the contrary it was characterized by optimism and a conviction of the value of classical methods. Its main element seem in its mature form to be as follows.

What is surveyable may be characterized without fear of contradiction. (Hilbert tends to talk of 'certainty' where I prefer 'evidently justified'.) Statements of basic arithmetic and logic - "real statements" - are certain because they are finitary, describing surveyable propositions. But the use of quantifiers leads out of the class of finitary statements into that of "ideal" statements. Ideal statements are not directly interpretable. Ideal statements are nevertheless used with justification as part of a consistent, conservative extension of the class of (true) finitary statements. The criterion for a consistent extension is not the exhibition of a model, but that $0 \neq 0$ should be demonstrably not derivable³¹. An extension is conservative if for every proof involving quantifiers there is one without them. To prove consistency and conservativity for arithmetic requires a formulation drained of all meaning - just as this was necessary for "point, line and plane" in the *Foundations of Geometry*. A "proof theory" is needed. The required theorems will therefore be theorems about a formal system - "meaningless marks on paper".

Hilbert puts it this way³²:

*"elementary mathematics contains, first, formulas to which correspond communications of finitary propositions (mainly numerical equations or more complex communications composed of these) and which we may call **real propositions** of the theory, and, second, formulas that - just like the numerals of contentual number theory - mean nothing but are merely things that are governed by our rules and must be regarded as the **ideal objects** of the theory."*

Examples Hilbert gives³³ are

³¹ Hilbert's criterion for consistency varied somewhat - sometimes the nonderivability of $0=1$ is sought. Later logicians have clarified this somewhat, but within the logical framework Hilbert used the differences are unimportant. Cf Routley [1980: 931] for an introduction to this literature.

³² Hilbert [1927: 470], emphasises this.

³³ The first two are from Hilbert [1925].

[E1] $1 + 3 = 3 + 1$

[E2] $3 > 2$

and

[E3] $1 \neq 1$

What characterizes true statements of this kind is that they are³⁴

"immediately intuitable and understandable without recourse to anything else". [They]³⁵ "can be verified by contentual considerations."

To understand such a statement correctly involves grasping its truth. This directness arises from the concreteness of the

"extralogical concrete objects which are intuited directly as directly experienced prior to all thinking.....For logical deduction to be certain, we must be able to see every aspect of these objects, and their properties, differences, sequences and contiguities must be given, together with the objects themselves, as something which cannot be reduced to something else and requires no reduction."

Here³⁶ Hilbert invokes what Curry calls proof "ad oculos" - that is, look and see! We can verify the truth of E1 by direct inspection of the content of this statement. It merely asserts that '3+1' and '1+3' are two names for the same thing, as becomes "evident" once the contents of '1' and '3' are exhibited to give

$$\backslash \backslash \backslash = \backslash \backslash \backslash$$

Hilbert goes on to discuss just how far the "intuitive method", as he calls it, can take us. It will deal with verifications of E1 and E2, certain uses of variables in meta-language schemata such as $\alpha + \beta = \beta + \alpha$, and even the argument of Euclid for the infinity of the primes with the conclusion that for any prime p , there is a larger prime between p and $p!+1$. But not, Hilbert maintains, the same argument with the weaker (!) conclusion that for any prime p there is a greater prime, or for genuine uses of variables as in the algebraic identity $a+b = b+a$, ie (a)(b) $a+b=b+a$. These statements are not finitary.

³⁴ Hilbert [1925: 196]

³⁵ Hilbert [1927:470]

³⁶ Hilbert [1925:192]

Finitary statements may be conjoined and disjoined yielding finitary statements, but the situation changes once generality is introduced through variables or quantifiers. Then, the class of real statements is closed under neither logical operations (negation) nor classical rules of inference (existential generalisation). Hilbert allows the schema³⁷

$$[G] \quad \alpha + 1 = 1 + \alpha$$

as finitary because it is viewed as asserting that any finitary equality obtained by substituting a numeral for α is true. Hilbert treats the universal generalisation implicit in this statement as a kind of hypothetical. Here there is a rather unclear aspect of Hilbert's notion of the finitary, apparently tied to the contrast in ordinary language between 'any' and 'every'. Hilbert's thought is that in any particular passage of mathematics such a schema is used with only finitely many instantiations. Thus pragmatically, it is not equivalent to the full universal generalisation with which it would generally be identified. This reckoning of such generalisations as finitary will be crucial to Hilbert's program, since the assertion that a specific formal system is consistent has this form. Nevertheless Hilbert will not allow that [G] can be negated with a finitarily meaningful result - "*one cannot, after all, try out all numbers*".

Hilbert regards the statement

$$[E] \quad a + b = b + a$$

as part of algebra and therefore an ideal structure. What is the difference from [G]? It seems to consist in the fact that in algebra this formula can be manipulated: "we can derive from that formula other formulas to which we do ascribe meaning." It is a "formal structure" because we manipulate it according to its form, not its content³⁸.

³⁷ N.B. With **German** letters in Hilbert.

³⁸ There is no magical difference between German and Latin letters (or Greek and Latin as I have presented it); the difference must really consist in the different uses to which mathematicians put these similar signs. Hilbert is assuming that everyone uses his notation - a common ploy.

Another way to leave the domain of finitary statements is, paradoxically, by weakening existential claims. From

$$1 + 3 = 4$$

we can infer

$$(\exists x)(x + 3 = 4)$$

which is not finitary! Hilbert also gives the example of the two formulations of Euclid's theorem mentioned above.

Obviously the nonfinitary statements can only be used if they do not conflict with the finitary ones we accept, so that some kind of consistency must be established if their use is to be justified. This is not enough for Hilbert, though, since having granted that they are not "meaningful", there is clearly a puzzle about their usefulness even if we are assured that they are not actually dangerous. Hilbert's (partial) answer to this is the idea of conservativeness taken up later by Field: not only do the nonfinitary statements not lead to contradictions, they don't even lead to finitarily inaccessible statements. That is, any finitary statement with a proof involving nonfinitary statements also has a proof which is free of them. Such a purely finitary proof might be very long or hard to find, and that is the essential point of nonfinitary statements for Hilbert - to streamline our proofs. Hilbert adhered to classical logic in the sense that he accepted that a contradiction implies everything. So the consistency of the nonfinitary statements with the finitary statements is equivalent for formal systems as Hilbert conceived them to the unprovability of the negation of some one theorem, say $0=0$. The aim of the Hilbert program then is to give a finitary proof of the theorem that $0 \neq 0$ cannot be proved in formal arithmetic. It turns out that consistency is the key property required to establish conservativeness in most interesting cases, so that the main aim of the Hilbert program became the creation of a finitary consistency proof for formal arithmetic³⁹. Kitcher blithely says that "of course a consistency of this special type

³⁹ It is not quite clear to me why Hilbert was so sure of this, though he indicates why it is a reasonable belief in his discussion about Fermat's Last Theorem.

(a finitary proof that $0 \neq 0$ is not provable)) proves much more, for it shows that ideal statements do not generate any new finitary theorems (conservation)". But this is a nonsequitur; a finitary proof that any proof of $0 \neq 0$ has a finitary counterpart, as sought by Hilbert, does not guarantee that *any proof whatever* has one. That depends on the system in question. I do not propose to pursue the Hilbert program here, I am interested only in the epistemology underlying it.

The objection to Hilbert's position, that his finitary statements cannot really be as secure as he claims, has been made recently by Resnik⁴⁰ and by Kitcher⁴¹. Kitcher interprets Hilbert, following Kant, to claim that we obtain knowledge about numbers by looking at stroke-symbols, for example

[*] \ \ \ \ \ \ \ \ \ \ \

provides us with a way to see at once that

[**] $2 + 3 = 3 + 2$

This is because '3' refers to \ \ \ and '2' to \ \, while presumably (though he does not say this) '+' refers to juxtaposing. Kitcher claims that this is implausible because there are only two plausible answers to the question exactly how [*] gives us knowledge of [**], neither of which will do. They are that \ \ is the number 2, or else that it is an "accurate representation" of it. Difficulties are easily raised for both of these suggestions, such as: are there then more than one 2? How are different instances of '\ \ ' related? Or else, what exactly is the relation between \ \ and '2', and how do we apprehend it?

Resnik similarly poses the problem that Hilbert's concrete objects cannot really be heaps of ink, else apart from the problems already posed by Kitcher, there would be a pragmatic limit to what was surveyable reliably without tools and inferences, and so an upper bound to the n and m for which $n+m=m+n$ was finitary. He suggests we avoid this problem by reinterpreting Hilbert to take our cognisance of concrete numerals to give us knowledge of abstract symbol types through

⁴⁰ Resnik [1980: 98ff]

⁴¹ Kitcher [1976]

apprehension of the patterns their instances show. Nothing in Hilbert supports or refutes this suggestion; he hardly discusses abstraction. Kitcher's dilemma relies on the standard semantics for number terms, for it relies on the conception of numbers as individual specific objects to which reference may be secured as it may to chairs. As he concludes himself, all that it shows is that Hilbert needs a nonstandard semantics, and I agree with that. He does not pursue the issue further whether such a thing is possible, because he thinks that there is a knockdown argument against the concept of intuition on which Hilbert is relying and which leads to the need for such a non-standard semantics. The reason is that Hilbert requires "contentual induction", that is, induction used informally in a general argument about proofs, to be finitary, so that a finitary consistency proof of $0 \neq 0$ is in principle possible. Kitcher denies that induction can be finitary. Hilbert's claim is that we can know

$$\alpha + 1 = 1 + \alpha$$

because in any case where α is replaced by a numeral and then that by a strokesymbol, we know how to reduce each side stroke by stroke until we return to $1+1=1+1$. For Hilbert this is merely the converse of the "intuitive" construction of the numerals. Kitcher claims that this procedure requires us to be able to intuit a "general stroke symbol $\backslash \dots \backslash$ ". Then he asks how the dots '...' are to be interpreted. If we think of them as being replaced by strokes in a given particular case, then this only yields knowledge for that particular case, and we don't get knowledge about all numbers. Even this may be granting too much, since some strokesymbols are too long to intuit. Alternatively, if we leave the '...' in the symbol then we don't have concrete details to survey and intuit to enable us to secure the desired evident equality $1+1=1+1$. How would we know we hadn't started from $100+1$ and $99+1$ with both 100 and 99 represented by $\backslash \dots \backslash$?

Hilbert says nothing about this kind of point, but I believe he could consistently reply as follows. First, he did not introduce the "general strokesymbol", Kitcher did, and its difficulties are Kitcher's not Hilbert's. Second, such a device is not necessary, any more than it is necessary as an intermediary between a symbol

such as α and a number such as 100 which may be substituted for it in some other context. Either we wish to treat of a particular instantiation, or we don't; if we are reasoning in general, we use algebra and letters; if we have a specific case such as $100+1=1+100$ then we may write it out in full. There is no third situation. Third, Hilbert could reiterate the formal/contentual distinction between knowing the schema $\alpha+1=1+\alpha$, and knowing the general thesis $(a)a+1=1+a$. A fourth reply would be to expand Hilbert's point that his procedure is merely the converse of the construction of numerals. How do we know what the next numeral is going to be? Here we make contact with Wittgenstein on following a rule, and I now propose to deal with these objections to Hilbert's epistemology by pointing out some features of the use of text which contrast sharply with the observational paradigm which Craig proposed as the refutation of alternative mathematics. So in the next section I shall return to the argument which was broken off in Chapter 11 about the observation and creation of necessity.

§148 Observing pawns and surveying signs

Craig objects to the Wittgenstein account of necessity that we have no choice about $5+7=12$, for in a situation where we have the best possible evidence of 5 white pawns and 7 black ones then reporting 13 altogether *must* be wrong. On the other hand, Kitcher objects to Hilbert that observing 5 strokesymbols and 7 strokesymbols to be the same as 7 and 5 does not have the security which Hilbert pretends, and on which he founds the necessity of mathematics. In this section I shall argue that my textual way of looking at 'looking at' vindicates both Wittgenstein and Hilbert.

The fundamental point is this. As I already pointed out in discussing Craig's argument in Chapter 11, physical, three-dimensional objects are observationally inexhaustible. On the contrary, signs in notation systems are not. N-signs are such that we can "see every aspect of these objects, and their properties, differences...as something which cannot be reduced to something else and requires no reduction."

This was Hilbert's requirement on the 'concrete objects', signs, used in mathematics. His use of the term 'object' is confusing; he is speaking of instances of signs. As physical objects, these are three-dimensional; as signs, they are not.

Surveying N-signs is not like observing physical objects because the result can be final. That character at the end of this sentence just *is* a capital P. But that object I see may be a pawn, but it is black one of such and such a height, so and so density, and so on. Signs are fully determinate as objects never are in our reports of them. This determinateness is the basis of the realms of bivalence we create using notation. The basis for the discrimination of signs from one another is their geometrical design. That is why Wittgenstein insists that proofs are designs, and why Hilbert insists that numerals like diagrams are expressions of our space intuition⁴². Notation consists of written diagrams, figures are graphic formulas. Notice the excellent "ambiguity" of the word 'figure' in English which expresses this point perfectly. Here Hilbert and Wittgenstein are at one.

The specifics of mathematical praxis which I have spelt out in Part II explain how it is possible for us to achieve these cognitively advanced perceptions of text. Mathematical text is composed so that Words draw attention to the Diagram or Notation which displays the facts in question. As a Reader one is acquainted with the fourfold mathematical sign-system and can use the W/N contrasts to grasp what is presented, and further paragraphic connections of the signs used in the present text to the rest of the archive where most of their sense is sedimented. To be a reader presupposes one's place in the processes of mathematical praxis, and the Archive and mathematised world in which readers of mathematics are produced. To be specific, one learns bivalence by learning the alphabet. This letter is a b not a p and that's all there is to it! The drill which permits children to learn their letters and numerals is the psychological basis for the many further layers of bivalence which are gradually built up in one's education. (Talk of right and wrong, of mine and thine, of mummy and daddy, and such like no doubt

⁴² Hilbert [1902: 443]

contributes to the force of these polychotomising practices. Good little boys learn their letters.) Children learning to count make lots of errors by not pairing numerals to objects as we do - but not only children. Many important political questions turn on what *counts* as a citizen (Aboriginals were not counted at a census in Australia until quite recently) or as a person (zygotes?) or as a case of tax evasion. It is not that concepts like person are inherently vague and those like pawn inherently sharp, all empirical predicates are vague. Rather, people learn to count by practising *subject to correction* on objects whose individuation is generally agreed upon.

Now Hilbert is in fact *wrong* to reduce the content of ' $3+2=2+3$ ' to the equinumerosity of certain strokesymbols, as I shall explain in moment. But this error is overshadowed by his realisation of two crucial facts. First, *written* signs are the foundation for arithmetic. Only, not strokesymbols: these were already displaced by more advanced notation by the time writing and accounting had evolved to recognisable form. Second, our reports on these exhaustively surveyable signs are *certain* in a way observation never is.

What Hilbert overlooked is the true role of introduced notation in the consolidation of this certainty. Being paranoid about the ontological implications of some notation, he tried to remove all of it from playing any real role. But the strokes themselves are still notation, so this attempt is doomed to failure; and the reductionist attitude which continues in his instrumentalism about ideal sentences represents mathematics as no more than a device to abbreviate thoughts which we could have anyway. Whereas Klein is quite correct about the creative role of notation: our present concepts, particularly those of mathematical physics with which Hilbert was particularly concerned, are *inexpressible* without their actual mathematical form.

To argue for this point I turn to definition again, and take up the account of definition as a text act from Chapter 5. We must now consider in what ways mathematical definitions are a specific kind of text act. The short answer is that

they are so when they introduce new notation. In such cases they are *not* mere abbreviations. This argument constitutes the third leg of the plan outlined in §136 for my account of the cognitive peculiarity of mathematics.

§149 The logicist muddle about definition

The official viewpoint of Frege, Russell and other logicians⁴³ is that definitions are merely abbreviations, that is, they are theoretically superfluous introductions of symbols to stand in for long expressions which occur often. As such they are not part of the development of the subject matter of the theory, being purely syntactic devices. They are a gesture of convenience to our limited cognitive grasp. As Robinson points out⁴⁴ about Russell, this is flatly contradicts other things which he says about definitions, such as that they analyse important concepts into already accepted terms. A definition cannot both simply abbreviate a string of signs and also analyse a concept. Similarly, Frege has some difficulty reconciling the same official view of definition with two different roles of a defining sentence. At its first appearance a definition is trivially analytic; but later it somehow becomes informatively so. How can this be? This disquiet hints at there being something crazy about the logicist project - surely we don't seek really an understanding of mathematics in which neither "number" nor "multiplication" figure?⁴⁵ Perhaps the logicist is only saying that they really do, but in a way we find hard to recognise. It is common for logicians to try to have it both ways about their definitions, to be keen that they be accepted as advances in understanding of the concepts they define, while insisting that it is indeed our pre-existing concepts which have been analysed. What the logicist definition attempts is *simultaneously* to "merely" introduce a symbol to abbreviate a symbolic expression, and also to give the essence of an

⁴³ Cf Frege [1971: 159-181], Quine [1936: 329,330].

⁴⁴ Robinson [1954: 195]. This book is valuable in that it recognises and clarifies various kinds of definition without trying to conflate them, in particular several different kinds of definition which occur in mathematics.

⁴⁵ A rather similar problem about the role of definition in Intuitionist mathematics was pointed out in §33.

important concept. This point has been made⁴⁶ against Quine's claims about "translation into canonical notation", but he admits⁴⁷ that regimentation does require a change in our concepts, but an advantageous one: we can do with the replacement all the things and more that we could do with the replaced, without the philosophical worries of the latter. 'Replacement' is a misleading term however, since the old discourse is not removed and usually continues to expand.

But cannot the logicist avoid all this criticism by saying that the analysis consists in creating the right-hand side of the definition and the abbreviation is the introduction of the left-hand side to abbreviate it? This latter step is surely epistemologically harmless, and so the former is what should be discussed. But this will not do, for it presupposes that the definitions in question are of the form "man is a featherless biped", whereas many interesting ones cannot be put into this form. Rather, we are given a sign for a known term, e.g. x for multiplication, and a procedure for eliminating it from discourse (the recursive definition that $a * 1 = a$ and $a * (n+1) = a * n + a$). (In the logicist endeavour it is an old sign, but in mathematics it is usually a new one. See below.) This does not make it possible for us to consider the definition for its accuracy to the prior concept of multiplication; rather, it enables us to judge whether the interpreted system including x has the same properties as the system of natural numbers. But of course it doesn't have all the same properties⁴⁸ - or the logicist program would be pointless. So how can we judge the worth of a definition in the logicist discourse? And is it the same as the typical use of definition in mathematics? Let us start from the point that, like the logicist program, definition has a point - an illocutionary point. A definition is not simply an assertion. It is not possible to see the whole of *Principia Mathematica* as a sequence of assertions, it is necessary to recognise there linguistic items which have

⁴⁶ I have discussed this in Chapter 7.

⁴⁷ Quine [1960]

⁴⁸ Recall that arithmetic without multiplication is decidable!

other functions than those of statement. Russell himself points out⁴⁹ the *volitional* aspect of definition.

In the next section I will suggest that the notion of text act can clear up this muddle about definitions in mathematics.

§150 Definition in mathematics

Definitions in mathematics are always text acts, written. But written definitions may be simply verbal or notational. Mathematics requires the latter, and this has some further consequences which we will bring to bear on our consideration of proofs.

Early mathematical definitions were actually often otiose. Euclid never uses his definitions of a line or a point⁵⁰. So why were they made? They are made because of an illusory idea. Not the idea that everything can be defined, as is sometimes suggested - I doubt that anyone ever thought that. Rather, the idea that the elements of knowledge must be known perfectly in order that the knowledge based on them shall be as good as possible, for surely knowledge necessarily falls off in point of certitude as one moves from elementary to complex. The basis for this idea is the practice of Socrates as reported and refined by Plato, of seeking definitions of "basic terms" like justice and knowledge. The "definitions" which the Platonic school sought would, if correct, be useful, since a touchstone for these things is greatly sought. Who does not wish for a sure way to distinguish justice and knowledge from their opposites? In contrast there is no difficulty in recognising a point or a line. And, of course, the definitions of points and lines in Euclid, and equally the definitions of space and time in Newton, are neither needed for their identification, nor useful.

Definition in mathematics evolved from reports of the philosophical speech of the Academy (no doubt the Pythagoreans and others provided

⁴⁹ Whitehead and Russell [1962: preface]

⁵⁰ Cf exhibit #12A.

foreshadowings), but progressed to a useable role in mathematics as it became expressed in usable form, as it became a matter of notation. In Euclid, the things that are used again are the axioms and theorems, not the definitions. Euclid's theorems are all about equality of angles, lines etc; but his definitions do not bear on this at all. In modern definitions such as those on exhibit #12B, no attempt is made to characterise elements. On the contrary, nowadays mathematicians are given to boasting that they don't care what the elements are: if they form a group they must obey Lagrange's theorem, and so on. This modern kind of definition, that of the group but not of its elements, is an invention of the nineteenth century really; perhaps Dedekind's definition of function is the origin of it. Its twin features are that it is *reductive* (the subjects of the theory are defined entirely in terms of entirely undefined elements), and it is *formal* (the definiendum is given as a structure of elements via symbolic form). The point of the use of definitions for comparison is that they permit the checking of a symbol's correct construction. In order to be used in this way as a resource a definition in mathematics needs to be written, because the defined term is introduced by it and has no external use. Or rather, any prior use is being rejected. In mathematics one can do that; but not with terms of ordinary language. So subsequent use is to be governed by the definition because of its consequences, what you can do with it, not because it is "correct".

Of course, it is easy to find intermediate examples, definitions which partly characterise by ostension and partly by axiom. Lakatos' *Proofs and Refutations* charts in great detail the way in which successive generations of mathematicians were forced to change the acceptable *kind* of definition for 'polyhedron'. The role of examples in the definition was gradually reduced to zero⁵¹. In that case, polyhedra were in the end defined entirely in terms of their elements - which are in the final formal theory left undefined. In teaching about geometry or polyhedra, it is necessary partly to retrace the phylogeny of the concept ('new maths' ignored this

⁵¹ Though of course it remains in the background as motivation - this is a characteristic process of mystification in mathematics, encouraged by replacement ideas.

fact); so definitions of all those kinds which have been needed historically can be found in current texts of different levels of sophistication. It would be a mistake to think that only those at the highest level of sophistication count as mathematics.

Frege objected to definitions through sets of axioms, because they do not allow one to "solve" the definition to express it in the form $A = BCD$. (There is some theory about when this can be done.) This is an objection with the same motivation as the Intuitionist rejection of a description of number which cannot be effectively evaluated. The two objections are objections to "creative" definition⁵¹, to taking objects to exist in mathematics if their definitions lead to no contradiction. The objection is to introduced signs which cannot be eliminated; yet these are the only signs which are really important! *New ideas need new signs.*

So there are two reasons why definitions in mathematics need notation. First, defined signs need to be free of independent meaning, for which empty signs are needed. That is to say, in order to achieve the complete characterisation of the signs by the axioms and explicit rules for inferences, which together determine their proper use in the construction of complex signs, they must not be familiar terms with established meanings. Secondly, they need to be surveyable, so that these later complex signs can be justified by reference to the axioms, definitions and rules.

§151 What do definitions and notations achieve?

The use of definitions in mathematics, deploying new notation in significant cases, is one way in which the demand for surveyability is met. The effect of this procedure is the extension of our concepts. Thus the most obvious feature of notation, abbreviation, and the most significant, the Klein point that it permits standing on the shoulders of giants, are actually two aspects of one process. There are two importantly different kinds of definition in mathematics, those like the definition of the dimension of vector space, or the definition of the field of

⁵¹ For Frege's views on definition, see Frege [1970:159-181]

complex numbers, which specify an individual or individuals, and those like that of 'ideal', or topological space, which define a *kind* of individual. In this section I will illustrate the foregoing discussion with examples of both kinds.

First, numerals. Numerals were of course not defined, but introduced over a long time informally as part of the process of inventing writing. But their advantages over strokesymbols (which are a modern idealisation of forms of tally which were widespread before writing⁵³) are at least two. First, they permit addition to be carried out for much larger numbers because larger numbers are surveyable in decimal notation. Strokesystem arithmetic loses its evidence fairly rapidly. Of course, numerals too are inadequate for calculations for very large numbers, so that further notational developments become necessary (for example scientific notation - see below). But it does not show that numerals are not capable of providing evidence for a range of calculations, that it is a *limited* range, nor that they are not an improvement on strokesymbols because they are not perfect. Their abbreviatory character is of value. But it also creates new possibilities, and this is their second advantage. The specific form which the abbreviatory nature of numerals takes, the place value system, is essentially recursive. It makes possible the development of algorithms for the recursive operations of addition, multiplication and so on. Stroke symbols can be added, though the surveyability is limited. To multiply them is perfectly possible, but to do it one has to be in possession of the ability to count groups. The algorithms for addition and multiplication can be mimicked in strokesymbols - but what's important is introducing these practices, newly drawing attention to these possibilities (text acts can be mimicked in speech too.) (It is the basic illusion of reductionism to show how to explicitly do simple things with reduced resources - when the one showing has the more complex cognitive resources. This doesn't really show what doing the simple thing needs.)

⁵³ Cf Menninger [1969]

Let us secondly consider multiplication. Reductionists say that it is "only" repeated addition. And so it is, originally. Yet once again it has two kinds of advantage over performing the same calculations by repeated addition. The calculation using multiplication is shorter and more perspicuous (to one who has learned to multiply!) and therefore extends further the domain of the surveyable. And the use of explicit signs for multiplication creates new possibilities - the introduction of exponents, for example. Or multiplication by *fractions*. Very many children have a problem at just this point, because they are given too strongly to understand that multiplication is repeated addition. "How can you add something two and a half times?" The most important criticism of reductionism may well be that it disguises the fact that in learning mathematics one has to make leaps like those which were made historically in the introduction of new ideas through new notational devices. Very many people cannot deal with negative numbers because their concepts of number have been bound too closely to counting and matching.

Third, exponentiation. No doubt, one could reductively find an extremely cumbersome and unperspicuous way to operate with exponents in strokesymbols. The extension of the domain of the surveyable calculation is now very obvious however. It is easy to write down in a handful of signs using repeated exponentiation, numbers whose ordinary numerals would fill the observable universe if written out; their strokesymbols would be even harder to use. Once again we can point to the two sides of this innovation⁵⁴. The abbreviatory power is considerable and in itself important; yet the new possibilities which it opened up are even more significant. The inventions of exponential functions and of logarithms was constitutive of vast areas in modern mathematics and underlay the possibility of calculations which made possible the Newtonian science. The point here is again that once the notation is in use, it can be used in new ways: the exponent

⁵⁴ In this case the introduction of the new notation is more closely related to definition, because its first users, particularly Descartes, were aware of its novelty and had to choose ways to represent it. The gradual assimilation of notational innovation into definitions is almost complete in mathematics now.

can be a fraction (logarithms) or a variable (functions). In retrospect, one might see that strokesymbols can be used in ways which mimic logarithmic calculations. But not by creatures like you or me; and no-one would have thought of it without the notational history leading to logarithms. To say that these possibilities are really there in the strokesystem is like saying that chess already existed, it just needed to be written down. The only sense in which this is true is a trivial one: you can only write down what can be written down.

My fourth example is the dimension of a vector space. This case is one where the definition and associated notation (see exhibit #12B) very plainly draw explicit attention to something which is essentially a theorem - the fact that any two bases of a (finite-dimensional) vector space are equinumerous. The definition would be nonsense if this were not true. So the definition highlights this fact by explicitly *labelling* that common number of basis elements. It turns out that this is a good way - perspicuous - to organise facts about vector spaces.

Fifth, the definition of an ideal. The original introduction of ideals by Kummer was a radical generalisation of the concept of number which made possible great advances in the understanding of divisibility. Yet it could be said to be no more than an abbreviation for "set of numbers obtainable by linear combinations from a given starting set" (the modern definition is a revised form). Its importance derives from the fact that a definition like the one in the exhibit makes it possible to manipulate such sets of numbers in informative ways, according to rules which can be derived from those in the definition. The concept of prime number was *extended* by Kummer's new notation.

To say that ideal theory can be reduced to strokesymbols might be in some sense true, but is nevertheless epistemologically absurd. This example and the others, illustrate how the use of definitions involving notation extend our cognitive grasp, extend what we can survey. Bachelard made popular the term 'epistemological break' to refer to those irruptions in the semantics of scientific discourse which Kuhn later in a different tradition termed paradigm shifts. In

mathematics this seems inappropriate because, as Kitcher was forced to concede, old theories live on in mathematics. Recalling the "shoulders of giants" idea, I propose the term *epistemological hoist* for these innovations to our signsystems which have the effect of creating the kind of elevation of our grasp which later must be rationalised by reductions. Reductions are hindsight's compliments to epistemological hoists. Notation's most important role, and the one to which Klein was pointing⁵⁵, is that of mediating these hoists in mathematics.

§152 Semiotic ascent

Despite the fact that to a certain kind of mind it is attractive to declare that the rational numbers can be reduced to the natural numbers, and so on, the ontological miserliness which motivates such an attitude in fact disguises something more important for understanding mathematical knowledge. Such reductions are found only after the event and are, as they are often called, rational reconstructions. The process of *extending* our knowledge does not consist of such reconstructions. It is largely achieved in mathematics through new proofs. New proofs entrain new signs - if they do not introduce them, they will be introduced in later use of the proof (as in the Hansen example - "hansen rings").

The role of notation should be seen in historical perspective, as a particular technique continuing a number of modes of development which go back into the mists of prehistory. They begin with exclusion. Exclusion is a social practice at least as old as language, with examples among animals. The very differentiation of animal from environment is based on an exclusion. Among mammals, sexual exclusion is the rule. In language, mutual exclusion is evident from its earliest forms; the colour exclusion problem (why *must* what is red not be green?) is evidence of the workings of normativity and exclusion in the most ordinary of ordinary language. I take the obduracy of this problem against attempts to "explain" necessity in ordinary language as evidence that there are various degrees of

⁵⁵ Cf the quotation from Klein in § 24 above.

necessity. The standard appropriate to mathematics has been generally taken as its essence, misleadingly.

But in fact, my point is, there is no timeless essence of the necessity even of mathematics. The initial normativity in mathematics is introduced through counting, but it takes its special character from the systematic exclusions set up in the practices surrounding letters and numerals. As Goodman almost observed, it's primarily the marks we make into instances, only secondarily the works we label with them. But this first and most characteristic hoist, that into writing, is only the beginning of the development of mathematical necessity. The role of notation is to mediate that development. Notation must of course work to have any effect; it does so in proofs. This is my interpretation of Wittgenstein's idea that *proofs make essence*. Mathematics develops in mathematical praxis through the constant notational innovation which the mathematical sign-system $\langle W, D, N, P \rangle$ makes possible for workers in the mathematical context $\langle A, R, M \rangle$.

Thus I complete my argument for the claim that mathematics is cognitively peculiar in that it includes evidently justified, novel textual truths. Notation's contribution is in its making these truths bivalent, by its maintaining the surveyability of proofs, and by its making possible the introduction of novelty through mathematical definitions.

§153 Concluding remarks

Throughout this thesis and explicitly in Part III, I have argued that improved ideas about text, context and praxis, which I developed in Part II, can illuminate the problems of the philosophy of mathematics which I identified in Part I. In trying to address such a large canvas my brush has inevitably at times been rather broad, and one of the most obvious ways in which this work could be developed is in taking the individual arguments further. For example the notion of 'textual truth' could obviously be developed considerably further from my discussion in this chapter. Historical studies of the development of particular

systems of notation, further investigation of the way the logico-formalist hegemony arose, flourished and obtained its position, studies of mathematical praxis "in the field" , and studies of the actual deployment of the mathematical sign-system in important groups of text, are all avenues which I intend to go explore in the future. But for the present my text ends here.

Bibliography

- Alston W P
[1958] "Ontological commitment"
Phil Stud 9: 8-17
- Ambrose A and Lazerowitz M (eds)
[1972] *Ludwig Wittgenstein Philosophy and Language*
London, George Allen and Unwin
- Ambrose A
[1972] "Mathematical Generality"
in Ambrose and Lazerowitz [1972: 287-318]
- Annas J
[1976] *Aristotle's metaphysics M and N*
Oxford, Clarendon Press
- Asquith P and Kyburg H (eds)
[1979] *Current Research in philosophy of science*
East Lansing, Philosophy of Science Association
- Austin J L
[1962] *How to do things with words*
Oxford, OUP
- Ayer A J
[1936] *Language, truth and logic*
London, Gollancz
- Bach K and Harnish RM
[1979] *Linguistic Communication and Speech Acts*
Cambridge, Mass., MIT press
- Baker G and Hacker P
[1980] *Wittgenstein meaning and understanding*
(An analytical commentary on the Philosophical Investigations v1)
Oxford, Blackwell
[1984a] *Language sense and nonsense*
Oxford, Blackwell
[1984b] *Frege: Logical Excavations*
Oxford, Blackwell
[1984c] *Scepticism, rules and language*
Oxford, Blackwell
[1985] *Wittgenstein: rules, grammar and necessity,*
(An analytical commentary on the Philosophical Investigations v2)
Oxford, Blackwell
- Bar-Hillel Y (ed)
[1965] *Logic Methodology and Philosophy of Science*
Amsterdam, North-Holland
- Barthes R
[1967] *Elements of Semiology* (trans Lavers and Smith)
London, Cape
[1974] *S/Z* (trans R Miller)
New York, Hill and Wang
[1983] *The Fashion System* (trans Ward and Howard)
New York, Hill and Wang
- Barwise J (ed)
[1977] *Handbook of mathematical logic*
Amsterdam, North-Holland
- Beeson M
[1985] *Foundations of Constructive Mathematics*

- Berlin, Springer
- Bell E T
 [1945] *The Development of Mathematics*
 New York, McGraw-Hill
- Benacerraf P
 [1973] "Mathematical Truth"
 in Benacerraf and Putnam [1983: 272-294]
- Benacerraf P and Putnam H
 [1964] *The Philosophy of Mathematics*
 Cambridge, Blackwell
 [1983] *The Philosophy of Mathematics*, 2nd ed.
 Cambridge, CUP
- Bernays P
 [1959] Review of Wittgenstein [1956]
 Ratio 2: 1-22, reprinted in Benacerraf and Putnam [1964]
- Bertin J
 [1983] *Semiology of Graphics* (trans Berg W J)
 Madison Wisconsin, U Wisconsin Press
- Bishop E
 [1967] *Foundations of Constructive Analysis*
 New York, McGraw-Hill
 [1970] "Mathematics as a numerical language"
 in Kino [1970: 53-71]
- Black M
 [1937] "Vagueness"
 Phil Sci 4: 427-55
 [1970] "Reasoning with loose concepts"
 in *Margins of Precision*, Ithaca NY, Cornell UP
- Blackburn, S (ed)
 [1975] *Meaning, Reference and Necessity*
 Cambridge, CUP
- Block N (ed)
 [1981] *Imagery*
 Cambridge Mass., MIT Press
- Bloomfield, L
 [1933] *Language*
 New York, Henry Holt
- Bloor D
 [1978] "Polyhedra and the abominations of Leviticus"
 BJPS 11: 245-272
 [1983] *Wittgenstein: a social theory of knowledge*
 London, Macmillan
- Boolos G and Jeffrey R
 [1980] *Computability and Logic* 2nd ed.
 Cambridge, CUP
- Bostock D
 [1974] *Logic and Arithmetic I*
 Oxford, OUP
 [1979] *Logic and Arithmetic II*
 Oxford, OUP

Boyer C

- [1968] *A History of Mathematics*
New York, Wiley

Bracewell R N

- [1973] "The opening message from an extra-terrestrial probe
Astr. and Aeronautics 11.5: 58
[1974] *The Galactic Club*
San Francisco, San Francisco Book Company

Brittan G

- [1986] "Towards a theory of theoretical objects"
in *Fine* [1986 v1: 384-393]

Brouwer L

- [1907] *On the foundations of mathematics* (thesis)
in [1975: 11-101]
[1912] "Intuitionism and Formalism"
in Benacerraf and Putnam [1983: 77-89]
[1923b] "On the significance of the principle of excluded middle.."
in [1975: 34-345]
[1927] "On the domains of definitions of functions"
in [1975: 446-463]
[1927a] "Intuitionistic reflections on formalism"
in [1975: 490-492]
[1928] "Mathematik, Wissenschaft and Sprache"
in [1975:]
[1975] *Collected Works* v1 (ed Heyting A)
Amsterdam, North-Holland
[1981] *Brouwer's Cambridge Lectures, 1931* (ed van Dalen D)
Cambridge, CUP

Brown G and Yule G

- [1983] *Discourse Analysis*
Cambridge, CUP

Bunn J H

- [1981] *The Dimensionality of Signs, Tools and Models*
Bloomington, Indiana U P

Bynum T W (trans, ed and intro)

- [1972] *Gottlob Frege: Conceptual Notation and related articles*
Oxford, OUP

Cajori F

- [1929] *A History of Mathematical Notations*, 2nd ed
London, Open Court

Campbell N

- [1957] *An account of the principles of measurement and calculation*
New York, Dover

Campbell R

- [1974] "The sorites paradox"
Phil Stud 26: 175-191

Cargile J

- [1969] "The sorites paradox"
BJPS 20: 193-202

Carnap R

- [1937] *The logical syntax of language*
London, Routledge and Kegan Paul

- [1939] *Foundations of Logic and Mathematics*
in Int Enc of Unified Science v1#3
- [1950] *Logical foundations of probability*
Chicago, Chicago UP
- Castaneda H-N
[1959] "Arithmetic and reality"
Aust J Phil 37: 92-107
- Castonguay C
[1972] *Meaning and Existence in Mathematics*
Vienna, Springer (Library of Exact Philosophy 9)
- Chihara C S
[1973] *Ontology and the Vicious-circle principle*
Ithaca NY, Cornell UP
[1982b] "A gödellian thesis regarding mathematical objects:
do they exist? and can we perceive them?"
Phil Rev 91: 211-227
- Christiansen B, Howson A G and Otte M
[1986] *Perspectives on Mathematics Education*
Dordrecht, Reidel
- Church A
[1956] *Introduction to Mathematical Logic*
Princeton NJ, Princeton UP
- Cleveland W S
[1985] *Elements of Graphing Data*
Monterey California, Wadsworth
- Cleveland W S and McGill R
[1985] "Graphical perception and graphical methods
for analyzing scientific data"
Science 229: 828-833
- Cohen P
[1963] *Set Theory and The Continuum Hypothesis*
New York, Benjamin
- Coulmas F and Ehlich K (eds)
[1983] *Writing in focus*
Berlin, Mouton
- Craig E J
[1975] "The problem of necessary truth"
in Blackburn [1975: 1-31]
[1985] "Arithmetic and fact"
in Hacking [1985: 89-112]
- Cresswell
[1973] *Logics and Languages*
London, Methuen
- Currie G
[1982] *Frege an introduction to his philosophy*
Brighton, Harvester
- Curry H B
[1952] *Outlines of a Formalist Philosophy of Mathematics*
Amsterdam, North-Holland
[1963] *Foundations of Mathematical Logic*
New York, McGraw-Hill

- Curwin J and Slater R
 [1984] *Quantitative Methods for Business Decisions*
 Wokingham England, Van Nostrand Reinhold
- Davidson D
 [1968] "On saying that"
 Synthese 19: 130-146
- Davis P and Hersh R
 [1980] *The Mathematical Experience*
 Boston, Birkhäuser
- Derrida J
 [1971] "Signature Event Context"
 reprinted in Derrida [1982: 307-330]
 [1974] *Of Grammatology* (trans G C Spivak)
 Baltimore, Md., Johns Hopkins UP
 [1977] "Limited inc"
 Glyph 2: 162-254
 [1978] *Edmund Husserl's Origin of Geometry* (trans Leavey J)
 Stony Brook NY, Hays
 [1982] *Margins of Philosophy* (trans Bass A)
 Brighton, Harvester
- Detlefsen M
 [1979] "On interpreting Gödel's second theorem"
 J Phil Log 8: 297-313
 [1986] *Hilbert's Program*
 Dordrecht, Reidel
- Detlefsen M and Luker M
 [1980] "The four color theorem and mathematical proof"
 J Phil 77: 803-820
- Devitt M
 [1981] *Designation*
 New York, Columbia UP
- Diamond C
 [1984] "What does a concept script do?"
 Phil Q 34: 343-368, reprinted in Wright [1985]
- Dressler WU (ed.)
 [1978] *Current Trends in Textlinguistics*
 Berlin, deGruyter
- Dressler W and Beaugrande R de
 [1981] *Introduction to Text Linguistics*,
 London, Longman
- Dummett M
 [1959] Review of Wittgenstein [1956]
 in Dummett [1978]
 [1973] *Frege: philosophy of language*
 London, Duckworth
 [1975a] "The Philosophical Basis of Intuitionistic Logic"
 in Benacerraf and Putnam [1983: 97-129]; also in Dummett [1978]
 [1975b] "Wang's paradox"
 in Dummett [1978]
 [1977] *Elements of Intuitionism*
 Oxford, OUP
 [1978] *Truth and Other Enigmas*
 London, Duckworth

Eco U

- [1976] *A Theory of Semiotics*
Bloomington, Indiana UP
- [1979] *The Role of the Reader*
Bloomington, Indiana UP
- [1984] *Semiotics and the Philosophy of Language*
London, Macmillan

Edwards P (ed)

- [1965] *Encyclopedia of Philosophy*, 8 volumes
New York, Macmillan

Eisenstein E

- [1979] *The printing press as an agent of change*
Cambridge, CUP

Ellis B

- [1968] *Basic Concepts of Measurement*
Cambridge, CUP

Esenin-Volpin

- [1970] "The ultra-intuitionistic program and the antitraditional program
for foundations of mathematics"
in Kino [1970: 3-46]

Euclid

- [1957] *The Elements*, trans and ed by T H Heath,
New York, Dover

Fann K T (ed)

- [1967] *Ludwig Wittgenstein, the man and his philosophy*
New Jersey, Humanities Press

Ferrara A

- [1980] "Appropriateness conditions for entire sequences of speech acts"
J Pragmatics 4: 321-40

Field H

- [1980] *Science without Numbers*
Oxford, Blackwell
- [1982] "Realism and anti-realism about mathematics"
Phil Topics 13: 45-69
- [1984a] "Is mathematical knowledge just logical knowledge?"
Phil Rev 93: 509-552
- [1984b] Review of Gottlieb [1980]
Noûs 18: 160-5

Fine K

- [1975] "Vagueness, truth and logic"
Synthese 30: 265-300

(ed)

- [1986] *PSA 86*
East Lansing, philosophy of Science Association

Fogelin R J

- [1968] "Wittgenstein and Intuitionism"
Amer Phil Q 5: 267-274

Fraenkel A, Bar-Hillel, Y and Levy A

- [1973] *Foundations of set theory* (2nd ed)
Amsterdam, North-Holland

Frege G

- [1879] *Begriffsschrift*
Halle, Nebert
trans in Bynum [1972]
- [1882] "On the scientific justification of a conceptual notation "
trans in Bynum [1972: 83-9]
- [1884] *Grundlagen der Arithmetik*
Breslau, Koebner
trans in Frege [1959]
- [1893-1903] *Grundgesetze der Arithmetik*
Jena, Pohle
partially translated in Frege [1964]
- [1959] *Foundations of Arithmetic*
trans Austin J L
Oxford, Blackwell
- [1964] *The basic laws of arithmetic: exposition of the system*
trans and ed by M Furth
berkeley, University of California Press
- [1970] *Translations from the philosophical writings*
trans and ed Geach P and Black M
Oxford, Blackwell
- [1971] *On the Foundations of Geometry and formal theories of arithmetic*
trans Kluge E-H W
New Haven Conn., Yale UP
- [1984] *Posthumous Writings*
ed Hermes H, Kambartel F, trans Long P and White R
Oxford, Blackwell
- [1984] *Philosophical and Mathematical Correspondence*
abridged McGuinness B , trans Kaal H
Oxford, Blackwell
- [1984] *Collected Papers on Mathematics, Logic and Philosophy*
ed McGuinness B, trans Black M et al
Oxford, Blackwell
- Freudenthal, H
[1960] *Lincos*
Amsterdam, North-Holland
- Gainnes B
[1976] "Foundations of fuzzy reasoning"
Int J Man-machine Studie3s 8: 623-77
- Gardner M
[1982] *Logic machines and diagrams* 2nd ed
Chicago, University of Chicago press
- Gasking D A T
[1940] "Mathematics and the world"
Aust J Phil 18.2: 97-116
- Geach P
[1962] *Reference and Generality*
Ithaca NJ, Cornell Up
- Gelb I J
[1963] *A study of writing*
2nd ed, Chicago, U of Chicago Press
- Gödel K
[1931] "Über formal unentscheidbare sätze der Principia Mathematica
und verwandter systeme I"
translated in Gödel [1962]
- [1947] "What is Cantor's continuum problem?"
trans in Benacerraf and Putnam [1983: 470-485]

- [1962] *On formally undecidable propositions of Principia Mathematica and related systems*
trans B Meltzer from Gödel [1931]
Edinburgh, Oliver and Boyd
- Goguen J
[1969] "The logic of inexact concepts"
Synthese 19: 325-373
- Goodman N
[1981] *The Languages of Art*
Brighton, Harvester Press
- Goodman N D
[1971] "The Arithmetic Theory of Constructions"
in Mathias and Rogers [1971]
[1984] "The knowing mathematician"
Synthese 60: 21-38
- Goodstein R L
[1972] "Wittgenstein's Philosophy of mathematics"
in Ambrose and Lazerowitz [1972: 271-286]
- Goody J
[1977] *The Domestication of the Savage Mind*
Cambridge, CUP
[1982] "Alternative Paths to Knowledge in Oral and Literate Cultures"
in Tannen [1982: 201-216]
[1983] "Literacy and achievement in the Ancient World"
in Coulmas and Ehlich [1983: 83-97]
[1986] *The logic of writing and the organisation of society*
Cambridge, CUP
- Goody J (ed)
[1968] *Literacy in traditional societies*
Cambridge, CUP
- Goody J and Watt I
[1962] "The consequences of Literacy"
reprinted in Goody [1968: 27-68]
- Griffiths H B
[1976] *Surfaces*
- Haack S
[1978] *Philosophy of logics*
Cambridge, CUP
- Hacking I
[1985] "Rules, scepticism, proof, Wittgenstein"
in Hacking I (ed) *Exercises in Analysis*
Cambridge, CUP
- Hadamard J
[1954] *The psychology of invention in the mathematical field*
New York, Dover
- Hallett M
[1979] "Towards a theory of mathematical research programs I"
BJPS 29: 1-25
[1980] "Towards a theory of mathematical research programs II"
BJPS 30:1-25

- Hansen D
 [1984] "Positive derivations on partially ordered strongly regular rings"
 J Aus Math Soc A 37: 178-180
- Hardy G H
 [1929] "Mathematical proof"
 Mind 38: 1-25
 [1963] *A Course of Pure Mathematics* 10th ed
 Cambridge, CUP
 [1967] *A Mathematician's Apology*
 Cambridge, CUP
- Hardy G and Wright E
 [1975] *An Introduction to the Theory of Numbers* (4th ed)
 Oxford, OUP
- Hartley J
 [1985] Designing instructional text
 London, Kogan Page
- Hatcher W
 [1982] *The Logical Foundations of Mathematics*
 Oxford, Pergamon
- Havelock E A
 [1963] *Preface to Plato*
 Oxford, Blackwell
- Heath T (trans and ed)
 [1956] *The Thirteen Books of Euclid's Elements*, 3vols
 New York, Dover
- Heidegger M (trans MacQuarrie and Robinson)
 [1962] *Being and Time*
 Oxford, Blackwell
- Heller P
 [1987] "The best candidate approach to diachronic identity"
 Aust J Phil 65: 434-451
- Heyting A
 [1931] "The intuitionist foundations of mathematics"
 trans in Benacerraf and Putnam [1983: 52-61]
 [1980] *Intuitionism: an introduction* 3rd ed.,
 Amsterdam, North-Holland
- Hilbert D
 [1902] "Mathematical Problems" (trans Newson M W)
 BAMS 8: 437-479, reprinted in Proc Symp Pure Math 28: 1-34
 [1905] "On the foundations of logic and arithmetic"
 trans in van Heijenoort [1967: 129-138]
 [1926] "On the infinite"
 trans in Benacerraf and Putnam [1983: 183-201]
 [1928] "The foundations of mathematics"
 trans in van Heijenoort [1967: 464-479]
 [1971] *Foundations of Geometry* 10th ed (trans Unger L)
 La Salle, Ill., Open Court
- Hintikka J
 [1968] "Are mathematical truths synthetic a priori?"
 J Phil 65: 640-51
 [1973] *Logic, Language Games and Information*
 Oxford, OUP

- (ed) [1969] *Philosophy of mathematics*
Oxford, OUP
- Hodges W
[1986] "Truth in a structure"
PAS 86: 135-151
- Holmes N
[1984] *Designer's Guide to Creating Charts and Diagrams*
New York, Watson-Guptill
- Hookway C
[1985] *Peirce*
Boston Mass., Routledge and Kegan Paul
- Hurford JR and Heasley B
[1983] *Semantics, a coursebook*
Cambridge, CUP
- Husserl E
[1936] "The Origin of Geometry"
translated in Derrida [1978]
[1970] *Logical Investigations* (trans J N Findlay)
London, RKP
- Innis H A
[1972] *Empire and Communications*
Toronto, U of Toronto Press
- Innis R E
[1986] *Semiotics; an introductory reader*
London, Hutchinson
- Jackson F
[1980] "Ontological commitment and paraphrase"
Philosophy 55: 303-315
- Jennings R C
[1986] "Tarski - a dilemma"
Inquiry 30: 155-172
- Jennings R E
[1986] "The punctuational sources of the truth-functional 'or'"
Phil Stud 50: 237-59
- Jubien M
[1977] "Ontology and mathematical truth"
Noûs 11: 133-468
[1981] "Formal semantics and the existence of sets"
Noûs 15: 165-176
- Kant I
[1933] *Critique of Pure Reason* (trans N Kemp Smith)
London, Macmillan
- Keenan E L (ed)
[1975] *Formal Semantics of Natural Language*
Cambridge, CUP
- Kennedy H (trans and ed)
[1973] *Selected works of Guiseppe Peano*

London, George Allen and Unwin

Kielkopf CK

[1970] *Strict Finitism*
The Hague, Mouton

Kino A, Myhill J and Vesley R

[1970] *Intuitionism and Proof Theory*
Amsterdam, North-Holland

Kitcher P

[1976] "Hilbert's epistemology"
Phil Sci 43:99-115
[1983] *The Nature of Mathematical Knowledge*
Oxford, OUP

Kleene S C

[1952] *Introduction to Metamathematics*,
Amsterdam, North-Holland

Klein J

[1968] *Greek Mathematical Thought and the Origin of Algebra*
Cambridge Mass., MIT Press

Klenk V H

[1976] *Wittgenstein's Philosophy of Mathematics*
The Hague, Martinus Nijhoff

Kline M

[1975] *Mathematics in Western Culture*
Harmondsworth, Penguin

Knorr R

[1986] *The ancient tradition of geometric problems*
Boston, Birkäuser

Kolers P et al (eds)

[1979] *Processing of Visible language 1*
New York, Plenum

Körner S

[1960] *The Philosophy of Mathematics*,
New York, Harper

Krantz D, Luce R, Suppes P and Tversky A

[1971] *Foundations of Measurement, Voll*
New York, Academic Press

Kreisel G

[1958] "Wittgenstein's remarks on the foundations of mathematics"
BJPS 9:135-158

Kripke S

[1975] "Outline of a theory of truth"
J Phil 72: 690-716
[1982] *Wittgenstein on Rules and Private Language*
Oxford, Blackwell

Kyburg H

[1984] *Theory and Measurement*
Cambridge, CUP

Lakatos I

- [1974] *Proofs and Refutations*
Oxford, Blackwell
- [1978a] *Mathematics, science and epistemology* (collected papers v2)
Cambridge, CUP
- [1978b] "What does a proof prove?"
in Lakatos [1978: 61-9]
- Latour B and Woolgar S
[1986] *Laboratory Life* (2nd ed)
Princeton NJ, Princeton UP
- Lear J
[1982] "Aristotle's philosophy of mathematics"
Phil Rev 4: 161-192
- Lefebvre L and Martin H-J (trans Gerard D)
[1976] *The coming of the book*
London, Verso
- Lehman H
[1970] "Queer arithmetics"
Aust J Phil 48.1:31-43
- [1979] *Introduction to philosophy of mathematics*
Oxford, Blackwell
- Lewis D
[1970] "General semantics"
Synthese 22: 18-67
- Littlewood J
[1986] *Littlewood's miscellany* (rev ed by Bollobas B)
Cambridge, CUP
- [1986a] "The mathematician's art of work"
in Littlewood [1986: 189-200]
- Lockwood A
[1969] *Diagrams*
London, Studio Vista
- Lycan W
[1984] *Logical Form in Natural Language*
Cambridge, Mass., MIT Press
- McGinn C
[1984] *Wittgenstein on Meaning*
Oxford, Blackwell
- McLuhan M
[1964] *Understanding Media*
London RKP
- Maddy P
[1980] "Perception and mathematical intuition"
Phil Rev 89: 163-196
- Margalit A
[1976] "Vagueness in vogue"
Synthese 33: 211-221
- Marshack A
[1972] *The Roots of Civilisation*
London, Weidenfeld and Nicolson

- Maxwell J C
[1910] "Diagrams"
in *Enc Brit*, 11th ed.
- Mayberry J
[1980] "A new Begriffsschrift"
BJPS 31: 213-254, 329-358
- Mehlberg H
[1962] "The present situation in the philosophy of mathematics"
in Kazemier B and Vuysje D (eds) *Logic and Language*
Dordrecht, Reidel
- Mellor H
[1966] "Inexactness and explanation"
Phil Sci 33:345-359
[1967] "Imprecision and explanation"
Phil Sci 34: 1-9
- Menninger K
[1969] *Number words and number symbols*
Cambridge Mass., MIT Press
- Meyer R K
[1987] "Curry's philosophy of formal systems"
Aust J Phil 65.2: 156-171
- Mill J S
[1843] *A system of logic*
London, Longmans
- Miller G A
[1967] *The psychology of communication*
New York, Basic Books
- Miller J P
[1982] *Numbers in Presence and Absence*
The Hague, Martinus Nijhoff
- Miller L W
[1975] "Kant's philosophy of mathematics"
Kantstudien 66: 297-308
- Montague R
[1970] "English as a formal language"
repr in Montague R *Formal philosophy*
(Papers ed by R Thomason)
New Haven Conn., Yale UP
- Morris A and Stewart-Dore N
[1984] *Learning to learn from text*
North Ryde, Addison-Wesley
- Mortensen C
[1988] "Quantities"
forthcoming
- Mueller I
[1970] "Aristotle on geometrical objects"
Arch für Gesch der Phil 52:
- Mulkay M
[1985] *The Word and the World*
London, George Allen and Unwin

- Mumford, L
[1961] *The city in history*
London, Martin Secker and Warburg
- Nagel E
[1979] *Teleology Revisited*
New York, Columbia UP
- Newman J R
[1956] *The world of mathematics*
New York, Simon and Schuster, 4 vols
- Northrop E
[1960] *Riddles in mathematics*
Harmondsworth, Penguin
- Novitz D
[1977] *Pictures and their use in Communication*
's-Gravenhage, Martinus Nijhoff
- Ong W J
[1982] *Orality and Literacy*
London, Methuen
- Otte M
[1986] "What is a text?"
in Christiansen [1986:173-204]
- Peano G
[1915] "On the importance of symbols in mathematics"
in Kennedy [1973: 227-234]
[1921] "Definitions in mathematics"
in Kennedy [1973: 235-246]
- Peirce C S
[1976] *New elements of mathematics*
ed C Eisele
The Hague, Mouton
- Phillips M
[1985] *Aspects of Text Structure*
Amsterdam, North-Holland
- Poincaré H
[1956] "Mathematical creation"
In Newman [1956: vol 4, 2041-52]
- Ponnamperuna C and Cameron A G W
[1974] *Interstellar Communication*
Boston, Houghton Mifflin
- Poincaré H
[1956] "Mathematical creation"
trans in Newman [1956: III 2041-52]
- Popper K R
[1947] "Logic without assumptions"
P A S n.s. 47: 251-292
- Priest G
[1979] "A note on the sorites paradox"
A J P 57: 74-5

Putnam H

- [1967a] "Mathematics without foundations"
in Benacerraf and Putnam [1983: 295-311]
- [1967b] "The thesis that mathematics is logic"
in Putnam [1975a: 12-42]
- [1972] *Philosophy of Logic*
London, George Allen and Unwin
- [1975a] *Mathematics, matter and method*
Cambridge, CUP
- [1975b] "What is mathematical truth?"
in Putnam [1975a: 60-78]
- [1979] "Philosophy of mathematics: a report"
in Asquith [1979: 386-398]

Quine W V

- [1936] "Truth by convention"
in Benacerraf and Putnam [1983: 329-354]
- [1948] "On what there is"
in Quine [1961: 1-19]
- [1951] "Two dogmas of empiricism"
in Quine [1961: 20-46]
- [1960] *Word and object*
Cambridge Mass., MIT Press
- [1961] *From a logical point of view*
New York, Harper and Row

Ramsey

- [1925] "The foundations of mathematics"
in Ramsey [1978] *Foundations*
ed D Mellor
London, RKP

Resnik M D

- [1980] *Frege and the philosophy of mathematics*
Ithaca, New York, Cornell UP
- [1981] "Mathematics as a science of patterns: Ontology"
Noûs 15: 529-50
- [1982] "Mathematics as a science of patterns: Epistemology"
Noûs 16: 95-105

Richards I

- [1942] *How to read a page*
Boston, Beacon Press

Robinson R

- [1954] *Definition*
Oxford, Clarendon Press

Routley R

- [1980] *Exploring Meinong's jungle*
Canberra, Research School of Social Sciences, A.N.U.
- [1982] "On what there is not"
Phil and Phen Res 43: 151-177

Rossi-Landi F

- [1983] *Language as Work and Trade*
South Hadley, Mass., Bergin and Garvey

Russell B

- [1905] "On denoting"
Mind 14: 479-493
- [1919] *Introduction to Mathematical Philosophy*
London, George Allen and Unwin

- [1923] "Vagueness"
Aust J Phil 1: 84-92
- [1937] *Principles of Mathematics*
London, George Allen and Unwin
- [1959] *My philosophical development*
London, George Allen and Unwin
- Rucker R
[1982] *Infinity and the mind*
Boston, Birkhäuser
- Sampson G
[1985] *Writing Systems*
London, Hutchinson
- Sanford D
[1975] "Borderline logic"
Amer Phil Q 12: 29-39
- Saussure F de
[1959] *Course in General Linguistics*
New York, McGraw-Hill
- Scheffler I
[1979] *Beyond the letter*
London, Routledge and Kegan Paul
- Schmandt-Bessarat D
[1978] "The earliest precursor of writing"
Sc Amer Jun 78: 38-47
- [1979] "Reckoning before writing"
Archaeology May/June 1979: 22-31
- Schmid C F
[1983] *Statistical Graphics*
New York, Wiley
- Schroeder M R
[1986] *Number Theory in Mathematics and Communication* 2nd ed
Berlin, Springer
- Searle J R
[1969] *Speech Acts*
Cambridge, CUP
- [1977] "Reiterating the differences"
Glyph 1: 198-208
- [1979a] *Expression and Meaning*
Cambridge, CUP
- [1979b] "A taxonomy of speech acts"
in [1979a: 1-29]
- Searle J R and Vanderveken D
[1985] *Foundations of Illocutionary Logic*
Cambridge, CUP
- Seidenberg A
[1962a] "Ritual origin of geometry"
Arch Hist Ex Sci 1: 488-527
- [1962b] "Ritual origin of arithmetic"
Arch Hist Ex Sci 2: 1-40
- Shapiro S

- [1983] "Mathematics and reality"
Phil Sci 50: 523-548
- Shuard H and Rothery A (eds)
[1984] *Children Reading Mathematics*
London, John Murray
- Shwayder D S
[1969] "Wittgenstein on Mathematics"
in Winch [1969: 66-116]
- Sless D
[1986] *In search of semiotics*
London, Croom Helm
- Sluga H
[1980] *Gottlob Frege*
London, RKP
- Sneed J D
[1979] *The Logical Structure of Mathematical Physics* 2nd ed
Dordrecht, Reidel
- Sohn-Rethel A
[1978] *Intellectual and Manual Labour*
London, Macmillan
- Staten H
[1985] *Wittgenstein and Derrida*
Oxford, Blackwell
- Steiner M
[1975] *Mathematical Knowledge*
Ithaca, NJ, Cornell UP
[1978] "Mathematical Explanation"
Phil Stud 34: 135-151
[1983a] "Mathematical Realism"
Noûs 17: 363-385
[1983b] "The Philosophy of Mathematics of Imre Lakatos"
J Phil 80: 502-521
- Stenlund S
[1974] "Analytic and Synthetic Arithmetical Statements"
in Stenlund S (ed) [1974] *Logical Theory and Semantic Analysis*
Dordrecht, Reidel
- Stubbs M
[1980] *Language and Literacy*
London, RKP
[1983] "Can I have that in writing please?
Some neglected topics in speech act theory"
J Pragmatics 7: 479-94
- Szabo A
[1978] *The Beginnings of Greek Mathematics*
Dordrecht, Reidel
- Tait W W
[1983] "Against intuitionism:
constructive mathematics is part of classical mathematics"
J Phil Log 12: 173-195
- Tannen D (ed)

- [1982] *Spoken and written language*
New York, Ablex
- Tarski A
[1931] "The concept of truth in formalised languages"
trans and reprinted in Tarski [1956: 152-277]
[1956] *Logic, Semantics, Metamathematics*
Oxford, OUP
[1964] "What is elementary geometry?"
in Hintikka [1969: 164-175]
- Teller P
[1980] "Computer proof"
J Phil 77:797-803
- Thiel C
[1968] *Sense and Reference in Frege's Logic*
Dordrecht, Reidel
- Thom R
[1973] "Modern mathematics: does it exist?"
in Howson [1973: 194-212]
- Thompson T
[1983] *From error-correcting codes through sphere packing to simple groups*
no place, Math Assoc Amer
- Tieszen R
[1984] "Mathematical Intuition"
Nous 18: 395-421
- Tragesser R
[1977] *Phenomenology and Logic*
Ithaca NY, Cornell UP
[1984] *Husserl and Realism in Logic and Mathematics*
Cambridge, CUP
- Troelstra A
[1969] *Principles of Intuitionism*
Berlin, Springer (LNM 95)
[1977] *Choice Sequences*
Oxford, OUP
- Tufte E R
[1983] *The Visual Display of Quantitative Information*
Cheshire Conn., Graphics Press
- Twyman M
[1979] "A schema for the study of graphic language"
in Kolers [1979: 117-150]
- Tymoczko T
[1979] "The four color theorem and its philosophical significance"
J Phil 76: 57-83
(ed) [1986] *New directions in philosophy of mathematics*
Boston, Birkhäuser
- Unger P
[1979] "There are no ordinary things"
Synthese 41: 117-154
- Vachek J
[1973] *Written language*

The Hague, Mouton

- Van Dantzig D
 [1956] "Is $10^{10^{10}}$ a finite number?"
Dialectica 10: 237-7
- Van Dijk TA
 [1977] *Text and Context*
 London, Longman 1977
- Van Dormolen
 [1986] "Textual analysis"
 in Christiansen [1986: 141-172]
- Van Heijenoort J
 [1967] *From Frege to Gödel*
 Cambridge Mass., Harvard UP
- Waismann F
 [1982] *Lectures on the Philosophy of Mathematics*
 ed Grassl H
 Amsterdam, Rodopi
- Wang H
 [1974] *From Mathematics to Philosophy*
 London RKP
- Webb J
 [1980] *Mechanism, Mentalism and Metamathematics*
 Dordrecht, Reidel
- Whitehead A and Russell B
 [1925-7] *Principia mathematica* (2nd ed)
 Cambridge, CUP
 [1962] *Principia mathematica to *56*
 Cambridge, CUP
- Wiggins D
 [1980] *Sameness and substance*
 Oxford, Blackwell
- Wilder R
 [1952] *Introduction to the foundations of mathematics*
 New York, Wiley
 [1973] *Evolution of mathematical concepts* (2nd ed)
 New York, Wiley
 [1981] *Mathematics as a cultural system*
 Oxford, Pergamon
- Winch P (ed)
 [1969] *Studies in the Philosophy of Wittgenstein*
 London, Routledge and Kegan Paul
- Wittgenstein L
 [1922] *Tractatus Logico-Philosophicus*
 London, Routledge and Kegan Paul
 [1935] "Philosophy for mathematicians"
 in Wittgenstein [1979: 205-225]
 [1953] *Philosophical Investigations*
 Oxford, Blackwell
 [1956] *Remarks on the foundations of mathematics*
 Oxford, Blackwell
 [1967] *Zettel*

- Berkeley, UC Press
- [1974] *Philosophical Grammar*
ed Rhees R, trans Kenny A
Oxford, Blackwell
- [1976] *Lectures on foundations of mathematics*
ed Diamond C
Hassocks England, Harvester
- [1978] *Remarks on the foundations of mathematics* (2nd ed)
Oxford, Blackwell
- [1979] *Lectures, Cambridge 1932-5*
ed. Ambrose A
Oxford, Blackwell
- Wright C
- [1975] "On the coherence of vague predicates"
Synthese 30: 325-365
- [1980] *Wittgenstein on the Foundations of Mathematics*
London, Duckworth
- [1982] "Strict Finitism",
Synthese 51: 203-282
- [1983] *Frege's Conception of Numbers as Objects*
Aberdeen, Aberdeen UP
- [1984] *Frege: Tradition and Influence*
Oxford, Blackwell
- Yates F A
- [1966] *The Art of Memory*
London RKP
- Zadeh L
- [1965] "Fuzzy sets"
Information and Control 8: 338-353
- [1975] "Fuzzy logic and approximate reasoning"
Synthese 30: 407-426
- Zaslavsky C
- [1973] *Africa Counts*
Boston Mass., Prindle Weber and Schmidt

Exhibits



- 1 A page from the financial records of a small business.
- 2 A page of Cajori [1929]
- 3A Half a page from Dummett [1977]
- 3B Half a page from Heyting [1980]
- 4 A page from Curwin and Slater [1986]
- 5 A page from Heath's Euclid
- 6A A histogram from Schroeder [1986]
- 6B A scatterplot from Australian Journal of Statistics
- 6C A plan from a work on architecture
- 6D A map of part of Adelaide
- 6E A list
- 6F A form
- 6G A chess diagram
- 6H A table about diagrams
- 7A Complete copy of Hansen [1984] {slightly photoreduced}
- 7B Samples of Begriffsschrift and Peano/Russell notation
- 9 A circuit diagram
- 10 Two pages from Griffiths [1976]
- 11 Two pages of Thompson [1983]
- 12A Definitions from Euclid book VII
- 12B Some definitions from some modern texts

Each of these exhibits is first discussed in the chapter indicated by its number; thus, exhibit #4 is first discussed in Chapter 4, exhibit #6D in Chapter 6 and so on. There is no exhibit #8 because no new item is discussed in Chapter 8.

asc wages 23/12/87 new format

| worker | hours | rate | gross | tax | net |
|--------|-------|------|--------|-------|--------|
| | | | 0 | | 0 |
| B | 14 | 8.5 | 119 | 0 | |
| | 0 | 13 | 119 | | 119 |
| C | 12.5 | 8.5 | 106.25 | 0 | 106.25 |
| E | 2.5 | 8.5 | 21.25 | 0 | 21.25 |
| F | 2.5 | 8.5 | 21.25 | 0 | 21.25 |
| G | 6 | 8.3 | 49.8 | 15 | 34.8 |
| H | 9 | 8.5 | 76.5 | 0 | 76.5 |
| I | 37.5 | 13 | 487.5 | 80.6 | 406.9 |
| J | 8.5 | 8.5 | 72.25 | 21.6 | 50.65 |
| K | 2.5 | 8.5 | 21.25 | 0 | 21.25 |
| | | | | | 0 |
| | | | | | 0 |
| total | 95 | | 1094.1 | 117.2 | 857.85 |

By the aid of a quadrant is determined the angular elevation *ACE*, "que erit altitudo tropici hiemalis," and the angular elevation *ACF*, "que erit altitudo tropici aestivalis," it being required to find the arc *EF* between the two. "Arcus itaque *EF*, fiet distantia duorum tropi-

$$\begin{aligned} &\sqrt{60} \dots \frac{2}{60}; \quad \sqrt{5} \dots \frac{2}{5}; \quad \sqrt{12} \dots \frac{2}{12}; \\ &\sqrt{20 \frac{4}{7}} \dots \frac{2}{7 \cdot 20}; \quad \sqrt{6} \dots \frac{2}{6}; \quad \sqrt{\frac{3}{5}} \dots \frac{2}{5}; \\ &3\sqrt{6} \dots \frac{2}{6}; \quad \sqrt{54} \dots \frac{2}{54}; \quad \frac{1}{2}\sqrt{48} \dots \frac{2}{48}; \quad \sqrt{12} \dots \frac{2}{12}. \end{aligned}$$

FORMULES D'ÉQUATIONS. TRINÔMES.

$$\begin{aligned} x^2 + 10x = 56 \dots \text{ش } 10 \text{ م } 1; \quad x^2 = 8x + 20 \dots \text{ش } 8 \text{ م } 20; \\ x^2 + 20 = 12x \dots \text{ش } 20 \text{ م } 1; \quad x^2 + 16 = 8x \dots \text{ش } 16 \text{ م } 2; \\ 6x^2 + 12x = 90 \dots \text{ش } 12 \text{ م } 6; \quad 4x^2 + 48 = 32x \dots \text{ش } 48 \text{ م } 32; \\ 3x^2 = 12x + 63 \dots \text{ش } 12 \text{ م } 63; \quad \frac{1}{2}x^2 + x = 7\frac{1}{2} \dots \text{ش } \frac{1}{2} \text{ م } 7\frac{1}{2}. \end{aligned}$$

PROPORTIONS.

$$\begin{aligned} 7 : 12 = 84 : x \quad \dots \dots \text{ش } 84 \text{ م } 12 \text{ م } 7; \\ 11 : 20 = 66 : x \quad \dots \dots \text{ش } 66 \text{ م } 20 \text{ م } 11. \end{aligned}$$

FIG. 35.—Al-Qalasâdî's algebraic symbols. (Compiled by F. Woepeke, *Journal asiatique* [Oct. and Nov., 1854], p. 363, 364, 366.)

corum quesita. Hâc Ptolemaeus reperit 47. graduum 42. minutorum 40. secundorum. Inuenit enim proportionem eius ad totum circulū sicut 11. ad 83, postea uerò minorem inuenerunt. Nos autem inuenimus arcum *AF* 65. graduum 6. minutorum, & arcum *AE* 18. graduum 10.

of a non-constructive proof:

Theorem There are solutions of $x^y = z$ with x and y irrational and z rational.

Proof. $\sqrt{2}$ is irrational, and $\sqrt{2}^{\sqrt{2}}$ is either rational or irrational. If it is rational, put $x = \sqrt{2}$, $y = \sqrt{2}$ so that $z = \sqrt{2}^{\sqrt{2}}$, which, by hypothesis, is rational. If, on the other hand, $\sqrt{2}^{\sqrt{2}}$ is irrational, put $x = \sqrt{2}^{\sqrt{2}}$ and $y = \sqrt{2}$, so that $z = (\sqrt{2}^{\sqrt{2}})^{\sqrt{2}} = (\sqrt{2})^2 = 2$, which is certainly rational.

Thus in either case a solution exists.

Exhibit #3A

2.2.6. Order relations between number-generators

I shall be brief about order relations.

Definition 1. $a < b$, if n and k can be found, so that $b_{n+p} - a_{n+p} > 1/k$ for every p . $a > b$ means the same as $b < a$.

Theorem 1. If $a \neq b$, either $a < b$ or $b < a$.

Proof. Find n and k so that $|a_{n+p} - b_{n+p}| > 1/k$ for every p . Now determine $m > n$ so that $|a_m - a_{m+p}| < 1/4k$ and $|b_m - b_{m+p}| < 1/4k$ for every p . Either $a_m - b_m > 1/k$ or $b_m - a_m > 1/k$; in the first case $a_{m+p} - b_{m+p} > 1/2k$ for every p , so $b < a$; in the second case we obtain $a < b$.

Exhibit #3B

Theorem 2. If $a < b$, then $a \neq b$.

This follows immediately from the definitions.

Theorems 1 and 2 show, that $a \neq b$ is equivalent to $(a < b \text{ or } a > b)$.

Theorem 3. If $a < b$ as well as $b < a$ are contradictory, then $a = b$.

Proof. We derive a contradiction from the supposition $a \neq b$ by Th. 1; then 2.2.3, Th. 3 gives $a = b$.

Since the standard normal distribution is symmetrical about its mean of 0, an area to the right of a positive value of Z will be identical to the area to the left of the corresponding negative value of Z . (Note that areas cannot be negative.) Thus to find the area between $Z = -1$ and $Z = +1$, we have:

| | | |
|-----------------------------|--------|-----|
| area to left of $Z = -1$: | 0.1587 | |
| area to right of $Z = +1$: | 0.1587 | add |
| | 0.3174 | |

Therefore,

area between $Z = -1$ and $Z = +1$ is $1 - 0.3174 = 0.6826$

For **any** normal distribution, 68.26% of the values will be within one standard deviation of the mean. (Hint: it is often useful to draw a sketch of the area required by a problem and compare this with Fig. 11.2.)

P1

Exercise

What percentage of values will be within 1.96 standard deviations of the mean? (Answer: 95%.)

P2

Example

If a population is known to have a normal distribution, and its mean and variance are known, then we may use the tables to express facts about this population.

The attendance at rock concerts at a particular stadium has a normal distribution with a mean of 15 000 and a variance of 4 000 000. The promoters are able to break even at an attendance of 12 500; what proportion of concerts will make a loss?

Mean = 15 000, variance = 4 000 000:
standard deviation = $\sqrt{\text{variance}} = 2000$

Therefore, $Z = \frac{12500 - 15000}{2000} = -1.25$

(Note that we have used two scales in Fig. 11.3, one for Z and one for the original distribution; this may help in understanding some questions.)

From tables, area to right of $Z = +1.25$ is equal to area to left of $Z = -1.25$: 0.1056

Thus, 10.56% of the concerts will make a loss.

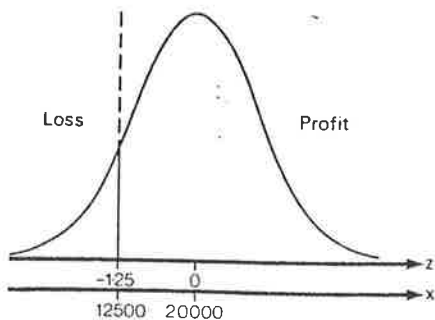


Figure 11.3

P3

P4

} W1

W2
W3

N1 N2
N3
N4

W4

} W5
W6

}
D

1. 46, 47]

PROPOSITIONS 46, 47

349

Therefore AG, CF are parallel.

Also, since each of the alternate angles AFG, FGC is half a right angle,
 AF, CG are parallel.

Hence $AFCG$ is a parallelogram; and AF, CG are equal.

Thus the triangles ABF, CBG have two angles and one side respectively equal;

therefore AB is equal to BC , and BF to BG .

PROPOSITION 47.

In right-angled triangles the square on the side subtending the right angle is equal to the squares on the sides containing the right angle.

Let ABC be a right-angled triangle having the angle
 5 BAC right;

I say that the square on BC is equal to the squares on
 BA, AC .

For let there be described
 on BC the square $BDEC$,
 10 and on BA, AC the squares
 GB, HC ; [1. 46]

through A let AL be drawn
 parallel to either BD or CE ,
 and let AD, FC be joined.

15 Then, since each of the
 angles BAC, BAG is right,
 it follows that with a straight
 line BA , and at the point A
 on it, the two straight lines
 20 AC, AG not lying on the
 same side make the adjacent
 angles equal to two right
 angles;

therefore CA is in a straight line with AG . [1. 14]

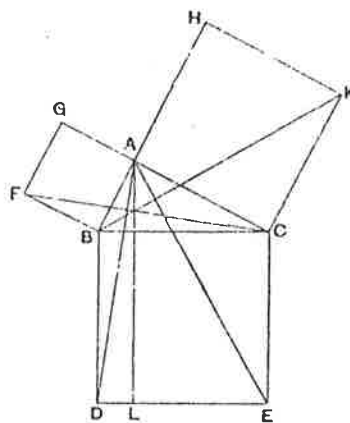
25 For the same reason

BA is also in a straight line with AH .

And, since the angle DBC is equal to the angle FBA : for
 each is right:

let the angle ABC be added to each;

30 therefore the whole angle DBA is equal to the whole
 angle FBC . [C. N. 2]



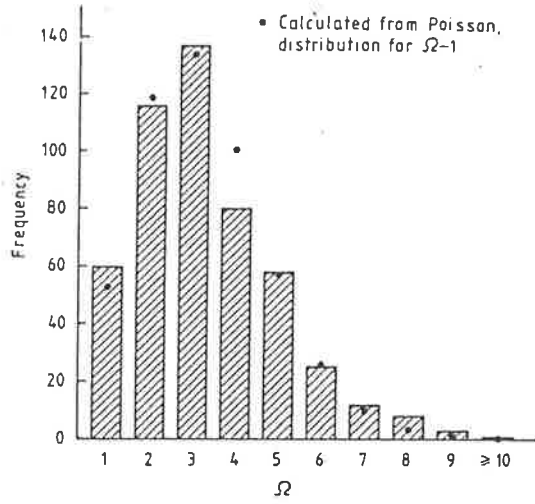


Fig. 11.2. The distribution of the number of prime factors (*bars*) in the interval $9000 \leq n \leq 9499$ and the Poisson distribution (*dots*) for the theoretical mean

Exhibit #6A

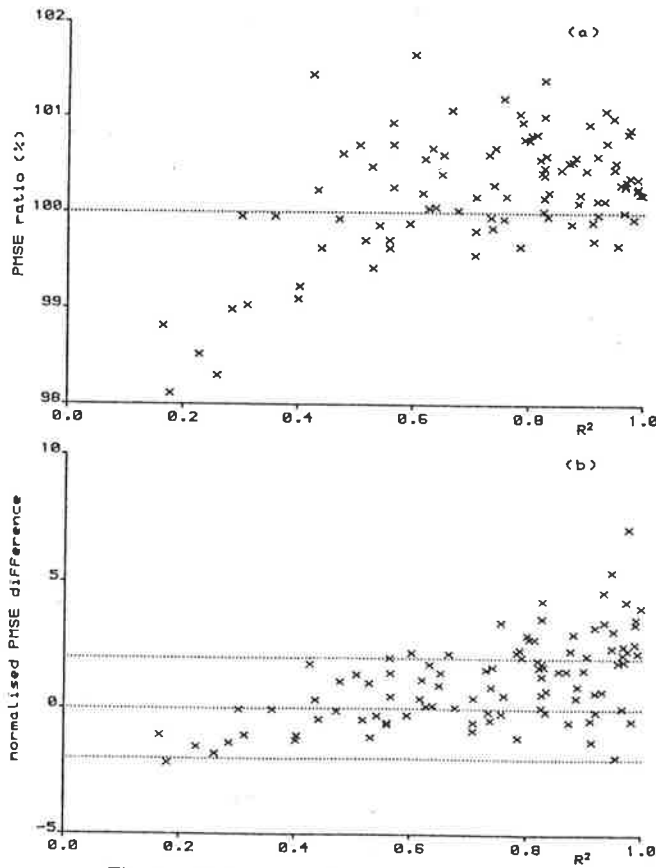
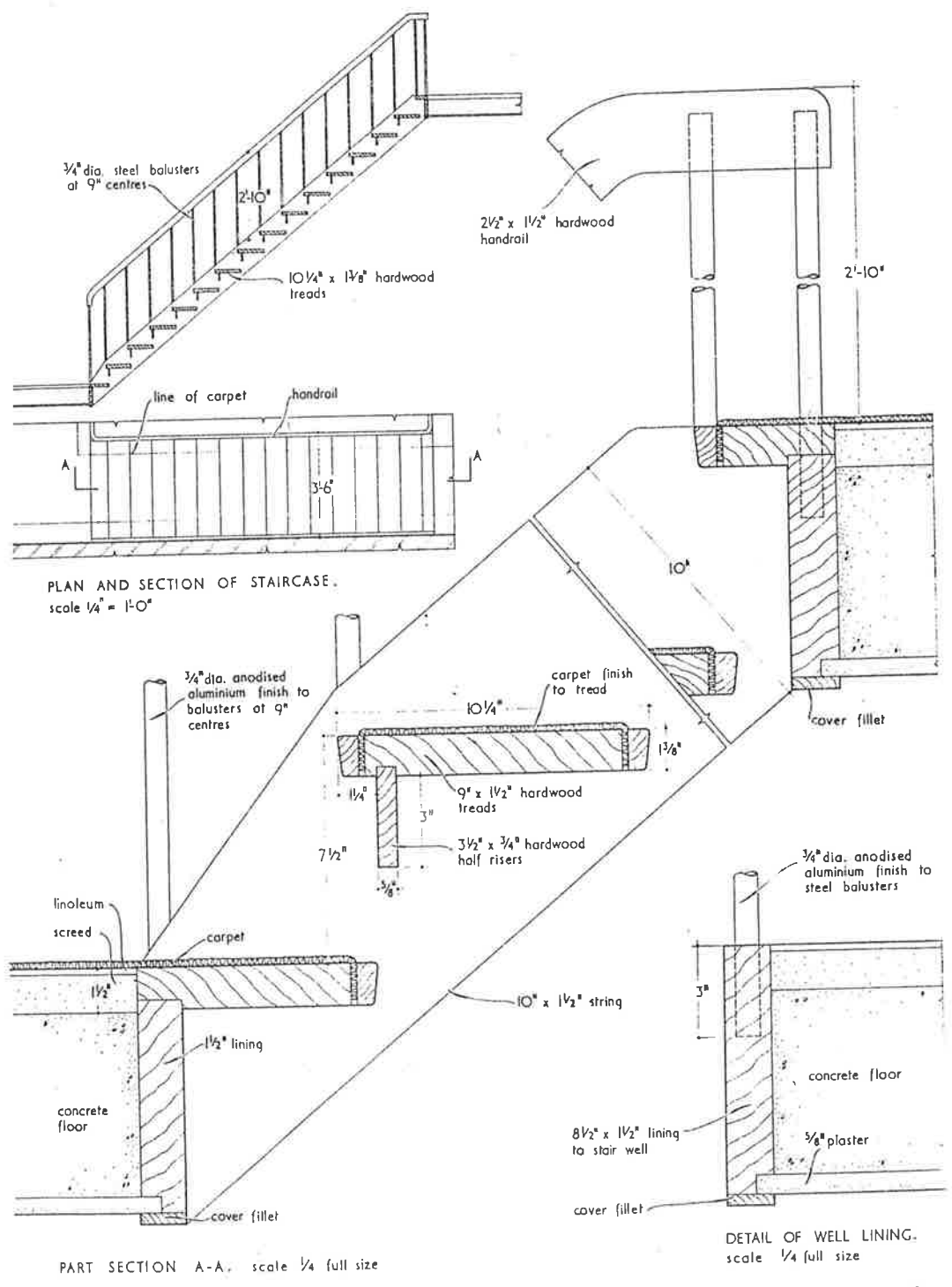


Fig. 2.—As for Figure 1, except that $p = 5$.

Exhibit #6E

STAIRCASES



PLAN AND SECTION OF STAIRCASE.
scale $\frac{1}{4}" = 1'-0"$

PART SECTION A-A, scale $\frac{1}{4}$ full size

DETAIL OF WELL LINING.
scale $\frac{1}{4}$ full size

MAP 69

Australian Map and 1:100,000 scale information shown on the
 Geographical Map of Adelaide parts of 1:100,000 series maps 1:20,000 1:40,000 1:60,000

JOINS MAP 59

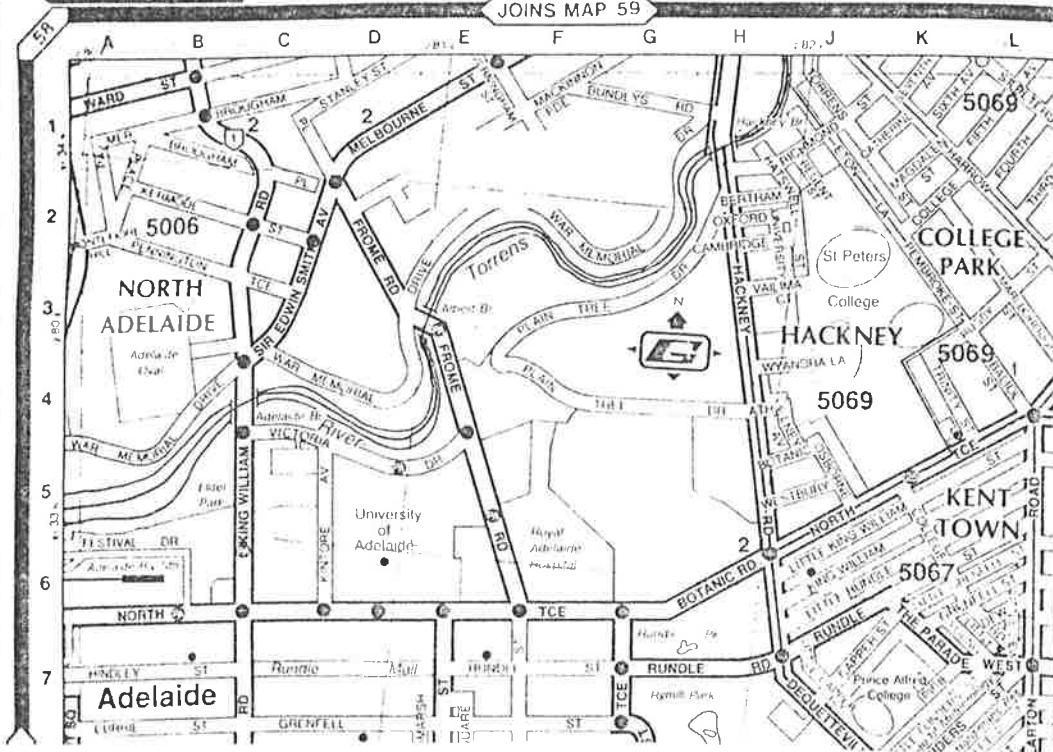


Exhibit #6D

5.6 Fibonacci, Rabbits and Computers

Fibonacci numbers abound in nature. They govern the number of leaves, petals and seed grains of many plants (see Fig. 5.7 [5.7,8]), and among the bees the number of ancestors of a drone n generations back equals F_{n+1} (Fig. 5.8).

Rabbits, not to be outdone, also multiply in Fibonacci rhythm if the rules are right: offspring beget offspring every "season" except the first after birth — and they never die (Fig. 5.9). As already mentioned, this was the original Fibonacci problem [5.9] considered in 1202 by Fibonacci himself.

Leonardo da Pisa, as Fibonacci was also known, was a lone star of the first magnitude in the dark mathematical sky of the Middle Ages. He travelled widely in Arabia and, through his book *Liber Abaci*, brought the Hindu-



- Buttercups : 5 petals
- Lilies and irises : 3 petals
- Some delphiniums : 8 petals
- Corn marigolds : 13 petals
- Some asters : 21 petals
- Daisies : 34, 55 and 89 petals

Fig. 5.7. Flowers have petals equal to Fibonacci numbers

Exhibit #6E

QUANTITATIVE METHODS 1987

STUDENT INFORMATION COLLECTION

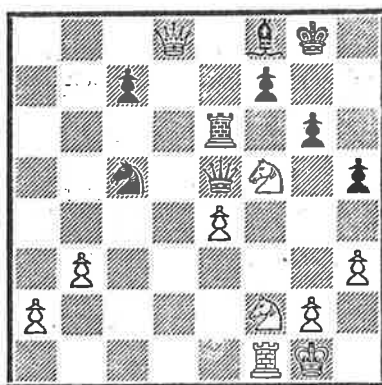
| | | |
|-------------------|-------------|---------------------------|
| NAME _____ | (surname) | _____ (personal name/s) |
|-------------------|-------------|---------------------------|

| | | |
|--|--|---|
| COURSE BBUS <input type="checkbox"/> BA <input type="checkbox"/> BED <input type="checkbox"/> OTHER <input type="checkbox"/> (specify...) | part-time <input type="checkbox"/> full-time <input type="checkbox"/> <hr/> internal <input type="checkbox"/> external <input type="checkbox"/> | ID _____ PHONE _____ ADDRESS _____ _____ |
|--|--|---|

| | | | |
|---|--|----------------------------------|--|
| MATHEMATICS BACKGROUND please tick one | | | |
| tertiary <input type="checkbox"/> | Matric maths 1, 2 <input type="checkbox"/> | year 11 <input type="checkbox"/> | |
| TAFE <input type="checkbox"/> | Matric maths 1S <input type="checkbox"/> | less <input type="checkbox"/> | |

| |
|---|
| <p>TUTORIAL TIMES</p> <p>You must attend one tutorial each week.</p> <p>To allocate you to a tutorial I need to know what times you prefer. Go to the notice board outside A67 and write your name and 1 on the sheet for the time you like best, and your name and 2 for your second preference. I will post a list of allocations by Monday 3aug87.</p> |
|---|

| |
|---|
| <p>DIAGNOSTIC QUESTIONS (show working)</p> <p>Spend no more than 5 minutes on these problems, using no books, and put your form into the box outside A67.</p> <p>1 What is $(-3)^{-3}$?</p> <p>2 Solve $7a + 25 = 2a$.</p> <p>3 Simplify $2*(3 - 6m) - (2+ m)$.</p> |
|---|



analysis of some diagrams

| | Form elements | | | | | | Structuration | | | | | Function elements | | | Function | | | | |
|-----------------------|---------------|------|-------|---|------|---|---------------|-----------|--------|---|-----|-------------------|----------|---------|---------------|---|----------|---|---|
| | Points | | Areas | | Text | | Juxt | | Layout | | Mea | PLot | Frans | Case | Specn | | Resource | | |
| | Lines | Char | Pix | | | | Align | Incidence | | | | Label | | Illustn | Display | | | | |
| CS174 | x | v | v | v | v | x | v | v | v | v | v | v | v | x | v | x | v | v | |
| Euclid 1.47 | v | v | v | v | x | x | v | v | v | v | v | v | x | v | v | x | v | v | |
| Schroeder histo p146 | x | v | v | v | v | x | v | v | v | x | v | v | v | x | v | x | v | v | |
| Scattergram | v | v | x | v | v | x | v | v | v | v | ? | v | v | v | x | v | x | v | v |
| London Underground | x | x | v | x | v | v | v | v | v | v | v | v | x | x | v | v | v | v | |
| Plan of staircase | x | v | v | v | v | v | v | v | v | v | v | v | v | x | x | v | v | v | |
| map of Adelaide | x | v | v | v | v | v | v | v | v | x | x | v | v | x | x | v | v | v | |
| list of petal numbers | x | x | x | x | v | x | v | v | v | x | x | v | x | v | x | x | v | v | |
| table of N(0,1) | x | v | x | v | v | x | v | v | v | x | x | v | v | x | x | x | v | v | |
| Stinform 87 | x | v | v | v | v | x | v | v | v | x | x | v | v | x | x | v | v | v | |
| chess diagram | x | v | v | x | x | v | v | v | v | x | x | v | x | v | v | x | x | v | v |
| this table | x | v | x | v | v | v | v | v | v | ? | x | v | v | v | v | v | v | v | |
| common | | | | | | | V | V | V | | | V | # | # | * | * | * | V | V |
| | | | | | | | | | | | | | # L or F | | * C or I or S | | | | |

Hansen, D. J. (1984). Positive derivations on partially ordered strongly regular rings. *Journal of the Australian Mathematical Society*, 37(2), 178-180.

NOTE:

This publication is included in the print copy
of the thesis held in the University of Adelaide Library.

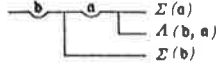
It is also available online to authorised users at:

<http://dx.doi.org/10.1017/S1446788700022011>

Then



or



is the circumstance that each child of a human being is also a human being, or that the property being a human being is hereditary.

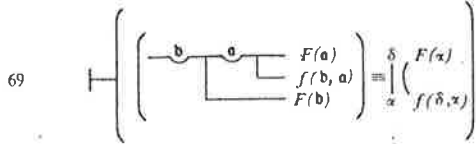
[Although this example was easy to put into words,] we can see, however, that it can become difficult and even impossible to give a rendering in words if very complicated functions occur in the places of F and f . Sentence (69) could be expressed in words this way:

"If from the proposition that b has the property F , whatever b may be, it can always be inferred that each result of an application of the procedure f to b has the property F ,

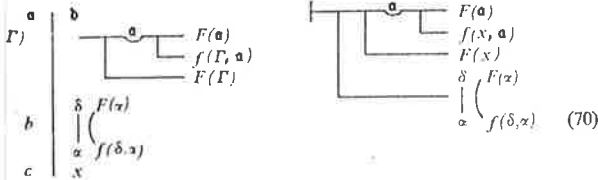
then I say:

'The property F is hereditary in the f -sequence'."

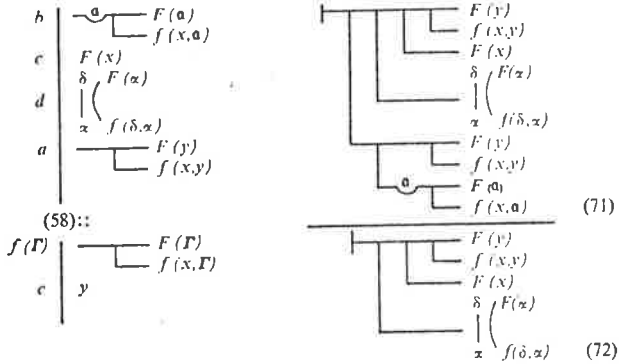
§25.



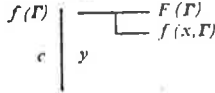
(68):



(19):



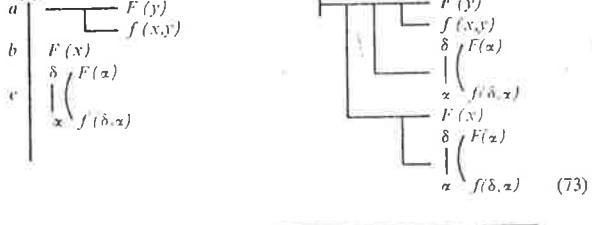
(58):



If the property F is hereditary in the f -sequence; and if x has the property F , and y is a result of an application of the procedure f to x : then y has the property F .

72

(2):



*90. ON THE ANCESTRAL RELATION

Summary of *90.

If R is any relation, " xR_*y " is to mean " x is an ancestor of y with respect to R ," where a term counts as its own ancestor provided it belongs to the field of R . The definition of R_* is as follows:

90-01. $R_ = \hat{x}\hat{y} \{x \in C^R R; \check{R}''\mu \subset \mu \cdot x \in \mu \cdot \exists \mu \cdot y \in \mu\}$ Df

That is, xR_*y is to hold when x belongs to the field of R , and y belongs to every hereditary class to which x belongs; a hereditary class being a class μ such that $\check{R}''\mu \subset \mu$, i.e. such that all successors of μ 's are μ 's.

90-02. $\check{R}_ = \text{Cnv}'R_*$ Df

This definition serves merely to decide the ambiguity between $(\check{R})_*$ and $\text{Cnv}'R_*$, either of which might be meant of \check{R}_* . It will be shown, however, that the two are equal (*90-132).

The most important propositions of this number are the following:

*90-112. $\vdash : xR_*y; \phi z \cdot zRw \cdot \supset z \cdot \phi w \cdot \phi x : \supset \cdot \phi y$

I.e. if xR_*y and if ϕz is a hereditary property belonging to x , then it belongs to y .

*90-12. $\vdash : x \in C^R R \equiv xR_*x$

I.e. R_* is reflexive throughout the field of R , but not elsewhere.

90-14. $\vdash : D^R R_ = C^R R_* = C^R R_* = C^R R$

90-15. $\vdash : I \uparrow C^R R \subset R_$

90-151. $\vdash : R \subset R_$

90-16. $\vdash : R_ \uparrow R \subset R_*$

90-163. $\vdash : \check{R}''\check{R}_^*x \subset \check{R}_*^*x$

I.e. \check{R}_*^*x is a hereditary class.

90-17. $\vdash : R_^* = R_*$

90-21. $\vdash : a \subset C^R R \equiv a \subset \check{R}_^*a \equiv a \subset R_*^*a$

90-22. $\vdash : \check{R}''a \subset a \equiv \check{R}_^*a \subset a$

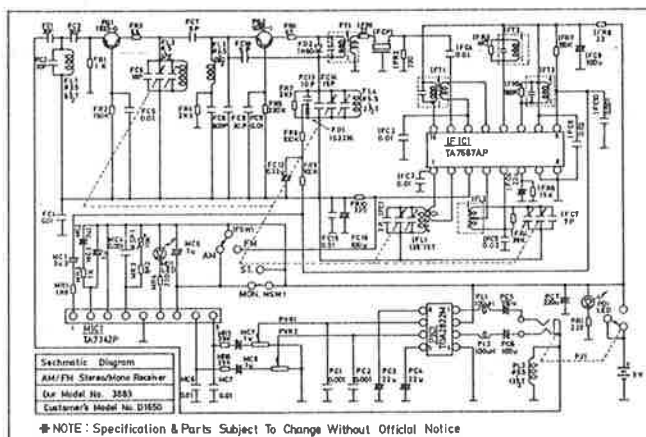
I.e. the classes that are hereditary with respect to R are the same as those that are hereditary with respect to R_* .

90-31. $\vdash : R_ = I \uparrow C^R R \cup R_* \cdot R$

90-32. $\vdash : R \cdot R_ = R \cup R' R_* \cdot R = R_* \cdot R$

90-33. $\vdash : R_^*a = (a \cap C^R R) \cup R_*^*R''a = (a \cap C^R R) \cup R''R_*^*a$

90-4. $\vdash : (R_)_* = R_*$



1. How to make surfaces and talk about them

1.1 WHAT IS A SURFACE?

What do *you* mean by a \textcircled{a} surface? In real life we are surrounded by surfaces – those of furniture, tools, utensils, buildings, fluids, our bodies – and yet most people have surprising difficulty in being able to say what they mean by a surface in general. They know a particular surface when they see one, but how can we tell a computer or a blind man what it is about surfaces that they all have in common? Mathematicians began to be faced with this problem in the late nineteenth century as mathematics grew more complicated, and it took them many years to find a way of saying what they meant by a surface, that all mathematicians would understand. They had to be able to agree on what they were talking about before they could begin to *prove* things about surfaces, to do mathematics about surfaces. Their agreed statement about what a surface is, is called a *definition*; but it is not easy for beginners, so we shall approach the question in a different way. We shall eventually *make* a definition of our own, but in ordinary language that does not look mathematical. One can do mathematics in many dialects of a ‘professional’ language, and a mathematician chooses a particular dialect to suit his immediate purposes. Indeed, the following discussion is designed to show how mathematicians look at things they wish to study, and decide on the right words to use, in order to make their study easier.

Let us therefore ask a different question. How would you *make* a surface? (This might then help you to say what you mean by a surface: anything assembled according to your instructions would be a surface, although perhaps some exceptional ‘surfaces’ would not be made that way.) Now, most people only make surfaces as the ‘skin’ on some solid, by baking dough, moulding clay, or assembling wood or concrete forms. It is hard to say what we mean by a ‘solid’ and its ‘skin’ (has a jelly a surface; is it a solid?). But a seamstress makes a sort of surface when she stitches together pieces of cloth to make a dress, and an engineer makes a surface when he joins metal sheets to make the hull of a ship, the fuselage of a plane, or the body of a car. In all these cases, certain simple bits of surface – pieces of cloth, panels of metal – are being joined to make more complicated ones. We may not have the skill or tools of a seamstress or an engineer, but instead of pieces of cloth or metal we can use sheets of paper cut into polygons. These polygonal panels can be joined with sticky tape along edges, instead of being stitched or welded, to form more complicated things that most people will agree are surfaces. Some people might say that these paper surfaces are rather special, for various reasons, but let us consider

Appendix A. Mathematical theory of surfaces

The purpose of this appendix is to outline a rigorous mathematical foundation for the theory expounded in the text. It is assumed here that the reader is acquainted with the elements of topology (such as in Griffiths and Hilton [7], ch. 25) and can read some of the mathematical texts which we cite. To save space, proofs of results are omitted, and instead we give sources where they may be found.

A.1 NOTATION

We use \mathbb{R} , \mathbb{N} respectively to denote the sets of real, and natural, numbers. In the Euclidean plane \mathbb{R}^2 , we let S^1 and D denote respectively the unit circle $x^2 + y^2 = 1$ and the unit disc $x^2 + y^2 \leq 1$. Each is compact. If $n \in \mathbb{N}$ and $n \geq 3$, we let D_n denote the unit disc together with its *vertices*, the n points with angular coordinates $2\pi k/n$ ($k = 0, 1, \dots, n-1$). A *Jordan curve* in a metric space X is a mapping $f: S^1 \rightarrow X$ that is a homeomorphism onto its image $f(S^1)$. (It suffices to say that f is a continuous injection, since $f(S^1)$ is compact.) We do not distinguish here between f and $f(S^1)$.

A.2 JORDAN AND SCHOENFLIESS

Any rigorous theory requires statements and proofs of the Jordan Curve Theorem (‘Every Jordan curve separates the plane \mathbb{R}^2 into exactly two components, of which it is the common frontier’) and the Schoenflies Theorem (‘Given a Jordan curve in \mathbb{R}^2 , there is a homeomorphism $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ which carries the curve onto the unit circle, and hence its interior onto the unit disc’). Proofs are in Newman [14], pp. 137 and 173. For an illuminating elementary proof of the polygonal Jordan Curve Theorem, see Courant and Robbins [5], ch. 5. Note that the Moebius band cannot be planar, because of the Jordan Curve Theorem; for, its centre-line (a Jordan curve) does not separate it (see Section 1.5).

A.3 MATHEMATICAL SURFACES

Mathematical models \textcircled{a} of the panels and paper surfaces of the text can now be made as follows. A ‘paper panel in our 3-dimensional world’ can be modelled by a continuous injection $p: D_n \rightarrow \mathbb{R}^3$ (for some $n \in \mathbb{N}$, $n \geq 3$). Reference to \mathbb{R}^3 can be eliminated, if desired, by taking ‘a paper panel P ’ to correspond to a homeomorphism $p: D_n \rightarrow P$, where P denotes a topological space. The *edges* of the panel then correspond to the restrictions $p|_{e_i}$, where e_i is the i th edge of D_n . The *vertices* are the images of those of

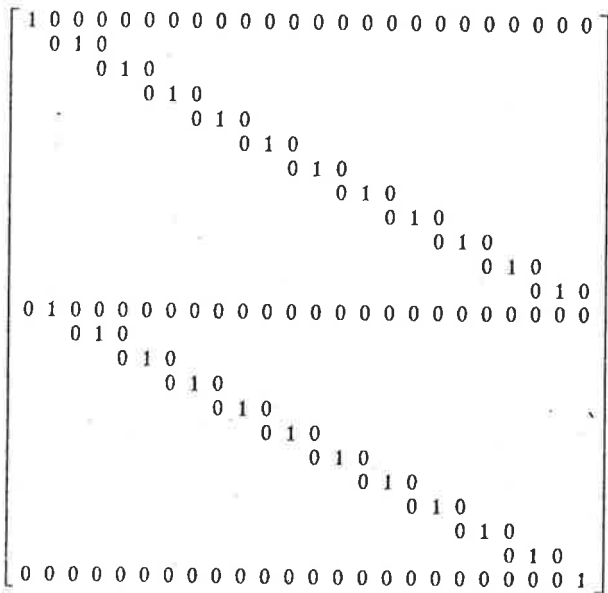


FIG. 3.13. The matrix of order 24 representing β .

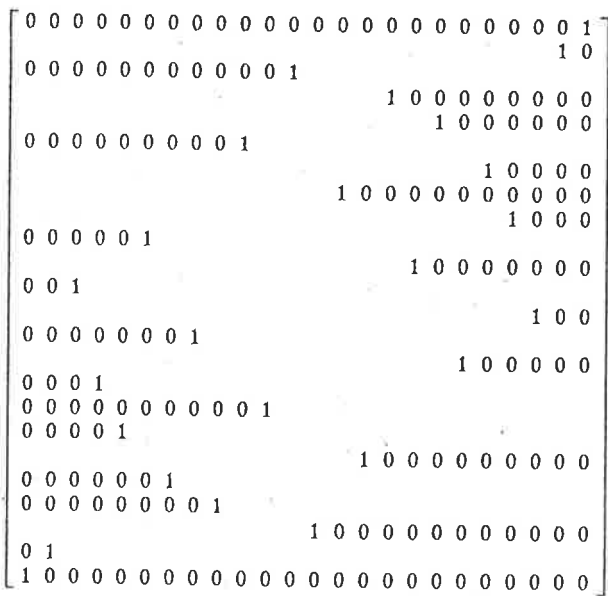


FIG. 3.14. The matrix of order 24 representing γ .

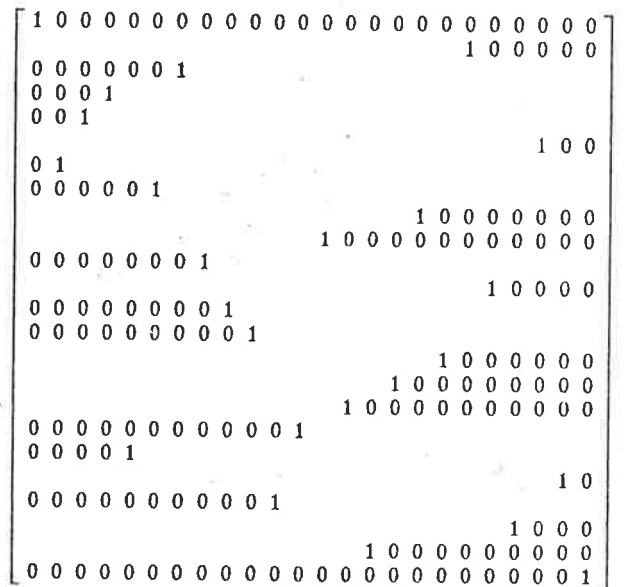


FIG. 3.15. The matrix of order 24 representing δ .

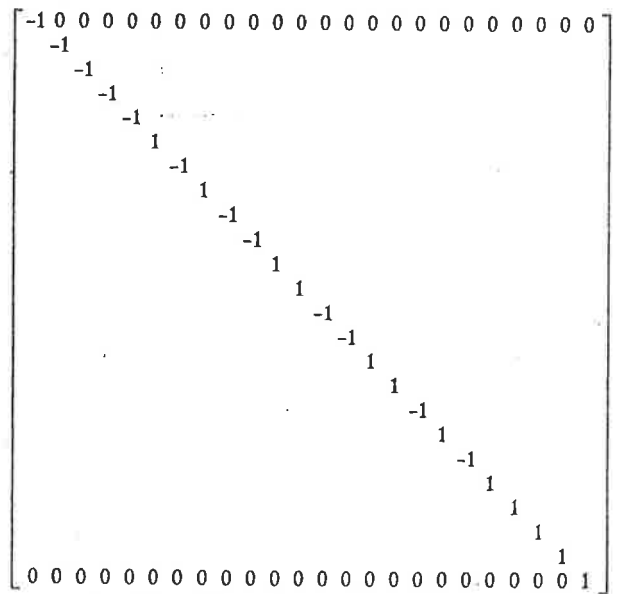


FIG. 3.16. The matrix of order 24 representing ϵ .

BOOK VII.

DEFINITIONS.

1. An **unit** is that by virtue of which each of the things that exist is called one.
2. A **number** is a multitude composed of units.
3. A number is a **part** of a number, the less of the greater, when it measures the greater ;
4. but **parts** when it does not measure it.
5. The greater number is a **multiple** of the less when it is measured by the less.
6. An **even number** is that which is divisible into two equal parts.
7. An **odd number** is that which is not divisible into two equal parts, or that which differs by an unit from an even number.

BOOK I.

DEFINITIONS.

1. A **point** is that which has no part.
2. A **line** is breadthless length.
3. The extremities of a line are points.
4. A **straight line** is a line which lies evenly with the points on itself.
5. A **surface** is that which has length and breadth only.
6. The extremities of a surface are lines.
7. A **plane surface** is a surface which lies evenly with the straight lines on itself.
8. A **plane angle** is the inclination to one another of two lines in a plane which meet one another and do not lie in a straight line.
9. And when the lines containing the angle are straight, the angle is called **rectilinear**.
10. When a straight line set up on a straight line makes the adjacent angles equal to one another, each of the equal angles is **right**, and the straight line standing on the other is called a **perpendicular** to that on which it stands.
11. An **obtuse angle** is an angle greater than a right angle.
12. An **acute angle** is an angle less than a right angle.
13. A **boundary** is that which is an extremity of anything.
14. A **figure** is that which is contained by any boundary or boundaries.
15. A **circle** is a plane figure contained by one line such that all the straight lines falling upon it from one point among those lying within the figure are equal to one another ;

11. A **prime number** is that which is measured by an unit alone.
12. Numbers **prime to one another** are those which are measured by an unit alone as a common measure.
13. A **composite number** is that which is measured by some number.

Chapter I

The Complex Number System

§1. The real numbers

We denote the set of all real numbers by \mathbb{R} . It is assumed that each reader is acquainted with the real number system and all its properties. In particular we assume a knowledge of the ordering of \mathbb{R} , the definitions and properties of the supremum and infimum (sup and inf), and the completeness of \mathbb{R} (every set in \mathbb{R} which is bounded above has a supremum). It is also assumed that every reader is familiar with sequential convergence in \mathbb{R} and with infinite series. Finally, no one should undertake a study of Complex Variables unless he has a thorough grounding in functions of one real variable. Although it has been traditional to study functions of several real variables before studying analytic function theory, this is not an essential prerequisite for this book. There will not be any occasion when the deep results of this area are needed.

§2. The field of complex numbers

We define \mathbb{C} , the complex numbers, to be the set of all ordered pairs (a, b) where a and b are real numbers and where addition and multiplication are defined by:

$$(a, b) + (c, d) = (a + c, b + d)$$

$$(a, b)(c, d) = (ac - bd, bc + ad)$$

It is easily checked that with these definitions \mathbb{C} satisfies all the axioms for a field. That is, \mathbb{C} satisfies the associative, commutative and distributive

that the y 's are linearly independent. In other words, after the argument has been applied m times, we obtain a set with the same property the x 's had, and this set differs from the set of x 's in that m of them are replaced by y 's. This seemingly innocent statement is what we are after; it implies that $n \geq m$. Consequently if both \mathfrak{X} and \mathfrak{Y} are bases (so that they each have both properties), then $n \geq m$ and $m \geq n$.

DEFINITION. The *dimension* of a finite-dimensional vector space \mathfrak{U} is the number of elements in a basis of \mathfrak{U} .

92 Topology

general topological space. We also construct some machinery which will be useful in the detailed study of these spaces.

Our main special interest in the four chapters that follow will be in continuous real or complex functions defined on particular types of topological spaces, and we shall develop the point of view that there is a constant illuminating interplay between the structure of these spaces and the properties of the continuous functions which they carry.

16. THE DEFINITION AND SOME EXAMPLES

Let X be a non-empty set. A class \mathfrak{T} of subsets of X is called a *topology* on X if it satisfies the following two conditions:

- (1) the union of every class of sets in \mathfrak{T} is a set in \mathfrak{T} ;
- (2) the intersection of every finite class of sets in \mathfrak{T} is a set in \mathfrak{T} .