Re-counting Plato

A Computer Analysis of
Plato's Style

Gerard R. Ledger

To Elisabeth

Why write I still all one, ever the same,
And keep invention in a noted weed,
That every word doth almost tell my name,
Showing their birth, and where they did proceed?

Shakespeare, Sonnet LXXVI

Preface

THIS book originated as a thesis presented at Reading University in 1987. The core of the work, the essential statistical analysis of the Platonic corpus, remains unchanged apart from corrections of detail which were necessary in the sections dealing with the chronology of the dialogues. Apart from this, the outlying chapters were rewritten. Chapter 10 especially, which was scarcely adequate in the original, owes much to a suggestion of Dr Wojtek J. Krzanowski, and I hope it gives a better account than the previous version offered of why the methods employed do indeed work.

My debts of gratitude are many, but mostly I owe thanks to my two supervisors, Dr Tessa Rajak and Dr Anthony J. Woods. Dr Rajak always showed interest and critical enthusiasm for a project which, in all fairness, cannot claim to be other than on the fringes of the classical world. Dr Woods encouraged me to look at the more sophisticated techniques of statistical analysis and opened my eyes to their potential.

Many in the various departments at Reading—Classics, Applied Statistics, and the Computer Centre—gave freely and generously their advice and assistance, and I ask them to accept this general acknowledgement as thanks for their involvement.

Outside the bounds of my immediate contacts at Reading I would like to thank especially Dr Anthony Kenny for the inspiration of his own work in this field, and for his interest and his help in enabling me to find a publisher.

Mrs Sybil Lowery typed cheerfully and tolerantly a difficult manuscript and made the final stages of preparation far less difficult than they might otherwise have been. I would also like to thank Professor Andrew Wallace-Hadrill, Head of Classics at Reading, for allowing me space and time to complete the book.

Finally to my wife Elisabeth, who has given me support and encouragement through eight years of continuous study, more thanks are due than I can adequately express. To her the book is dedicated.

G.R.L.

Contents

List of Figures

List of Tables

Abbreviations

The following abbreviations used throughout the books will be unfamiliar to most readers:

MVA   Multivariate (statistical) analysis
OCP   Oxford Concordance Program: a textual concordance package produced by the Oxford University Computing Service
SAS   Statistical Analysis Systems: trade name of a computer statistical package
TLG   Thesaurus Linguae Graecae: a library of machine-readable Greek texts produced at the University of California, Irvine, CA 92717.

Abbreviations for classical authors and works follow LSJ, except for the following:

Ant.   Isocrates, *Antidosis*
Archid.   Isocrates, *Archidamus*
Era.   Lysias, *Against Eratosthenes*
Ora.   Isaeus' orations
Pac.   Isocrates, *De Pace*
Pan.   Isocrates, *Panegyricus*
Panth.   Isocrates, *Panathenaicus*
Tma.   Aeschines, *Against Timarchus*
Xen.   Xenophon.

1
Introduction

THE aim of stylometry is the quantification of style. There is little agreement, however, as to how this objective might be achieved. Over the last 20 years, with the advent of large computers, more exhaustive research has become possible, research which has shown that detectable differences in the minutiae of language are present in the works of various authors. But whether or not these differences are adequate to define one author's style in a way which distinguishes it from all others in all circumstances is not yet known.

If one were restricted to the mention of a single weak point which undermines so many enquiries in this field, it must surely be the fact that at the level of word frequency so much variation is possible that reliable predictions cannot be made. All words, including the most common, occur at such low levels of frequency that in order to form reliable judgements about an author's performance with regard to vocabulary usage extremely large sample sizes would be required. Thus the word 'but' has occurred in this Introduction once only (in 170 words approximately), the word 'frequency' twice. This does not indicate that as an author I am twice as prone to use the latter word rather than the former. It may imply something in a general way about my interests. Alternatively—and this is more likely—it is a fortuitous result stemming from the nature of the subject and the smallness of the sample.

Statistics, however, is not really concerned with such irregularities. It is only because the world as we perceive it presents us with phenomena which have an underlying pattern of regularity that the science of statistics can operate, often with extremely powerful effect. Paroxysms and the unpredictable it is less able to deal with, although even in these areas of seemingly impenetrable chaos mathematics has some success in coping with and predicting the range of possible disturbance. But statistics is more at ease with orderly and regulated phenomena. It does not anticipate a world without variation, but it knows or finds out how great this variation will be and on such findings bases its predictions.

I make these simple points because it is not uncommon for students to be daunted by the machinery of statistics. The various tests and distributions are fearful concepts to many and may be awkward to work with. But they are only the means by which a great mass of discrete events may be shown to be reducible to an underlying pattern. The task of statistics is to clarify, not to

obfuscate. This it does by showing that amidst the chaos of individual measurements an abiding order prevails.

Yet language is an excellent example of a constrained, regulated, and ordered system. No one can break its rules without risk of becoming unintelligible, of failing to communicate, of ceasing to use language at all. Any 10 or 20 words strung together at random do not constitute language. What matters is that the sequence has meaning, and this is usually only achieved if the very complex system of checks and balances which conventional usage dictates is observed. If we accept that language is a regulated system (few would deny it), it is inconceivable that statistics could have nothing to say about it. Stylometry must have a basis in reality because of the orderly nature of the material with which it is dealing, not because authors might differ in their particular usage of selected words, or are idiosyncratic in other ways, although it is very probable that this is the case. What matters more is that there is inherently, in the very structure of language, an orderly matrix, a network of predictability.

This book attempts to show that statistics can work extraordinarily well in these uncharted fields. I have analysed the works of various ancient Greek authors, especially those of Plato, using techniques which have been developed for use in other areas of scientific enquiry. They are largely techniques which are concerned with species differentiation. The underlying assumption is that it is legitimate to regard works from different authors, or from different periods of an author's career, as different species of language. The statistical analysis then attempts to show that this hypothesis may be sustained.

I have also departed from the traditional approach of stylometry by ignoring entirely meanings and grammatical functions, measuring instead the frequencies of words according to their orthographic content. This has the effect of detaching the enquiry from any reference points it might have had in common with conventional linguistic analysis. I hope to show that the basis for such an approach is sound and that the evidence obtained which illustrates the existence of quantifiable differences between works and authors is sufficient justification for using such (at first sight) unconventional methods.

Most of the detailed analysis relates to Plato. Because of the intervening centuries and the lack of biographical information we are not in a position to state with certainty which of the peripheral works of the Platonic canon were in fact written by him. Tradition, however, and current scholarly consensus do allow us to accept most of the major works as being of his authorship. There remains the problem of the authenticity of the *Epistles, Hippias Major*, and various minor works, the provenance of which is often disputed. An attempt is made to solve these problems and to show the general relationship

of all the Platonic works to each other, as well as the relationship of the style of Plato to that of other authors.

Finally, a detailed study is made of the chronology of Plato's dialogues. It is here perhaps that the greatest interest lies, because many uncertainties still remain and, despite the generally accepted division into early and late works, most commentators adhere to their own preferred order of composition within these two groups. Statistical methods, however, can lay claim to some sort of objectivity, and, although I do not put forward the results of my own research as being the definitive solution to the problem of Platonic chronology, yet I think that they do offer a more reliable guide to the sequence of composition than has been available till now.

For these reasons I hope that the work will be a stimulus to those whose interest lies in stylometry, as well as to those, *O terque quaterque beati*, whose study is Plato. If the latter can make the effort to grapple with the statistical concepts involved, which are not as daunting as a first glance might imply, I think they will find their toil amply rewarded.

2
The Choice and Measurement of Variables

In order to build up a definition or description of style to use in a statistical enquiry it is necessary to select something, or a group of things, which may be measured. These objects are called 'variables' and they could be anything which is accessible as a physical quantity within the text and are in some way related to what the author wrote or, in other words, to his or her style. Examples would be the average length of words or sentences, the number of nouns or adjectives, the number of function words, and so on.

In our case the criteria which are to be applied to the selection of suitable variables are threefold.

1. They must be easily recognizable and hence easily counted by the computer without any preliminary enhancement of the machine-readable texts.
2. They must score at a sufficiently high rate to be free of the uncertainties associated with low-level distributions and also to enable us to use small samples.
3. There must be a fair chance that these variables be linked to stylistic features and not be just measurements of random and haphazard events.

Of these three conditions the first one is of practical importance, because it is simpler to work on a raw, unembellished text (these are readily available for ancient Greek authors). Any additional work, such as tagging all grammatical features or reducing words to their dictionary form, will not necessarily produce more favourable results. Usually such work has to be done manually and is very time-consuming. It is preferable to adopt a system which allows for the gathering of information with the minimum of fuss and textual intervention. The method then becomes more readily applicable to other texts, wherever and whenever one might wish to extend the enquiry.

The second and third criteria are of theoretical importance. It is necessary to escape from the restrictions imposed by the reliance on the measurement of individual words according to their meaning, since this produces readings of 1% or less, except in the case of a few common words.[1] Such low scores

---

[1] See e.g. A. Ellegard, *A Statistical Method for Determining Authorship* (Gothenburg, 1962), and F. Mosteller and D. Wallace, *Inference and Disputed Authorship: The Federalist* (Wok-

have to be compensated for by using larger samples—complete works in most cases.

In most studies it is found that the frequently occurring words do not act as good discriminators of style.[2] There is too much overlap between authors and, although one might find cases where differentiation seems to be possible, it is usually only between two selected authors, and the characteristic chosen does not have universal validity as a unique mark belonging only to the chosen author's style.

With regard to sample size, inequality of length (usually measured by the number of words) has been a recurrent problem in nearly all stylometric enquiry. Ideally, a system which can deal with small samples is preferable, since many works do not satisfy the minimum length requirements that would give statistical reliability when using variables of low frequency. A word occurring at an average rate of 0.5% should, in theory, score 50 in a sample of 10 000 words. But any one such sample selected at random from that author might give a reading as high as 71 or as low as 29, or be even more extreme.[3] It is true that in the majority of cases, assuming a large number of 10 000-word samples to be available, the figures would be considerably closer to 50, yet in general the approach could not inspire a belief that what had been measured could act as a pointer to the authorship of that sample, for it is evident that there would be much overlap between authors on this one variable, even though their 'true' averages were not in fact identical.

Besides, it is my intention to use much shorter samples, all of uniform length, namely 1000 words. Such an approach was recommended as long ago as 1897 by Lutoslawski.[4] This will enable direct comparisons to be made between all samples and, by splitting longer works into samples of this length, a better idea of the homogeneity of individual works or authors could be obtained. Such relatively small samples will obviously introduce a much greater level of fluctuation than with longer samples, and the presence of high-scoring variables becomes crucial.

I have set, therefore, an arbitrary lower limit of 5% for the variables to be measured. This is only to be taken as a rough guide, as in practice it is found that, for some authors, the score on a particular variable might drop below this value but, since the idea is to measure the same quantities for all samples, such readings have been retained. The nature of the subsequent statistical

---

ingham, Berks., 1964), for fairly extensive word lists. In the latter book the authors find only four words which score consistently above 2%, these being 'and', 'of', 'the', and 'to' (p. 244).

[2] The four common words mentioned in n. 1 were found to be inadequate as discriminators between the styles of Hamilton and Madison.

[3] See A. J. Kenny, *The Computation of Style* (Oxford, 1982), ch. 7. The figures relate to the distribution of sample means. 99% would be in the range quoted.

[4] W. Lutoslawski, *The Origin and Growth of Plato's Logic* (London, 1897), pp. 141–3, 185.

analysis will ensure that only those readings which contribute to the grouping together of similar samples will be used and all other readings will be ignored.[5]

The other criterion, that the variables should relate to style and be capable of giving us some sort of quantification of that elusive concept, is something which cannot be insisted upon too rigorously at this stage. Some general observations, however, can be made.

Greek is a highly inflected language. Words change their form according to their function in the sentence. With verbs the changes are linked to voice, mood, tense, and person; with nouns, to case and number. Other parts of speech vary in a similar manner. Most commonly, the alteration is to the word ending, but sometimes the stem alters, and sometimes a prefix is added to the word. If we could count words according to their orthography, therefore, it seems probable that we could pick out certain stylistic features, possibly even isolate them if we so wished.

This suggests that there are two possible approaches. One is to predefine certain categories which are closely linked to grammatical features, specifying in advance the orthography (usually word endings) which will define these features and then counting the words which exhibit the required patterns. The other is to do a blanket count of all words according to their alphabetic content, paying special attention to word endings. Of the two the former method presents problems due to the difficulty of adequately defining grammatical features by orthography; the danger of applying too rigorous an interpretation of grammar and syntax; the need for frequent intervention in the process of selection of suitable categories; and the probability of neglecting significant stylistic features because of the bias imposed by a predefined concept of style.

The latter method is not without its disadvantages, for it is likely that we will gather more information than we need; that some of it will be irrelevant or superfluous; and it will be difficult to relate the variables directly to any recognizable stylistic feature. However, on balance, since style is such a nebulous concept, it seemed better to adopt this method. Better, in fact, to count too much than to miss important facts through a harsh selection process. By doing so we may be fortunate in filtering out some of the unsuspected features of style which a more rigorous method, with all its preconditions and assumptions, would be likely to miss.

The basis of the word count, therefore, is to be quite simply their letter content, and three categories are defined as follows:

1. words containing a specified letter;
2. words ending in a specified letter;
3. words with a specified letter as penultimate.

[5] This is a simplification of MVA. Readings which do not contribute much to the information which we are seeking are given a lower weighting.

All categories were recorded as a percentage rate of occurrence and a fuller description of each now follows.

1. ALET1–ALET19. These 19 variables record the percentage rate of occurrence of words according to their overall letter content. That is, a word is counted only if it contains the specified letter which that particular variable requires. If the letter occurs twice or more times in a word, the count for that word is still only considered to be one. Thus ALET1 counts the percentage of words containing a, ALET2 the percentage containing γ, and so on. The Greek alphabet contains 24 letters, and the reason we have only 19 variables is that six of the letters occur so infrequently that it is desirable to combine them all into one reading. Thus ALET18 records the percentage of words containing β, ζ, ξ, ϕ, χ, ψ. A summation of ALET

$$\sum_{i=1}^{i=19} \text{ALET}i$$

gives an approximation of average word length. This becomes apparent if we imagine an extreme situation in which every word of the sample contains every letter of the alphabet occurring once only. In that case every ALET would be 100%, and $\sum$ ALET would be 2400%, indicating an average word length of 24.[6] In practice, there is a certain amount of repetition of letters in words, and $\sum$ ALET underestimates the average word length, depending on the frequency of letter duplication within the sample. Mathematically,

$$\text{Average word length} = \kappa \sum \text{ALET}i/100$$

where κ is a constant, greater than unity, which corrects for the effect of replication of letters in words in the sample. We may use this information to calculate letter distribution. For each ALET we could calculate a corresponding variable, say ZLET, which measures the distribution of each letter in the sample, irrespective of word length.[7] Thus

$$\text{ZLET} = \left[\text{ALET}i/\kappa \sum \text{ALET}i \right] 100 \quad \%$$

In fact we do not know the value of κ, so that we canot make the calculation. But it is probable that it does not vary much from sample to sample, so that, if we assume $\kappa = 1$, we may calculate the series of ZLETs and these will be an approximation of letter distribution.

It is possible, of course, to measure letter distribution directly, using a

[6] Strictly speaking only 1900% since there are only 19 categories, with ALET18 counting the rates for all six low-scoring letters combined. My illustration shows what would occur if we used all 24 ALETs.

[7] I have simplified the problem considerably. In fact κ would not be a constant but would depend on both the rate of replication of all letters in words of the sample and the rate of replication for the individual letter being measured. These two quantities would vary from sample to sample and the relationship between ALETs and ZLETs would not be linear.

suitable computer program. The Oxford Concordance Program (OCP), which was used for all the data collection undertaken in this thesis, does not have this facility, but it is well adapted to counting words according to their orthographic content. Consequently, letter distribution, or ZLET as we have chosen to name it, only occurs as a secondary variable. In most cases, in the subsequent analysis, it is the ALETs which have been used.[8]

It has been suggested that letter distribution shows only such variation as would be expected from a random selection of samples taken across the whole spectrum of the language and that idiosyncrasies of style would not affect this pattern. This may be true of an individual letter taken in isolation, but certainly not of the whole range when combined in multivariate statistical analysis (MVA). The ZLETs can be shown to have considerable discriminating power. In practice, however, the ALETs were found to work better as discriminators and these were always used in preference to the ZLETs.[9]

2. BLET1–BLET9. It was found that only nine letters of the alphabet occur sufficiently frequently as word endings to justify individual measurement. These letters are α, ϵ, η, ι, ν, o, σ, υ, ω. Each BLET records the percentage of words ending in one of these specified letters. If all 24 letters had been separately counted, then the sum of all possible 24 BLETs would have been 100%. This follows by definition, since a word must end in one of the 24 letters of the alphabet. It suggests that we could calculate a tenth variable, BLET10, such that

$$\text{BLET}10 = 100 - \sum \text{BLET}i \quad \%$$

This would record the total percentage of words which end in a letter other than that included in the set BLET1–BLET9. In the sense that this quantity would always be predictable from a knowledge of the previous nine variables, it is superfluous and does not add anything new to our information on the sample. Variables which are linear combinations of others in the same set create a singular matrix and as such are liable to be rejected in most applications of MVA. But it is worth bearing in mind the summation characteristics of the BLETs as a rough check on wild values and for help in the elimination of errors.

3. CLET1–CLET9. These measure the percentage of words occurring with the specified letter as penultimate. Again, only a limited range of letters is found in this position in the word with a sufficiently substantial frequency to justify measurement. The 9 letters are α, δ, ϵ, η, ι, o, τ, υ, ω. Similar

---

[8] In all cases which I describe subsequently I have used the ALETs (or BLETs or CLETs). The ZLETs were used occasionally in the early stages of the enquiry.

[9] I am uncertain if letter distribution, when correctly measured, could act as such a powerful discriminator of style as the ALETs. I suspect that it would approximate to the distribution of the CLETs, which represents, in a sense, a random dipping into the text.

summation characteristics are found as for the BLET variables, except that in this case a certain proportion of words, say $n\%$, contain only one letter and by definition cannot be included in this category. The equation for CLET10, equivalent to that shown for BLET10, would therefore be

$$\text{CLET}10 = 100 - (n + \sum \text{CLET}i) \quad \%$$

It would be appropriate to include CLET10 as one of the CLET variables, if one wished, since it is an independent quantity. However, for the present study, the necessary counts have not been made and we remain content with the nine specified CLET variables.

The full list of 37 variables which are to be measured is given in Table 2.1. It would have been possible to measure even more variables than those chosen, but time and the nature of the research prevented this: decisions had to be taken based mostly on intuition and without knowing in advance what measurements would contribute most to a knowledge of style. A fairly large number of samples had to be used to test the efficacy of the methods and to

TABLE 2.1   *List of 37 variables*

ALETs		BLETs		CLETs	
	% of words containing		% of words ending		% of words with penultimate
1	α	1	α	1	α
2	γ	2	ϵ	2	δ
3	δ	3	η	3	ϵ
4	ϵ	4	ι	4	η
5	η	5	ν	5	ι
6	θ	6	o	6	o
7	ι	7	σ	7	τ
8	κ	8	υ	8	υ
9	λ	9	ω	9	ω
10	μ				
11	ν				
12	o				
13	π				
14	ρ				
15	σ				
16	τ				
17	υ				
18	6*				
19	ω				

* Words containing any of the following: β, ζ, ξ, ϕ, χ, ψ.

do this over and over again with different variables would have been prohibitively costly in time and effort.

It is widely believed that the greater the number of variables the greater will be one's success in discriminating between styles.[10] I suspect that this is a fallacy, for what really counts is the level of correlation between each variable and the differences in style which interest us, whatever may be the cause of these differences. In addition, the multiplication of variables increases the computing time required for any calculation, while in many applications of MVA the number of variables which may be used is restricted by theoretical constraints.

Consequently, I decided to limit the number of variables to be measured to the 37 described, for even this number represents a considerable degree of superfluity and there is no guarantee that increasing the number to 100 would have produced better results.

---

[10] Lutoslawski used 500 features and recommended the use of 10 times that amount (*The Origin and Growth of Plato's Logic*, p. 161). More recently Birch uses 252 style characteristics to analyse texts by Thomas More (D. Birch, 'The Stylistic Analysis of a Large Corpora of Literary Texts', *ALLC Journal*, 6(1 and 2) (1985), 33–8. With MVA 252 variables could adequately classify 253 samples of text, even if all texts were by different authors. Using 1000-word samples, this represents a colossal mass of text, much larger than Birch himself proposes to deal with. MVA makes better use of a smaller amount of data by emphasizing the salient points. This is important, because it reduces the labour involved in the collection of data and makes the subsequent analysis more intelligible.

3
Technicalities

THE Greek texts were obtained in machine-readable form from Thesaurus Linguae Graecae of Irvine, California.[1] These texts use a transliterated Greek alphabet which is basically the equivalent in Roman upper-case letters of the Greek sounds, with some variation. Thus A = α, B = β, G = γ, D = δ, and so on. Accentuations, breathings, subscripts, capitals, punctuation, and a variety of other devices are dealt with by a series of symbols. Ancient Greek has 24 letters but it could be argued that the range of possible letters is effectively increased by the use of rough or smooth breathings added to a vowel, by the iota-subscript which sometimes occurs in combination with some of the vowels, and by the widespread use of accentuation.[2]

Whatever may be the classical scholar's interpretation of these matters, it is imperative for anyone hoping to use a computer for counting words according to their orthographic content to decide how these additional symbols are to be treated. Should they be ignored, or treated as supplementary letters, or considered as creating a new class of letters whenever they are added to any one of the pre-existing set of 24? In the end, it seemed simplest to ignore all but the original 24 letters and to treat all the rest as padding, the elimination of which would not effect the main trend of the study. Although such a brutal approach to language would certainly be inappropriate if the objective were to produce a concordance, in these circumstances it is simply a question of practical efficiency, for here our interest lies in the measurement of certain well-defined numerical quantities. If, alternatively, it had been decided to treat α with rough or with smooth breathing, with or without the iota-subscript, with accents acute, grave, or circumflex, or unaccented, or any combination of these forms; if all these various possibilities had been broken down and treated as different letter forms, then out of the single letter α about 10 other letters would have been created.

---

[1] Thesaurus Linguae Graecae intends to have all extant Greek authors (600 BC to AD 600 approx.) available in machine-readable form within the next few years. The project is already almost complete. Explanatory literature may be obtained from TLG, University of California, Irvine, CA 92717, USA.

[2] All alphabets depend on the conventional use of certain symbols to represent particular sounds and evidently some alphabets are more effective or more accurate in this respect than others. In Classical Greek the rough breathing corresponds approximately to the English letter 'h', the iota-subscript is a survival from an earlier stage in the language's development, while the system of accentuation seems to stem from the Alexandrian scholars of the 3rd to 1st cents. BC. It relates more to the written than to the spoken language. The Greek of the authors whose work is studied in this thesis was written originally without accentuation.

Clearly, it is important to avoid such fragmentation because we would be reintroducing the problems which the main impetus of the study has so far been seeking to avoid, that is, the problems associated with low-scoring variables. Many of the readings would be so low that we would be driven to recombine them, thus undoing the laborious work of separation. Therefore, throughout the study, each individual letter is treated as itself pure and simple, no matter what modifications or attachments may or may not be present. This will have the effect of keeping the readings of the main bulk of the variables at an acceptable level.

A theoretical justification for this approach may be sought both in the concept of style which we are attempting to measure mathematically, and in the use of multivariate analysis to cope with those measurements. For, in the count which measures the percentage of words containing a, or any other letter, there is no implicit statement that the feature in itself is a pure criterion of style. At this stage it is, in any case, impossible to know what is the relevance of any one variable to the final model or definition of style which the mathematical analysis will produce. All the variables are in fact regarded in precisely the same light and as having equal importance, since the underlying assumption is that they are the product of a multitude of hidden forces working within the confines of language. It is the interaction of these forces with the constraints inherent in language which produces style, but we cannot say that we are in a position to isolate any one of them for the purposes of measurement, any more than it is possible to isolate intelligence, or other human qualities, although we can measure their effect in a variety of situations. This effect may be distorted by various extraneous influences. The subsequent mathematical treatment will decide what importance is to be attached to the individual measurements, whether they are to be magnified or reduced, what weight or relevance they are to have in the context under investigation.

In fact it is precisely this freedom which MVA has in dealing with the recorded observations which encourages a robust attitude towards the collection of data. It is not necessary to be too chary in making arbitrary decisions or cutting the Gordian knot when faced with seemingly intractable linguistic problems, for the mathematical processes which MVA employs are quite capable of selecting the data which is valid and rejecting the superfluous. This will become more apparent and will be dealt with at greater length in Chapter 5, on the statistical background and MVA.

A recurrent problem is that presented by the occurrence of quotations in the text. Prima facie these are the work of another author and should not be included. However, relying on the principles outlined above, it seemed reasonable to assume that the methods used would be sufficiently robust to cope with such intrusions. Overall, the level at which quotations occur must be less than 1%, although certain sections of Plato are noted for the liberal

sprinklings of ancient authors, in particular Homer in the *Republic* and Simonides in *Protagoras*. But even were we to suppose that quotations accounted for 10% of the text, which is an over-generous allowance, it does not follow that readings of the 37 variables would be in error by 10% in absolute terms. Only if the quoting author used a totally different set of letters and word endings than those used by the author quoted would we be justified in making this conclusion. The natural constraints of language, however, make such an eventuality impossible, and the probable error is much lower than the bald figure for the proportion of quotations would suggest.

For any given variable, if the proportion of quotations is $x\%$, then the error in the observed readings for that variable will not be greater than $x\%$ of the greatest difference known to exist between readings on that variable of the original author and any other author comparable to the one quoted. This sounds rather complicated but, exempli gratia, let us suppose that in one sample of Plato 10% of the text consists of Homeric quotations. The greatest difference on the ALET1 variable for these two authors is, say, $52\% - 33\% = 19\%$. The greatest possible error for the ALET1 variable would therefore be $0.1 \times 19\%$ or 1.9%, which is well within the range of the standard deviation of the Plato samples. But this is an extreme reading, and the average error would only be the pro rata amount of the average difference found between these two authors for the variable in question.

In fact we do not know any figures for Homer, since this study deals with prose authors only, but it is reasonable to suppose that the possible errors will be of the order of magnitude of those quoted above. Obviously it is necessary to be prudent in such matters, and there must be a point (somewhere below 50%) at which the dilution of one author by another distorts the readings sufficiently to require the excision of quotations. MVA takes into account the total pattern produced by all the variables, and what appears to be a small shift in a small number of variables may loom large in the final analysis. It is not likely, however, that any of the samples used in this study have reached this critical level, and the general rule was adopted of including quotations.

An exception was made in the case of the treaties in book 4 of Thucydides' *History*. These appear to be treaties, some of considerable length, quoted verbatim from source. But, whatever their origin, it would be unwise to allow such large chunks of alien material to be taken as representative of Thucydides' style, and all of them were excluded from the relevant samples.

The funeral oration in Plato's *Menexenus* and the supposed Lysias speech in the *Phaedrus* (also by Plato) are special cases and are discussed fully in Chapter 8, which is devoted to the analysis of the Platonic corpus.

OCP and the collection of data

The texts were read on the IBM Amdahl machine at Reading University Computer Centre. This machine provides text-handling and -editing facilities as well as the two packages which were used in the analysis, namely, OCP (the Oxford Concordance Program, 1983 version) and SAS (Statistical Analysis Systems, 1983 version).

OCP was used for all measurements of variables, or word counts. Since all samples consisted of 1000-word extracts, it was a simple matter to convert the reading for each variable into a percentage. The output from OCP is considerable more extensive than that required for this sort of measurement, as it gives a list of the numbers of words occurring at the various possible frequencies between 1 and the maximum found for this variable. A considerable amount of editing is needed to obtain the one figure that gives the word count which we are seeking. There is no interface between SAS and OCP, so that it was not possible to do the necessary runs with OCP on the sample and transfer the output directly to a SAS data set.

In the early stages the data collection and transference to a SAS file involved a great deal of computer printout and manual entry of the figures. But gradually, with help from the Computer Centre, and with increasing expertise, it became possible to mechanize the whole process, so that from beginning to end there was no necessity to set pen to paper, thus reducing to a minimum the likelihood of error. Using a batch system and a series of Xedit macros, as many as 20 1000-word samples could be processed in a day. This entailed the handling of 740 separate OCP output files.[3] Beyond this figure one begins to tax the resources of the machine, and it becomes increasingly difficult to ensure that the appropriate results are attached to each sample. A systematic nomenclature for the output files is absolutely essential and helps considerably to ease the process of collection.

In addition to the values of each of the 37 variables, the following information for each sample was fed into the data table:

1. *Author*. The author's full name was used, namely, Aeschines, Isaeus, Isocrates, Lysias, Plato, Thucydides, or Xenophon (Greek authors of this period are generally known by a single name only). In some cases the computer printout truncates the name to the first eight letters.

2. *Work or document*. The common title of the work was used, sometimes in abbreviated form. The computer printout truncates this title to the first eight letters.

3. *Sample*. A coded alphanumeric marker, unique to each sample, was

---

[3] 37 variables are measured for each sample, giving a total of $20 \times 37 = 740$ separate output files. An Xedit macro is a program which may be used when processing files on the Amdahl computer. It can do a lot of repetitive operations on a large number of similar files, thus freeing the researcher from the necessity of manual handling of the data.

devised, which makes it possible to identify each one and to locate its position in the work to which it belongs. Thus *Lg.* 305 refers to the 5th sample of book 3 of Plato's *Laws*, *Ant.* 12 refers to the 12th sample of Isocrates' *Antidosis*, and so on. Since the samples were sequential and a record was kept of the initial line of each, it was possible to refer back to the original with comparative ease and relate the statistical conclusion to the character of that part of the work.

4. *Genre*. A code number was used which loosely describes the genre of writing to which the sample belongs. The genres were:

G1. Forensic oratory,
G2. Political oratory,
G3. Philosophical dialogue,
G4. Philosophical treatise,
G5. History,
G6. Biography,
G7. Miscellaneous.

This variable was introduced in case it should become necessary to correct for genre dominance in the statistical analysis. The boundary line between the various genres is not rigid and in places scarcely applies. However, it was deemed useful to retain the notion, not least because of the uncertain nature of stylometric investigation and also because, although it is always possible to ignore or discard superfluous information, it is often troublesome to provide what is lacking at a later stage in the proceedings.[4]

5. *Group*. A group category was introduced because it might prove useful to be able to isolate particular works in certain groups, as, for example, early, middle, or late Plato, or works dealing with a predefined topic. Having this facility it might then be possible to isolate the characteristics which identify the different styles used by authors at different periods of their lives. The coding for this variable takes the form GPx or GPPx where x is a distinguishing numeric variable. Thus GPP1 might refer to the group containing the putative first period Platonic dialogues.

In practice both genre and group identifiers could be altered comparatively easily to accommodate various interpretations of the nature or chronology of each work or sample.

As it turned out, the above two categories were little used in the subsequent investigation, since so much information was provided by the other variables and the main task of identifying author characteristics or plotting chronological stylistic change did not require the use of genre or group definition.

A typical printout of data, in this case for samples of Plato's *Apology*, is

---

[4] The list of genres is obviously inadequate, but the point is unimportant because in the end I did not make use of the information in any of the analyses.

shown in Figure 3.1. The initial line for each sample contains the information Author, Document, Part no., Genre, Group, in that sequence. The remaining 37 numeric variables, four lines for each sample, then follow, commencing with ALET1 and finishing with CLET9.

Sample size and the mechanics of division

At an early stage it was decided to use samples of 1000 words and all the works studied were divided into sections of suitable length. The great advantage of using equal sample sizes is that direct comparisons between samples can be made without making adjustments for differences in variance which would arise inevitably with different sample sizes. Yet it is surprising how rarely this approach has been used. The tendency has always been to compare one work with another, either ignoring variations in length or attempting to make an adjustment to correct for it. But it is important to know how much variation is found within a single work in order to be able to compare it with other works by the same or different authors. Much of the strength of MVA lies precisely in this comparison of within-group and between-group variances, and unless the work is split up into samples the first of these quantities cannot be calculated.

Evidently there is a need to strike a balance between using very short samples, where the fluctuations in variable readings would be high and difficult to work with, and the opposite extreme, when greater stability would be achieved, but with loss of flexibility, because many shorter works would automatically be excluded from the study.

1000 words is an obvious choice as it facilitates calculation and is likely to be safely above the critical size beneath which styles cannot be differentiated. For there must be a lower limit beyond which identification becomes impossible, and, although this limit cannot be fixed with certainty, probably it will not be less than a score or so of sentences. 1000 words offers convenience of computation and will be used throughout this study.

Finally, it should be emphasized that, although the word is the obvious unit for measurement and perhaps the natural one, it is not the only candidate. Letters, syllables, morphemes, phrases, sentences, lines of text, or pages could also be used. Each would demand a separate approach but it would be no less logical than the system which relies on word counts. Most text-handling packages on computers, however, work most easily with words as units, so that it is convenient to fall in with this convention, which in any case satisfies the most commonly held notions about language, and at the same time spares us the task of justifying any unusual or unorthodox linguistic theories.

The samples used were approximately sequential blocks of 1000 words of

```
    PLATO    APOLOGY    APO1    G1    GRPP1
40.0 11.7 11.3 41.0 17.5 5.5 39.9 15.0 13.0
14.8 40.0 40.8 14.5 15.1 26.8 32.2 25.6 9.7 19.8
6.2 7.1 5.0 20.2 26.7 2.4 16.9 3.5 6.7
14.2 3.9 7.6 6.3 9.9 15.7 10.6 4.9 10.4
    PLATO    APOLOGY    APO2    G1    GRPP1
40.9 9.4 11.5 38.3 15.3 6.6 42.4 12.0 10.4
15.5 38.1 41.2 13.6 12.8 23.8 32.6 22.8 11.8 16.5
8.4 5.3 4.6 23.4 25.8 2.4 12.9 4.3 6.3
16.3 3.5 8.1 5.4 8.1 16.1 10.9 2.9 10.0
    PLATO    APOLOGY    APO3    G1    GRPP1
36.7 9.2 13.0 39.4 14.4 7.0 42.6 14.0 12.8
13.4 35.2 43.9 13.2 13.2 28.0 33.4 24.3 10.8 19.1
6.8 9.0 5.0 22.5 23.4 2.1 17.2 3.8 5.8
12.7 4.9 6.8 3.8 11.5 13.5 12.6 6.3 10.1
    PLATO    APOLOGY    APO4    G1    GRPP1
41.0 8.8 12.6 39.5 16.1 8.8 43.0 14.8 9.6
16.3 36.1 41.7 11.4 12.8 26.8 28.3 22.1 10.1 14.5
8.2 7.1 5.8 21.3 19.7 2.1 17.3 6.1 6.3
14.5 3.7 7.9 6.1 10.7 15.1 8.2 5.9 7.5
    PLATO    APOLOGY    APO5    G1    GRPP1
42.8 8.3 11.0 43.9 16.8 7.5 40.9 12.1 12.4
19.6 34.6 37.8 15.2 13.5 23.9 30.0 23.2 11.4 17.0
8.1 9.1 7.0 19.8 22.2 3.0 13.4 3.6 7.5
14.2 4.0 7.5 6.4 7.5 13.6 11.1 3.2 9.1
    PLATO    APOLOGY    APO6    G1    GRPP1
42.5 7.4 16.2 42.0 14.3 7.8 43.2 14.1 10.6
14.4 35.0 43.0 12.5 13.9 26.2 33.2 25.7 11.9 16.3
6.9 8.0 4.8 21.1 23.2 2.6 16.9 4.8 5.6
12.7 5.0 7.4 3.9 10.8 17.1 11.0 4.2 8.9
    PLATO    APOLOGY    APO7    G1    GRPP1
39.8 9.9 11.3 38.7 15.2 7.7 41.4 11.3 12.9
18.9 38.2 39.0 11.6 12.2 24.3 30.3 23.5 9.6 18.1
6.2 7.0 4.1 22.5 25.6 2.3 14.6 6.5 6.3
14.7 4.2 5.6 4.7 8.2 18.5 9.8 3.9 10.6
    PLATO    APOLOGY    APO8    G1    GRPP1
45.2 8.3 12.5 39.4 15.9 9.2 43.8 14.9 10.4
16.8 37.9 39.5 11.4 12.2 25.4 33.3 23.4 10.3 13.8
6.1 5.9 6.6 23.2 22.8 3.1 15.5 5.9 5.4
17.7 3.2 7.3 3.8 10.9 17.0 10.9 3.4 6.8
```

FIG. 3.1   Typical computer printout of data

text. The intervening gap between samples was of the order of 100 words, usually less, so that at least 90% of the works in the study were analysed. This is a far greater proportion than that demanded by the statistics of random sampling,[5] but, since it is a comparatively easy matter to extend literary and stylistic measurements indefinitely without incurring prohibitive costs, it seemed best to remove at the outset the possible doubt that literary production lay outside the bounds of material which could be properly studied by statistical methods. By demonstrating the homogeneity which characterizes most of the variables measured, and this even when almost the entire work has been scrutinized, this doubt has been removed.

The mechanics of text division by which samples of at least 1000 words were obtained was such that the ideal of entirely consecutive samples could not be achieved. From a preliminary knowledge of the number of words in a work divisions were made in the text of such a size as to contain slightly more than 1000 words. The calculation was based on the number of records in the file which contained the work. It was important that each sample contain at least 1000 words, since OCP, if instructed, will count only the first 1000, but if there is a shortfall it cannot make up the deficiency. Consequently, samples which were found to be short were adjusted to include the full number of words by stripping some from neighbouring samples which had an excess. If necessary, when too many samples were found to be short, the entire division process was rerun to achieve a better distribution.

No doubt a FORTRAN program could have been created to divide the text accurately into the proper lengths, but the *ad hoc* methods adopted, using a simple program, were found to be more or less trouble-free and easy to use, despite the description given above, which might suggest otherwise. The format of the Thesaurus Linguae Graecae texts is based on that of the original edition from which the machine-readable texts were taken (often the Oxford Classical text). But there is not a complete line-for-line correspondence because marginal information from the original, such as Stephanus page references for Plato, might occupy a whole record in the computer file. Consequently, considerable variation existed in word density and no simple standard of division could be adopted.

Care was taken, however, to avoid any overlapping of samples and, except in the case of one speech of Lysias, no words of text were included more than once, and all the samples were unique.

---

[5] In fact this procedure of splitting the text into sequential units cannot be considered as true sampling. This should be a random process if classical sampling theory is to apply. It seems probable, however, that the random element is introduced by the exigencies of subject and linguistic constraints operating on the author, forces over which the researcher has no control. Consequently, it may not be entirely inappropriate to assume that this element of constraint at the primary level of composition replaces the random-selection process which would normally be required, and that the central-limit theorem and other parts of sampling theory may be relevant.

To summarize the foregoing: all samples consisted of 1000 words of text taken as a block; apart from an intervening gap averaging less than 100 words, the samples were sequential; at least 90% of each work was sampled in this way and overlaps were not permitted.

Effectively, therefore, the entire population will have been surveyed, rather than a fragment of it, and, whatever other defects this study may be found to have, it may at least claim to have made a full study of each of the works analysed and to be in a position to show how much variation does in fact occur in individual works.

4

The Authors

THE seven authors studied were all either Athenians or resident in Athens in the fifth and fourth centuries BC. Their writings show considerable diversity of style, but all are in the Attic dialect and may be described reasonably as prose. The decision to avoid the work of poets was deliberate and sprang from the presupposition that prose works exhibit a greater degree of regularity and uniformity than do the productions of poets. Perhaps this may seem surprising to anyone accustomed to considering the metrical and formal aspects of poetry, but those who have studied languages will know that poets stretch syntax and vocabulary to their limits. They work, so to speak, at the edge of the known world, and the variety and extremes inherent in their use of language may be very difficult to quantify by any known system of measurement. Prose seemed to be a safer, more appropriate, and more tractable choice and, rightly or wrongly, the study is limited to the works of seven prose authors of ancient Athens. They are Aeschines, Isaeus, Isocrates, Lysias, Plato, Thucydides, and Xenophon.

I have assumed that the texts are substantially the same as those which were available to ancient readers and that textual corruption is not a major problem in respect of the authors studied. To a certain extent this is an act of faith, since we cannot consult the original manuscripts. However, there is some evidence in the form of papyrus remains to justify the inference that, except in the cases where only fragmentary remains are available, textual corruption has been fairly limited.[1] It would not have been sufficient to invalidate the concept of individuality of style as applied to any of these authors. Besides it will be possible, using the data which has been gathered, to test for textual integrity.

The essential test will be that of homogeneity, for it is unlikely that this would be discovered in a work which had suffered massive textual corruption. Where homogeneity is found we could conclude that any corruption is slight or insignificant, or, at its worst, widespread but uniformly distributed.[2]

---

[1] See e.g. B. R. Rees, *The Use of Greek* (Cardiff, 1960), p. 16. The texts of Plato are generally thought to have suffered little in transmission. In any case we shall not be concerned with the alteration here and there of a single word, but with a much broader and (hopefully) more robust measurement of style which uses every word in the text.

[2] The differences of style which are implied by such results as are obtained in ch. 6 and illustrated in Fig. 6.1 indicate that homogeneity does exist within individual works, despite the vagaries and uncertainties of textual transmission.

Ideally the new stylometric methods introduced in this study should have been tested in a more congenial environment, where the historical tradition was less tenuous and verification of the results could be achieved with certainty. The advantages of using classical Greek are that it is readily available in machine-readable form; that it is an inflected language which probably responds more satisfactorily to this form of approach than the only slightly inflected modern European languages; and that my own background as a classicist added a personal interest to the venture over and above that afforded by the purely stylometric content of the investigation.

Some brief biographical details of the individual authors now follow.

Aeschines (c.397–322 BC)

Aeschines was active in political affairs in the days when Philip of Macedonia was trying to extend his hegemony over the Greek mainland. The three speeches which are attributed to him are all related to the political rivalry and enmity which arose between himself and Demosthenes, the most famous of the Greek orators. Technically these speeches were for delivery in the law courts, but they have a strong political flavour and serve to illustrate the close connection between the democratic institutions of Athens and its law courts. As Aeschines himself remarks, 'very often private enmities correct public abuses' (*Against Timarchus*, sect. 2). The concept of 'isonomia' or equality before the law, which in practice meant that anyone could bring a prosecution against a third party whom he felt was transgressing the laws and institutions of the State, was crucial to the democratic tradition.

Aeschines suffers under the commanding reputation of Demosthenes, who championed the cause of Greek freedom. Stylistically he is a master of clear exposition and vivid narrative, but he never quite reaches the dizzy heights of rhetoric which we associate with Demosthenes.

Of his speeches only *Against Timarchus* was available at the time from TLG, the rest of his work still being in an uncorrected state and not available to subscribers. In the analysis the laws and depositions included in the speech, which would have been read out before the court by an official, have been excluded, as they are generally regarded as later interpolations and, even if genuine, would not be relevant to a study of Aeschines' style.

Isaeus (c.420–350 BC)

Nothing much is known for certain about this author. It is even doubted whether he was an Athenian citizen and some traditions suggest that he was Chalcidian. Traditionally he was a pupil of Isocrates and a teacher of Demosthenes. But, whatever his background, it is clear that he wrote speeches for pleaders in the Athenian courts. The surviving speeches are all connected with testamentary disputes and interpretations and indeed are the chief source of our knowledge of such matters for the period. Isaeus is noted

for his ability to explain complex and tangled legal situations pertaining to inheritance in an easy and flowing style.

Other speeches dealing with cases of real property, guardianship, sureties, adoption, assault, and rights of citizenship were known in antiquity. According to the pseudo-Plutarchian *Lives of the Ten Orators*, Isaeus left 64 speeches, of which 50 were regarded as genuine. Only 11 survive, plus some fragments and, inevitably, doubts as to authenticity must arise. However, no special attention has been given to the problem in this study.

Some work has been done by R. F. Wevers on Isaeus' chronology using stylometric methods. It is based on the use of particular clausulae rhythms and the changing preferences for certain forms exhibited in various speeches.[3] My own work on Platonic chronology will show, however, that the results obtained by reliance on one variable, in this instance clausulae rhythms, are unreliable and will indicate a sequence which is only one of many possible chronological sequences which might be obtained using other variables. Wevers's study represents a considerable simplification of the problem and it would be unwise to accept his conclusions. A brief investigation which I conducted on the Isaean corpus using multivariate methods revealed that *Oration* 3 was highly distinctive but beyond that point it was difficult to draw any satisfactory conclusions about the chronology of composition of the speeches (see also Chapter 6).

Isocrates (436–338 BC)

Isocrates is one of the most distinguished of the Athenian rhetoricians, although he did not play an active part in politics. He appears to have distrusted his abilities as a public speaker and his speeches are therefore written for a selected audience or possibly directly for a reading public. In *c.*390 he opened a school of rhetoric at Athens, and he represents the commonsense view of rhetoric which sees it as the art which aids the politician, rather than the more extreme view of Plato which represents it as being the corrupter of moral standards.

Five of Isocrates' speeches were included for analysis. These were *Antidosis*, *Archidamus*, *De Pace*, *Panathenaicus*, and *Panegyricus*. I had hoped to do more detailed work on the entire Isocratean corpus, alongside that of Plato, but found that the demands of coping with even a moderate stylometric study of the latter author were sufficient to occupy all the available time. One thing that does emerge with great clarity from all the analyses in which these Isocrates speeches were included is the remarkable homogeneity of the Isocratean samples relative to those of the other authors. Always these were found to occupy the extreme range of any series of groups, with no danger that they could ever be confused with works by any of the other six authors,

[3] R. F. Wevers, *Isaeus: Chronology, Prosopography and Social History* (The Hague, 1969).

whereas among those other six there remained the possibility at some point of overlap and confusion.

Why this distinctiveness should be so apparent for this one author was a problem which I failed to solve, although the results do correspond to a certain extent with one's experience in reading Isocrates. He gives more attention to style than most other writers and as a teacher of rhetoric his achievements in oratory sought to match his lofty ideals. He was responsible for the development of a style of oratory which seeks to avoid hiatus, a feature which apparently influenced Plato and has been used by some scholars to date the dialogues.[4]

Though born before Plato, Isocrates lived until 338, some nine years after Plato and died in his 97th or 98th year.

Lysias (459–380 BC)

Lysias was of a Syracusan family which had been persuaded to settle in Athens by Pericles. Under the reign of the Thirty Tyrants the family fortune was seized and Lysias made his escape to Megara. His brother, however, was killed and their property confiscated. Owing to his unswerving loyalty to the democracy Lysias was granted citizenship but this was subsequently rescinded because of some technical error discovered in the original award. From that date (403) onwards he appears to have made a living by writing speeches for those engaged in lawsuits.

According to Dionysius of Halicarnassus his genuine speeches numbered 230. Of the odd 30 or so which survive, apart from a few, it is not possible to know which belong to the genuine corpus and which to the spurious, for the total number attributed to him in ancient times exceeded 400. A recent stylometric study by Stephen Usher and Dietmar Najak attempted to solve the problem of authenticity.[5] However, it would have been more useful if a few forensic speeches by authors known to be other than Lysias had been included. Comparisons could then have been made with the speeches from the Corpus Lysiacum thought most likely to be genuine. But the absence of any control against which the performance of a supposedly genuine speech could be assessed gave the impression that the foundations of the study were insecure.

Only the speech *Against Eratosthenes* (12) is used in this study as it is one of the longest and also the one which, owing to its autobiographical content, is regarded as having the greatest claim to authenticity. Samples 4 and 5 of this speech contain a small amount of textual overlap, because the overall length of 4870 words did not allow for the creation of five full-length

[4] L. Brandwood, 'The Dating of Plato's Works by the Stylistic Method: A Critical and Historical Survey', Ph.D. thesis (London, 1958).
[5] 'A Statistical Study of Authorship in the Corpus Lysiacum', *Computers and the Humanities*, 16(2) (Oct. 1982), 85–102.

samples. This results in a marginal distortion and makes Lysias appear to be a slightly less varied author than is really the case. To correct this error it would have been simple to have omitted either sample 4 or sample 5 from the enquiry, but since Lysias was only being used to provide background to the analysis I did not consider it necessary to make any special adjustments.

Plato (c.429–347 BC)

This author hardly requires an introduction. To modern readers he is best known as the author of the *Republic* and the *Apology*, the latter purporting to be Socrates' defence against the charge of corrupting the young and not accepting the city's traditional gods, a charge which subsequently led to his execution. The *Republic* sets out to be a treatise on justice and then expands to describe the entire constitution of an ideal city. To some it has appeared to be the precursor of communism, to others the blueprint for the ultimate oligarchical tyranny, to still others the depiction of the only *State* in which goodness and piety could flourish. Plato was much admired by Christians both early and later and, alongside Virgil, was considered to be one of the few pagans worthy of a place in heaven.

His work is immensely varied both in style and subject matter and probably it spanned a considerable period of his life, during which his style and philosophical thought continued to evolve. Authenticity is of relevance chiefly with regard to *Epistle* 7 which gives a great deal of biographical information. Other dialogues have been challenged at various times in the last hundred years, usually because the apparent philosophical content does not fit in with some preconceived notion of what Plato's philosophy should be. Often stylistic evidence is adduced, but this has rarely been objectively assessed or presented. Campbell, Ritter, and Lutoslawski established the stylistic argument on more objective criteria, but the dating of Plato's works is still rather insecure. Ryle most recently attempted a radical revision of the accepted order, but he does not have many followers.[6]

In this study the majority of the works are accepted as genuine, and the question of authenticity is dealt with in Chapter 8. Attention is then given to the chronological order of composition, and this is the subject of Chapter 9. The entire corpus, including the disputed works, was used in these analyses and a comprehensive survey of the affinities of the various dialogues is attempted.

Chapter 7 gives further biographical information and attempts to relate the stylometric analysis to the wider field of Platonic studies.

Thucydides (c.460–400 BC)

Thucydides was an Athenian citizen who served in the Peloponnesian War.

[6] G. Ryle, *Plato's Progress* (Cambridge, 1966).

In 424 he was appointed as a general but his failure to save the Thracian city of Amphipolis from capture resulted in his exile, an exile which lasted for 20 years. He devoted his time to writing a history of the war, a history which he hoped would be 'a possession lasting for all time and not a mere ephemeral showpiece' (I. xxii. 4).

Objectivity has been claimed for his *History of the Peloponnesian War*, sometimes even beyond what is reasonable. His most likely sympathies were, however, with the upper classes and he blends this leaning with a reserved admiration for the democratic traditions of his native city.

His style is generally considered to be difficult, sometimes even abstruse, but certainly exciting because of the compression of the thought and the novelty of the ideas expressed.

Only books 3, 4, and 5 are analysed and these are used in the background study which seeks to establish the validity of multivariate methods for determining differences of style.

All the treaties from book 4 have been excluded from the text.

Xenophon (428–354 BC)

Xenophon was from a wealthy Athenian family. In his twenties he joined an expedition led by Cyrus against his brother Artaxerxes who was then King of the Persian Empire. The expedition met with disaster when Cyrus was killed, and the story of the Greek mercenaries' flight of several thousand miles northwards to the Black Sea is told by Xenophon in the *Anabasis*. On his return to Greece he offended Athens by siding with Sparta and was banished. For many years he lived in exile. His works include the *Anabasis* already mentioned; a continuation of Thucydides' *History* which goes under the title of *Hellenica*; various works on instruction, estate management, and the like; short historical works and a number of pieces which use Socrates as the chief character and parallel, in some ways, the Platonic works, although not in their philosophic content. These works are the *Memorabilia, Oeconomicus, Apologia,* and *Symposium*.

His work is of interest to the stylometrist because he wrote in several genres and also because of his potential proximity to other authors, notably to Plato and Thucydides. I used, as part of the overall analysis and in the tests of author against author, the whole of the *Memorabilia* and *Oeconomicus* as well as book 1 of the *Hellenica*.

5
Multivariate Analysis

IF we were faced with a problem of classifying some naturally occurring objects into their correct generic categories, it would be appropriate to use some of the techniques of multivariate statistical analysis (MVA) to assist us. Suppose, for example, we wished to identify various trees, perhaps for use in a tree-planting scheme, to ascertain which species grew best in certain types of soil and climate, and that, because of cross-pollination and the possible existence of various subspecies, identification by simple visual inspection had proved impossible. Since MVA requires data to work on, the first requisite would be the collection of information in the form of measurements of various accessible features for each tree that was the subject of the enquiry. We could choose to measure such things as weight of seed, volume of seed, heartwood density, density of bark, hardness of bark, weight of leaf, number of leaf stomata per unit area, and so on. Obviously it would be best to avoid measurements which relate to age, such as height and girth of trunk, since these variables would not be linked directly to species, except perhaps as extremes, and would be misleading. An oak tree is no less an oak whether it is 2 metres or 50 metres high. We will also assume that only ordinal variables are to be measured, that is, those which may be expressed directly as a number which forms part of a scale, as weight, for example, or density, or length. Categoric variables, such as whether or not the seed is winged, whether or not the tree is deciduous, will not be included. This is for the sake of simplicity, because it is more problematic to deal with a mixture of variables of different sorts, and at this stage it is best to avoid complications. Obviously, the knowledge that a tree is coniferous or otherwise, deciduous or evergreen, smooth-barked or prickly, would be of great use for identification. But we will assume that none of these obvious differences exist in the trees which we are studying, for, in any case, it is always easy to identify types which exist at either end of a spectrum, and problems only arise when it is a question of breaking an apparent continuum into a number of discrete groups or clusters.

So, in the progress of this notional problem, which we hope is not too far removed from the problems of stylometry, we collect a series of measurements for each subject on a chosen set of variables. This series could be expressed in general terms as

$$x_1, x_2, x_3, x_4, ..., x_n,$$

where x_1 might represent weight of seed, x_2 volume of seed, and so on. Generalizing, we could say that x_i represents the measurement for each subject on the ith variable.

The total series of measurements for each subject,

$$x_1, x_2, x_3, x_4, ..., x_n,$$

is called an observation, so that if we had taken measurements on 100 trees we would have 100 observations, or, in general terms, measurements taken on m trees would produce m observations. The final result therefore is an $m \times n$ data table or matrix of the form shown in Figure 5.1.

Here each row represents the full series of measurements for one observation, or one tree, and x_{ji} is the ith measurement on the jth tree. Such a data table is shown in Figure 3.1, slightly modified because it is not possible to fit all 42 variables on one line (37 numeric plus five non-numeric variables). Consequently, each set of five lines represents the full series of measurements for one observation. The SAS program reads each group of five lines as a single line and it is only the necessity of fitting the data into the prescribed file format in which the record length is 80 spaces, thus allowing no more than 15 variables per line,[1] which gives it this particular appearance. We have to regard each five lines as being one notional line of data entry.

To return, however, to the trees, we have for each tree in the study a set of measurements on n variables. For the jth tree it will be of the form:

$$x_{j1}, x_{j2}, x_{j3}, ..., x_{ji} ..., x_{jn}.$$

It is worth noting at this stage that these variables are not necessarily all in the same metric—some may be in gms, others in gm cm^{-3}, others in cm^2, and so on. It may be that we find, or that MVA requires us to find, a combination of all these variables which produces a number which is then seen to be very

x_{11}	x_{12}	x_{13}	$\cdots$	x_{1i}	$\cdots$	x_{1n}
x_{21}	x_{22}	x_{23}	$\cdots$	x_{2i}	$\cdots$	x_{2n}
x_{31}	x_{32}	x_{33}	$\cdots$	x_{3i}	$\cdots$	x_{3n}
$\cdots$						
x_{j1}	x_{j2}	x_{j3}	$\cdots$	x_{ji}	$\cdots$	x_{jn}
$\cdots$						
x_{m1}	x_{m2}	x_{m3}	$\cdots$	x_{mi}	$\cdots$	x_{mn}

FIG. 5.1.   Typical MVA data table

[1] This depends on how many decimal places are required for each variable. I have recorded the percentages used to one place of decimals only.

useful as a discriminator of the various species. If we should choose to add all these variables together, are we in fact justified in using such disparate quantities in this way? For in normal circumstances it is meaningless to add together density and volume or area and weight, for they are distinct entities and may only be added directly to their like. Apples and pears may not be added together to give pearapples, but they will always remain separate entities. So, does the expression

$$x_1 + x_2 + x_3 ..., x_n$$

have any meaning?

To answer this it is necessary to return to a consideration of the purpose for which the measurements first were taken. For we did not measure the heartwood density of all these trees, or any of the other quantities, out of the mere interest to know what it would be in each case, but because we supposed that these quantities were all indicators, sometimes clear, sometimes obscure, of that more elusive entity which is the object of our search, namely generic affinity. If some obvious measurable characteristic were attached to each tree which gave unequivocally an indication of the species to which each belonged, then, clearly, we would look no further and not trouble ourselves with additional measurements of more and more features which could not in any way improve on the knowledge obtained by consideration of the value of the first characteristic. Unfortunately (or fortunately), nature is not quite so simple and the concepts which most interest us are often those which are not accessible by direct measurement. We have the choice of taking measurements of features which are related, possibly only tenuously, to the concept which we are seeking to define, or abandoning the search entirely.

Consequently, it seems valid to regard these measurements not as the physical quantities which they purport to be, but as pure numbers which we may manipulate to help us to define that more elusive concept, be it type, or species, or genre, or style, or whatever it is that happens to be the object of our search.

However, the question of the optimum use of the data must be carefully defined because in different situations that which is best has different meanings and the words themselves—'best', 'optimum', 'most advantageous'—have different interpretations. The purpose of the enquiry must be scrupulously defined. For example, if the intention is to classify subjects according to species it is essential to decide at the outset whether or not the characteristics of each species are to be predefined, so that observations which fall close to the predefined parameters, or within a certain range, may then be classified in the relevant species. Or perhaps the question is to be left more open-ended and the concept of species is to emerge from the investigation, if the observations are found to fall within well-separated clusters. Thus oak, ash, elm, or whatever, may be species for which it is possible to state a

clear and unambiguous series of typical characteristics which will enable us to identify and classify each individual and to resolve doubtful cases of attribution. On the other hand, it may be more important to attempt to find what the distinguishing characteristics are from a prior knowledge of the groups to which each observation belongs. Or we may suspect an underlying, more simple structure which would explain the data more satisfactorily, and we would wish to test various hypotheses about this structure.

All these situations would require a particular application of MVA, one which was designed to fulfil the function which the investigator has in mind. The main types of MVA which are available in most main-frame computer statistics packages are multiple regression, principal component, factor, discriminant, and canonical correlation analyses, as well as multivariate analysis of variance and covariance (Manova and Canova). The forms of analysis relevant to the stylometric enquiries of this work will be explained later in the chapter.

On first acquaintance with MVA it gives the impression that its use of data is somewhat arbitrary and high-handed. For the layman's approach to a series of measurements is to treat them as sacrosanct and not to be tampered with. Hence, it is important to realize that any use of data, however innocent it may seem, will inevitably involve arbitrary judgements and the use of techniques which are difficult to justify from first principles. This is so because measurements themselves are not based on absolute scales but on those which time and practice have shown to be convenient and, in any situation in which more than one variable is used, a decision is necessary with regard to the relative importance of each variable in the series. In the hypothetical case of the classification of trees, for example, we might decide that the physical characteristics of the seeds are of much greater importance for our purposes than any measurements which can be taken of the foliage or the wood of the trees. Consequently, we might choose to increase the weight attached to such characteristics by multiplying them by a certain factor, or by ignoring all the other data, thus effectively giving it a zero rating. But, whatever approach is used, it is likely to be based on hunches and instinct rather than on any strictly rational criteria.[2]

Lutoslawski in his study of the Platonic corpus investigated 500 different stylistic features and then separated them into four groups according to the importance he attached to each. The second group was given twice the weight of the first, the third group three times, the fourth four times that of the first. This enabled him to calculate, for each work, an affinity factor, as he termed it, which gave an indication of the closeness of that particular work to the

---

[2] This is true of purely human intervention, to suppress or emphasize certain characteristics of measured data. MVA is strictly mathematical and only optimizes the data according to carefully pre-defined requirements.

Laws, the latter being supposedly the last of Plato's works. The affinity factor (*A*) thus took the form

$$A = x_1 + 2x_2 + 3x_3 + 4x_4.$$

But if at the outset anyone had claimed that style could be quantified as the sum of all of a number of pre-specified accidental peculiarities, plus twice the occurrence of $\pi\acute{\epsilon}\rho\iota$ after the substantive, plus three times the presence of $\pi\tilde{a}s$ and its compounds, plus four times the absence of hiatus, we would have suspected him of insanity.

Of course, this is a parody of Lutoslawski's position, and he would probably have pleaded common sense as the basis for his approach, since he clearly suspected that there was a better way of using the raw data than simply adding it all together. Why not after all subtract alternate measurements and then sum the squares of the result, or use some other equally arbitrary method? In fact what Lutoslawski did without realizing it was to calculate a sort of discriminant function which enabled him to date, approximately, each work to a certain period of Plato's authorship. The criteria he used were arbitrary and he could with equal validity have used a different set of coefficients for calculating *A*, but the set he lighted upon seemed to work reasonably well. There is an infinite number of possible coefficients for x_1, x_2, x_3, and x_4, and how can we possibly know which is the best combination?

This is precisely the sort of problem to which MVA addresses itself. For, given a set of data about trees, or style, or fruit flies, or the intellectual abilities of children, it is reasonable to postulate that there is some unique combination of the variables within this data which will most effectively perform the task demanded of it, be it to classify the samples into the correct groups, or to find a smaller set of variables which, nevertheless, contains most of the information of the original set or whatever else it is that we require. Provided these requirements are stated with sufficient mathematical precision then MVA is capable of calculating that unique set of coefficients which, when applied to the variables, will produce that combination of them which performs most effectively the task requested of it.

A certain vagueness is necessary in these generalizations about MVA, but it is hoped that this will be dispelled by a perusal of the description of each type, where it will be possible to give more specific information about the mathematical and statistical criteria employed.

To summarize, therefore, it may be said that the problem of stylometric analysis is essentially one of classification and that such a problem is not unknown in other sciences such as biology, psychiatry, and various branches of sociology. In such areas the statistical techniques of MVA are frequently employed and it seems sensible to apply similar methods to stylometry, especially as we have identified a set of variables which it would be

convenient to measure and which are likely to contain information which is valid as a numerical description of style. It is not necessary to know in advance what is the relationship of each variable to the elusive concept which we entertain of style, or, indeed, whether or not it relates to it at all, although, in the latter case, a series of measurements containing information which is largely irrelevant to style is going to make the task of analysis very difficult if not impossible. The supposition is that most of these variables will contribute something to our knowledge of style and we may draw the parallel between this and measurements of intelligence. IQ, for example, is measured by the sum of scores on a series of tests, but it is too often forgotten that there is no absolute measure of intelligence, that it is only a concept which is found useful in descriptions of the way in which the mind works and in measurements of its performance. The tests are, of course, chosen with a view to the stimulation of responses which require the use of intelligence and in which intelligence has some role, but it is a mistake to identify the test with the concept, for of that each one is only a partial portrayal.

Similarly, with style, the variables which we are to measure are not themselves style, although it is presumed that they are related to it, even though the relationship may be tenuous, and it is hoped that the data will contain sufficient information to provide a comprehensive description of style which is both mathematically sound and linguistically probable.

One should not, of course, attempt to dodge the question of what is meant by style or how closely stylometry may hope to define it. In fact style, as far as the stylometrician is concerned, is no more nor less than the observed values to be found in the variables which he or she chooses to measure, provided these values are found to be distinctive or to show evidence of regularity. If it were possible to measure every feature, nuance, and variation of language, then I suppose such information would represent a comprehensive definition of style, but it would be so cumbersome as to be unmanageable. In the interests of clarity detail has to be sacrificed, and it is obviously better if one can reduce the number of variables which the analysis requires, since ultimately one would hope to be able to relate the distinctive measurements which emerge and which serve to isolate one work or one author from another to some concrete and recognizable characteristics of language.

The aim of MVA is to maximize the value of the information provided so that it can perform most effectively the task prescribed. In the case of style the task is mainly one of classification, but there may also be occasion to use principal component analysis as a data reduction technique. Factor analysis may be used to study the underlying structure of the data, but the main emphasis will be on the techniques of discriminant analysis and canonical correlation analysis used to separate the samples into the appropriate groups. These techniques are described more fully in the sections which follow.

Cluster analysis

As no mention of this has been made above it may seem strange to start with this branch of MVA. However, it is implicit in any attempt at classification that the observations will fall into certain groups or clusters. The object of cluster analysis is to discover what these clusters are without making too many assumptions about the distribution of the variables which contribute to the data matrix. These may or may not be multivariate normal, but, as with so many other branches of statistical analysis, the multivariate normal distribution does provide a mathematical model which is more tractable than those which underlie distributions which are skew or otherwise distorted from the symmetric curve of the normal. Such refinements, however, must necessarily be left to the mathematician, and we will confine ourselves here to a general description of some of the ideas involved.

Cluster analysis is a very general technique which is used to cluster objects according to their similarity or to separate them according to their dissimilarity. This perhaps appears to beg the question because it implies that we already know what these concepts mean (i.e. similarity and dissimilarity), whereas a moment's reflection will convince us that differences of any sort are notoriously difficult to quantify.

It is helpful in this context to consider two ways in which the stylistic data may possibly be visualized. Each observation will consist of a series of measurements, x_1, x_2, etc. which may be represented graphically as a humpy line joining a series of peaks, as in Figure 5.2. This line may be thought of as the stylistic profile of that particular sample, and the problem is that of comparing it with the profile of all the other samples in the study. A possible
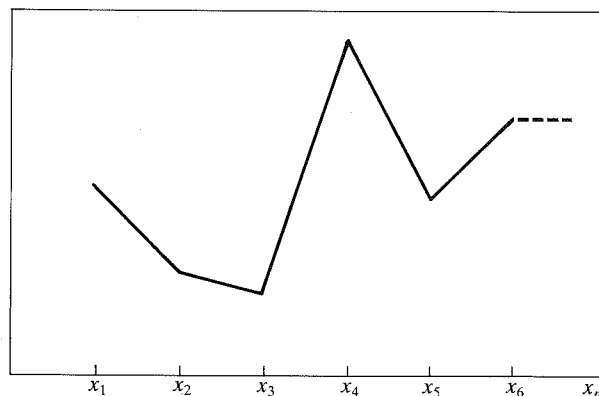

FIG. 5.2   Typical graph of multivariate observation

measure of dissimilarity would be the sum of squares of the differences between all the points x_1, x_2, etc. on the two profiles which we were comparing. Or it may be desirable to measure the differences in slope of comparable segments of the line and then sum these differences. From the point of view of stylometry it is hoped that families of profiles will emerge corresponding to the individual works or to the authors and that the methods of cluster analysis which SAS provides will prove capable of grouping these families correctly. If our hypotheses about the variables are correct, then the measurements taken should provide a characteristic profile for each author which should be recognizable in most contexts.

Another way of visualizing the data is to regard each observation as being a vector of style, say S, where

$$S = x_1, x_2, x_3, ..., x_n.$$

A vector represents the mathematical notion of positioning a point in space. In familiar three-dimensional space the three Cartesian coordinates x, y, z are sufficient to specify the position of any point uniquely. If we extend the idea to more than three dimensions it is possible to conceive of the vector S as positioning the sample of writing at a particular point in n-dimensional hyperspace and the coordinates of the position will be defined by x_1, x_2, x_3, ..., x_n. Samples which are similar in style will cluster around the same point in hyperspace, and by dividing this space into appropriate sections it should prove possible to isolate the clusters.

The SAS cluster procedure provides a choice of three agglomerative, hierarchical clustering techniques.[3] This means that they start from the position in which all the samples are in separate clusters and the two samples which are nearest are then joined to form a cluster. The process is then repeated with the two nearest clusters always being joined together until the position is reached where all the samples belong to one cluster. The adjective 'hierarchical' implies that none of the clusters overlap, so that an observation cannot belong to two clusters at the same time.

The three clustering techniques available with SAS are Ward's method, the centroid method, and average linkage over squared Euclidean distances. With Ward's method distance is computed as the sum of squares between the two clusters added up over all the variables. It is biased towards joining clusters with a small number of observations and tends to produce clusters of approximately equal size. The centroid method defines cluster distance as the Euclidean distance between their centroids or means. It is more robust to outliers than the other methods, but is less good at finding sharp divisions between the clusters. Average linkage computes distance as the average squared Euclidean distance between pairs of observations (one in each

[3] SAS Institute Inc., *SAS User's Guide: Statistics* (1982), p. 147.

cluster). It tends to join clusters with small variances and is biased towards producing clusters which have approximately the same variance.

Most clustering techniques have a tendency to assume that clusters will have an approximately spherical shape. Consequently, they might fail to detect clusters of unusual shape even though these might be distinctly defined. The human eye can often see patterns which are not so readily definable that computer programs may be written which will pick them all out. Widening the scope of a program will usually reduce its sensitivity so that all clustering algorithms have to compromise between competing claims.

However, it is not necessary to concern ourselves unduly with the mathematical processes of clustering. The methods will be used in the preliminary stages of the investigation to see whether or not clusters may be found which correspond roughly with our prior knowledge of the data. If that is found to be the case and all three methods produce approximately similar clusters, then that will be taken as a positive indication that the enquiry is proceeding along the correct lines and that the variables do indeed give information which is relevant to the concept of style. Otherwise, other methods will have to be tried, and it will be necessary to examine the data more closely to try to ascertain why it should fail to give the required information.

Factor analysis (FA) and principal component analysis (PCA)

Although I do not use these techniques in the subsequent investigation, I describe them briefly here since they are frequently encountered in studies within the social and behavioural sciences, and most textbooks of MVA devote large sections to them.[4] Factor analysis attempts to reduce the original set of variables to just a few factors which then replace the first set of variables. These factors explain the phenomena under investigation more effectively or more cogently and can often be related directly to the physical basis or the theoretical framework of these phenomena. For example, the growth rates of particular plants could conceivably be related to two or three factors, such as soil type, climatological conditions, and level of applied nutrients, even though the experimental measurements consisted of 20 or 30 variables recording disparate information about the plants' environment. In this way the data will have been compressed and probably rendered more intelligible, since it is easier to understand the interrelationships between two or three factors than it is to grasp the complex interactions between 20 or 30 variables which affect plant life.

[4] S. Bennett and D. Bowers, *An Introduction to Multivariate Techniques for Social and Behavioural Sciences* (London, 1976) devotes six chapters to various forms of FA and PCA, and only one to discriminant analysis.

Principal component analysis is also a data reduction technique, but one which is less flexible than factor analysis. It seeks to reduce the multiplicity of original variables to a few which contain almost as much information as the original set. It does this by combining them in such a way that for the first PC the maximum amount of information is extracted, and so on for all those which follow.[5] Thus the first few may contain nearly as much information as all the original variables. Typically the first five PCs might account for 80% or more of the information contained in the original data.

However, I have not employed these techniques in the following chapters since it was found that other methods were better adapted to dealing with the classification problems that lay at the root of the enquiry.

Discriminant analysis

In the case of a univariate study, if we were comparing two or more groups of samples on the basis of one measurement, for example the percentage rate of occurrence of $\tau\epsilon$, it would be appropriate to use the statistical technique developed by R. A. Fisher in the 1930s and known as Anova (analysis of variance).[6] Basically this consists of comparing the variance which is found to occur between the groups with that which occurs within them. The ratio of these two quantities is known as the F-ratio, and if it exceeds a certain value (which is determined by the group size, the number of groups, and the level of significance chosen), then the samples are deemed not to have been drawn from the same population.

The corresponding multivariate case is based on the same approach, but instead of the single variable it uses a combination of all the variables of the form

$$W = a_1x_1 + a_2x_2 + a_3x_3 + ..., + a_nx_n.$$

The objective of multivariate discriminant analysis is to maximize the F-ratio, that is, to maximize the ratio of between-group to within-group variance, and the calculation of the coefficients a_1, a_2, etc. is based on this objective. The unique set of coefficients which gives the maximum value of F is calculated and samples are then classed according to their score on this composite variable.

[5] I have simplified here and used the term 'information' loosely. Technically it is the variance which is maximized for each PC, each succeeding one in the calculation being orthogonal to those which precede. There are as many PCs as original variables. See e.g. Bennett and Bowers (n. 4 above); R. J. Harris, *A Primer of Multivariate Statistics* (London, 1975); F. H. C. Marriott, *The Interpretation of Multiple Observations* (London, 1974); A. E. Maxwell, *Multivariate Analysis in Behavioural Research* (London, 1977).

[6] See any textbook of basic statistics, e.g. H. T. Hayslett, *Statistics Made Simple* (London, 1974), ch. 12. Kenny illustrates the use of Anova on literary tests in ch. 10 of *The Computation of Style*.

Using the SAS Discrim option[7] it is possible to calculate the discriminant functions for a specially chosen set of samples in which we wish, for example, to distinguish between different authors or different periods of an author's life, and then apply this information to a set of test data in order to determine the provenance of the samples being tested. The success of this method depends on the general validity of the discriminant functions calculated in the first instance. If the samples chosen are not representative of the authors whose characteristics we are attempting to determine, then the discriminant functions also are likely to be unrepresentative and perform poorly when applied to test data. In addition, it should be remembered that in cases of uncertain authorship the correct classification will depend on whether or not samples of the putative author are available for comparison. A sample will be classed in the group to the mean of which its discriminant function comes closest, but it may in fact be considerably distant from all the group means. Sometimes the sample will be classed as 'other', indicating that it does not meet the threshold probability for classification in any one group, but in other instances an accidentally close resemblance may result in an unexpected classification.

More light will be shed on these problems when concrete examples are considered later in the study. In general it is perhaps true to say that discriminant analysis is better adapted to confirmatory classification, especially where the field is well defined a priori, but, in more open-ended situations where any one of a large number of possible origins may be claimed with equal probability for a given sample and background information about these origins is sparse, discriminant analysis may not be very helpful.

Canonical correlation analysis

This type of analysis was originally developed as a means of studying relationships between different sets of variables. In general it poses the question 'What linear combination of the first set of variables produces the maximum multiple correlation with a linear combination of the second set of variables?' Group membership may, however, be considered as a variable, and in this way a series of canonical variables may be calculated which gives maximum correlation with the pre-defined allocation in groups. Canonical correlation analysis thus used becomes a form of classification analysis which discriminates between samples according to the group to which they belong, and for this reason the SAS Candisc option is so called.[8] However, it differs from discriminant analysis in that the former maximizes the value of the F-ratio while canonical discriminant analysis maximizes the value of R^2, the

---

[7] *SAS User's Guide*, p. 381.   [8] *SAS User's Guide*, p. 369.

multiple correlation coefficient. There is a unique set of coefficients defining the first canonical variable, C_1, of the form

$$C_1 = a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + ..., a_{1n}x_n$$

such that R^2 is a maximum. When C_1 has been calculated it is possible to calculate a second canonical variable, C_2, subject to the condition that it will be uncorrelated with the first, but which again gives the maximum value of R^2. And so on for each successive canonical variable. The number of possible canonical variables is one less than the number of groups, or equal to the number of original variables, whichever is the smaller. Each C_n is uncorrelated with all the others and has a higher value of R^2 than all those which succeed it in the series.

In many cases it is possible to reduce a complex data set to a few canonical variables which will be found to summarize most of the information found in the original variables. For example, a Candisc run on 475 samples of Plato derived from 35 different works gave an R^2 for the first three canonical variables of 0.90, 0.63, and 0.58 respectively. The figures indicate that these variables account for 90%, 63%, and 58% of the variance of the original set of 37 variables. Effectively the original variables could be replaced by this new set of canonical variables since they tell us almost as much cumulatively about the separation of the samples into works. A study of even the first canonical variable, CAN1, may on its own be very informative, especially when R^2 is as high as in this example.

Canonical correlation analysis will be used extensively in the following chapters. It will be explained more fully on the occasions when it is used and its versatility as a tool for the investigation of literary phenomena should become apparent.

Conclusion

In the foregoing summaries the mathematical content has been kept to a minimum. This is necessitated by the nature of the material, which relies heavily on matrix algebra and the calculus of matrices. Principal component analysis, for example, requires the calculation of the latent roots and vectors of the variance/covariance matrix of original variables. Such matters, apart from being inappropriate within a work which is dedicated mainly to Classics, are beyond my competence to expound satisfactorily. Besides, there are already many excellent books available which deal with the subject (see note 5).

However, anyone intending to embark on a project which makes use of MVA in stylometric research would be well advised to acquire as much knowledge as possible of the underlying mathematics, if for no other reason than that of being able to respond intelligently to computer output.

Another point worth mentioning at this stage is that the main impetus of the study will be in the direction of descriptive statistics. This may seem cowardly to those statisticians who have decided that inferential statistics and the finding of significant differences is the chief purpose of statistical enquiry. In fact, as will emerge subsequently, significant differences between works are not hard to find and the problem is not the discovery of the difference, but its interpretation. For it is rarely possible to make the equation 'difference of population equals difference of authorship'. Without more knowledge of how much authors vary and how much genre and other similarities, possibly even direct imitation, may mask the differences due to authorship, it is safer to concentrate on similarities rather than differences and to devise methods of illustrating, visually or otherwise, how close some works are to each other and the degree of family relationship which exists between certain groups of works. The advantage which canonical correlation analysis and discriminant analysis have over the less rigorous techniques of cluster analysis or principal component analysis is that they offer a range of significance tests which are available if we should need them. This does not mean that such tests have to be employed and in many cases it will be found that the differences uncovered are so well defined, irrespective of the technique employed, that significance testing in the circumstances would be superfluous.

It should become obvious, as the work proceeds, that decisions concerning the choice of analysis were made on an ad hoc basis, depending often on the direction of preliminary results, and that since the application of SAS to the data, once that had been obtained, was a comparatively simple procedure, many different analyses were conducted on a large number of subsets of the data, often with only a few slight modifications of the parameters of the enquiry, such as the inclusion of an additional principal component as a variable in a cluster analysis, or the removal of samples of a work the inclusion of which seemed to muddy an otherwise clear result. The purpose of such juggling has been to elucidate as much information as possible from the data, and it is hoped that the succeeding chapters will show that this objective has been achieved.

6
Preliminary Survey

WE are now in the position of having to decide how to make the best use of the data available. Originally only 320 1000-word samples taken from the seven authors were subjected to analysis. Various tests were done on these samples, of which only a small selection may be given here in order to illustrate the methods and to demonstrate their efficacy. Space and economy do not permit the inclusion of a more extensive survey.

Strictly speaking the word 'sample' as used here is misleading because each work was included almost in its entirety, as explained in Chapter 3. The samples were 1000-word sections of text taken in sequence, beginning with the first word and continuing until the end of the work was reached. Thus the longer the work the greater the number of samples which were obtained from it.

From the outset Plato tended to be more heavily represented than the other authors. This was intentional since the study was orientated towards Plato and the attempted solution of some of the problems connected with authenticity and the chronology of his works. Hence, a greater effort has been made to include Plato in the analyses at an early stage and to devote more time to collecting data from his works rather than from other authors. At a later stage (Chapter 8) the entire Platonic corpus is included, giving a total of 493 samples for Plato compared with an overall total of 209 samples for the remaining six authors.

This does suggest a certain amount of imbalance in favour of Plato, but I have attempted in the early stages to maintain a reasonable equilibrium between the contributions made by various authors. In the case of Lysias the number of samples is low because only one speech was used from the Corpus Lysiacum, *Against Eratosthenes*, since this has the greatest claim to authenticity.[1] Throughout the entire study these five samples only were used as being representative of Lysias. To have included other speeches might have put too much strain on the concept of authorship, since there could be no guarantee that any other speech or speeches which I might have chosen from the Corpus Lysiacum would necessarily be by the same author.

This does not of course preclude a subsequent study being made of the

---

[1] It was reputedly delivered by Lysias himself. In this case, however, since we are using only the one speech, the name of the author does not matter. All that is required is that the whole of it was written by one individual, otherwise our assumptions of homogeneity may not hold.

Lysias speeches to see precisely where the affinities between them lie, but I was more concerned with restricting the number of unknowns at this early stage of the enquiry, rather than increasing them and thereby being forced to deal inadequately with a large number of problems simultaneously.

The same arguments also apply to Isaeus, where we have extant only eleven orations attributed to him plus various fragments. All of them which were of sufficient length were included, giving a total of 23 samples. I am well aware of the danger of assuming, without further proof, that these speeches are all by the same author, and in most cases I include only one or at most two of them in each analysis. This is a reasonably safe approach, especially as there is a unity of subject matter in the Isaean speeches—since they all deal with testamentary disputes—which is lacking from the Corpus Lysiacum, where the range of subjects and styles exhibits far greater variety. However, I have tried to bear in mind the fact that the imposition of a single author on the Isaean corpus may in some instances cause a distortion of the results if the underlying assumption (of Isaeus being the author of all of them) is in reality untrue.

For Aeschines the number of samples was artificially restricted to 12, since, at the time, *Against Timarchus* was the only work of this author available in machine-readable form from TLG.

The other three authors have been treated, I hope, with even-handed justice, and I have selected works, or parts of works, which are reasonably representative of their output. This gives a total of 55 for Isocrates, 49 for Thucydides, and 65 for Xenophon. A full list of the works included for each author is given in Table 6.1.

It should be apparent, therefore, that a sizeable amount of Greek is being investigated, 702 000 words from seven authors whose works approximately span the century 420–320 BC. In this preliminary survey selections of the works have been used, for the full range of works and samples is not employed until the study of authenticity of the Platonic corpus is commenced in Chapter 8. It should be remembered that, for all of the 702 samples, readings have been taken on a series of 37 orthographic variables as described in Chapter 2.

It will probably be most helpful if, in what follows, these 37 variables are visualized as forming a profile for each sample. It is to be expected, because of the innate variation and choppiness of language, that all of these profiles will be unique, since it is highly unlikely that the readings for 37 variables on any two samples would be identical. Nevertheless, we anticipate that there will be family resemblances, whether caused by concurrence of author, or subject, or genre, or other unforeseen causes. The problem, therefore, reduces to one of identifying profiles which resemble one another and defining similarity in a mathematical sense.

There are various mathematical models available for measuring similarity

TABLE 6.1   *Works included in the study*

Author (abbreviation)	Work (abbreviation)	No. of samples
Aeschines (Aeschin.)	*Against Timarchus (Tma.)*	12
Isaeus (Is.)	*Orations (Oras.)* 1–3, 5–9, 11	23
Isocrates (Isoc.)	*Antidosis (Ant.)*	17
	Archidamus (Archd.)	6
	De Pace (Pac.)	7
	Panathenaicus (Panth.)	15
	Panegyricus (Pan.)	10
Lysias (Lys.)	*Against Eratosthenes (Era.)*	5
Plato (Pl.)	All works longer than 1000 words*	493
Thucydides (Th.)	*History (His.)*,	
	bk. 3	16
	bk. 4	20
	bk. 5	13
Xenophon (Xen.)	*Hellenica (HG)*, bk. 1	20
	Memorabilia (Mem.)	30
	Oeconomicus (Oec.)	15

* A full list of Plato's works included in the study is given in ch. 8.

or its converse.[2] One simple approach is to sum the difference of scores on each variable for any two samples, giving a grand total of 37 individual differences which reduce to one single figure. This would be expressed mathematically as

$$D = \sum_{n=1}^{n=37} (\text{LET}_{nA} - \text{LET}_{nB})$$

where D is the distance (or difference) between the two samples, LET_{nA} is the nth variable of sample A and similarly for LET_{nB}.

In practice it is best to square the individual score differences so as to eliminate negative quantities, otherwise there are circumstances in which two totally different samples could end up with a zero total-score difference, indicating that they were identical (a dissimilarity score of zero is conventionally taken as describing complete likeness). In such cases the negative scores would have completely cancelled out the positive ones. By squaring these differences we are left only with positive scores and the final figure is a squared distance between two samples. The equation then becomes

$$D^2 = \sum_{n=1}^{n=37} (\text{LET}_{nA} - \text{LET}_{nB})^2$$

[2] See A. D. Gordon, *Classification* (London, 1981).

This is known as the Euclidean squared distance between the two samples, because it has obvious connections with geometry, although 'Pythagorean' might be a more appropriate title as the connection is more directly with the theorem of Pythagoras.[3]

This, therefore, is a basis for measuring dissimilarity. It is not the only method available, nor is it necessarily the best in the circumstances. We shall also be encountering the Mahanalobis distance,[4] especially in connection with discriminant analysis which is used later in the chapter. In this case the distance measurement attempts to take account of the fact that many of the variables correlate with each other, or, to put it another way, they are duplicating information about the samples. By using raw scores as the basis for our distance measurements, as in the Euclidean distance, we may be simply repeating the same information over and over again, for the variation in one variable, say the one which measures the percentage of words containing a, may be perfectly in step with any one of the remaining 36 and therefore either could be considered to be superfluous. If the standardized scores on two variables are identical from sample to sample, then their correlation is unity and we are entitled to assume that it is sufficient to use only one of them, since it is a reasonable guess that they are both measuring the same underlying phenomenon.

The Mahanalobis distance takes account of all the correlations between the variables and does not attach equal weight to all the scores. It is expressed by the formula

$$D^2 I{:}J = (\bar{X}_I - \bar{X}_J)' S^{-1} (\bar{X}_I - \bar{X}_J)$$

where the capitals D, S, and X stand for matrices.

The above formula is for a generalized squared distance between two groups I and J, each group consisting of a number of samples. Thus the groups in our case could consist of works and we could find the generalized squared distance between the *Republic* and *Protagoras* for example, or between one of these and Thucydides' *History*. Equally, with a slight variation of the above formula we can find the distance between any one

---

[3] In two-dimensional geometry the distance between two points A and B of coordinates $(x_1\ y_1)$, $(x_2\ y_2)$, may be calculated using Pythagoras' theorem. It is given by the equation $AB^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2$. This corresponds to the equation given for D^2: $D^2 = \sum (\text{LET}_{nA} - \text{LET}_{nB})^2$ for in the two-dimensional case there is only one summation of xs and ys. $(\text{LET}_{nA} - \text{LET}_{nB})$ is therefore the equivalent of $(x_1 - x_2)$. In geometrical terms the final positions for A and B in the multivariate case (A and B are two samples) are defined by 37 coordinates in each case. The distance which separates them corresponds to a distance in multidimensional space. Obviously one cannot visualize a 37-dimensional space, or hyperspace, but there is no reason why the mathematical approach should not be the same in principle as for the two- or three-dimensional space which we can visualize.

[4] Named after an Indian statistician, P. C. Mahanalobis. See references in D. F. Morrison, *Multivariate Statistical Analysis*, 2nd edn. (New York, 1976).

---

sample and any of the groups in the analysis, or between individual samples.[5]

S in the above formula stands for the variance/covariance matrix of all the variables and $\bar{X}_I$ is the matrix of mean scores for all the variables for the samples of group I and so on. From this it may be seen that the mathematics involved is rather advanced, but fortunately it is not necessary to become too much concerned with the details. A minimum understanding of what lies behind any approach that the various packages of multivariate statistics might offer is, however, very helpful, enabling one to avoid the hazard of misinterpretation of the results.

The fact that there are various ways of defining similarity and distance should, therefore, be borne in mind.

The simplest approach to the problem of determining whether any family resemblance is to be found among these 320 samples is to run a cluster analysis on all of them. This will effectively group the samples according to some predefined notion of similarity. In the case of SAS cluster analysis three options are available, Ward's method, average linkage, and centroid[6] (see Chapter 5). We need not concern ourselves too deeply with the mathematical differences between these methods, for the fact that substantially similar results are obtained whatever method is used is sufficient to guarantee that the cause is not some chance coincidence of the method used and a particular combination of variables, but relates to something more fundamental and underlying, something which is linked directly to the structure of the language which each author or work employs.

But before proceeding further it is necessary to supply a little background information to explain my approach to the methods used. In the first place I must stress that what I present in the following pages is only an edited and severely restricted version of the findings gathered from the various avenues which I explored in attempting to assess the validity of using variables of this type to determine author characteristics or to solve problems of chronological variation. For any selection of the 320 samples it would be possible to use a cluster analysis, a neighbour analysis, a discriminant or canonical correlation analysis to determine the groups, in each case using either all 37 variables or any subset of them selected at random or with some special purpose in mind.

The vast amount of output from each run of analysis on each set of data makes compression and précis of the results vital if the work is not to be swamped under a mass of detail. Even a straightforward cluster analysis

---

[5] Neighbour analysis will calculate the distance between any two samples based on either Euclidean or Mahanalobis distance scales. See *SAS User's Guide* under 'neighbour analysis'.

[6] This applies to the 1982 version of SAS. The most recent version offers 11 different clustering techniques. I have tested all these on various typical groups of samples and none offers significantly better results than I had previously obtained with Ward's method or average linkage.

produces two to three pages of printout for 50 samples and more if greater numbers of samples are used. Discriminant analysis and Candisc produce typically 40 or 50 pages, all of which contain information which is relevant to some aspects of the enquiry. I therefore summarize in two ways, firstly by selecting only a very limited number of cases of the many possible combinations of samples, variables, and types of analysis and, secondly, by reproducing only a fraction of the computer output which was available in each instance. This applies also to Chapters 8 and 9, which deal with authenticity and chronology, for the classification details alone contained in the discriminant analysis often run to 80 pages of computer printout when the number of samples is large.

In retrospect it is easy to see that any cluster analysis on a subset of samples which contained one work from each author would have been all that was necessary to demonstrate that the type of variable measured contains a great deal of the sort of information which we are looking for, namely information which helps to separate author from author or work from work. For it becomes clear immediately that authorship characteristics seem to dominate in many cases and result in clustering of samples from each work with the parent group. In many cases also it will be found that, where more than one work is included for some of the authors, the proximity of these works to each other usually reveals a bias in favour of authorship clustering.

However, this is to anticipate subsequent arguments and evidence which will be introduced later in the chapter. From the above it may be seen that it is necessary to familiarize ourselves with cluster analysis.[7] This is best approached by means of an example, and Figure 6.1 shows a tree diagram obtained by using a cluster analysis on 55 samples. Not all the samples have been used for each work, as the objective here has been to restrict the tree diagram in order to be able to read it more easily. The works used, together with the number of samples, are listed in Table 6.2.

As explained in Chapter 5 a cluster analysis groups samples according to some predefined concept of similarity. The process is repetitive and may be envisaged as operating in one of two ways, as a series either of continuous divisions, or of continuous accretions in which clusters repeatedly merge. Thus we may consider the process as commencing from the point at which all the samples are contained in one large cluster and divisions are repeatedly made between groups wherever the linkage is least strong, until the point is reached in which every sample is contained in a cluster on its own. Or, conversely, the start may be made with each sample forming an individual cluster and mergers are than made in successive steps between those clusters which show the greatest affinity, until finally all are contained in a single, all-
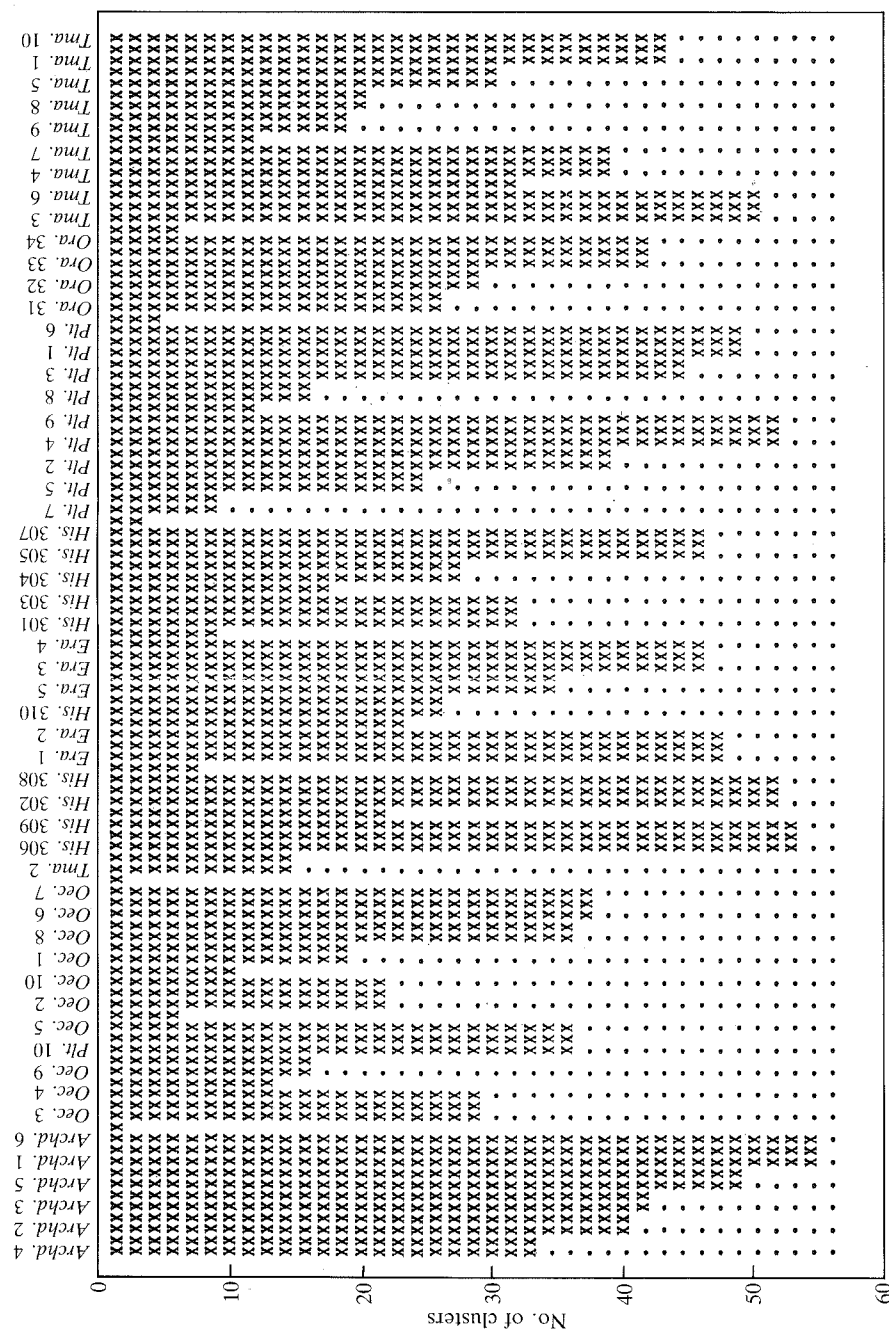
[7] *SAS User's Guide* under 'cluster'.

FIG. 6.1   Ward's minimum variance cluster analysis on 55 samples

TABLE 6.2   *Samples included in Fig. 6.1*

Author	Work	No. of samples
Aeschin.	*Tma.*	10 (of 12)
Is.	*Ora.* 3	4 (of 4)
Isoc.	*Archd.*	6 (of 6)
Lys.	*Era.*	5 (of 5)
Pl.	*Plt.*	10 (of 16)
Th.	*His.*	10 (of 16)
Xen.	*Oec.*	10 (of 15)
TOTAL		55

embracing cluster. The point is largely academic, however, since we are concerned mainly with the end-product rather than how it is obtained, although it does help to show us that certain questions which we might be tempted to ask about clustering are not necessarily relevant.

For example, a general enquiry such as 'How many clusters do these samples form?' is largely meaningless because the answer depends on where one chooses to interrupt the clustering algorithm. It could be any number between 1 and n, n being the total number of samples. The question 'Is there an optimum number of clusters?' is more meaningful as, for some types of cluster analysis (Ward's and average linkage), SAS provides a figure called the 'cubic clustering criterion' for each level of clustering, the point at which this reaches its maximum value being an indicator of the optimum number of clusters.

The tree diagram[8] of Figure 6.1 is a visual representation of the clustering history. It contains trunk, branches, and leaves, the leaf consisting of a single sample and the trunk being the top unbroken row of Xs, which corresponds to the group or cluster containing all the samples. In between are the branches which correspond to the clusters of various sizes. Above the top row of Xs are the identifying codes for each sample. Thus *Era.* 1 stands for Lysias, *Against Eratosthenes*, sample 1, *Plt.* 1 for *Politicus*, sample 1, and so on. Each sample is represented by a double vertical column of Xs, which changes to a column of dots at the point where the sample breaks off from the adjacent clusters and occupies a cluster entirely on its own. A blank column indicates that the clusters on either side of the gap are not joined, whereas unbroken rows of Xs show that all the samples covered by these Xs belong to that particular cluster until a blank space or a dot is reached.

[8] *SAS User's Guide* under 'tree'.

Thus a horizontal line drawn across the tree diagram at any point will reveal, on the left-hand axis, the number of clusters at that level and the samples which belong to each cluster. For example, in Figure 6.1 a line drawn across at the level of 10 clusters reveals the following groups, taken as they occur, working from left to right of the tree diagram:

1. *Archidamus*. All six samples. Does not begin to break until beyond the 30th generation.
2–3. *Oeconomicus*. A double cluster which contains all of the *Oeconomicus* samples and a stray one from *Politicus*. This main *Oeconomicus* cluster splits at about level 8.
4. *History*. This contains four Thucydides samples plus one from Aeschines.
5. *Lysias*. All the *Against Eratosthenes* samples are in this cluster plus a stray from Thucydides.
6. *History*. The remaining five Thucydides samples.
7. *Politicus*. One sample only, *Plt.* 7, joined on to the main *Politicus* cluster at level 9 (approximately).
8. *Politicus*. The main *Politicus* cluster.
9. *Oration* 3. Isaeus. Homogeneous to the 25th level.
10. Aeschines. *Against Timarchus*. Contains nine out of the 10 Aeschines samples.

The tree diagram therefore contains a great deal of information, and we may see at a glance how effectively the clustering history corresponds with our knowledge of the works themselves. It also gives a good general idea of the homogeneity of the samples within each individual work. A dense mass of Xs indicates excellent homogeneity within the group, whereas if the cluster is broken at an early stage by spaces or dots, the individual samples, though having an overall resemblance, display greater evidence of heterogeneity, which causes the cluster to split into individual components.

However, it is also important to remember that the clustering algorithm is repeated until all the samples form unit clusters, so that at each stage a new fragmentation *must* be produced, whatever the relative tightness or looseness of the clusters might be at that point, and this necessity inevitably results in a patchiness in some of the clusters. The picture that emerges is therefore one of relative homogeneity, for if the Isocrates cluster is found to fragment only at an advanced stage of its history, say the 30th generation, while the Lysias cluster commences to break much earlier, then we may conclude that, in relation to the works chosen for these two authors, Isocrates shows much greater homogeneity than Lysias. In general terms, however, the picture may not be the same, for it could change if different variables were used (e.g. only a subset of the 37 original variables) or if a different selection of works were used. One should, therefore, be wary of making generalized conclusions

based on the use of a small selection of works. In fact it does emerge subsequently that, in comparison with the other six authors, the samples taken from Isocrates seem to knit together much more tightly, and this is verified by many different combinations of individual works and by the use of different subsets of variables. For the present, however, we are concentrating on situations in which all 37 of the variables are included in the analysis.

Many questions arise in connection with the results shown in Figure 6.1 but we are principally concerned with establishing whether or not the use of orthographic variables of this type tells us enough about the different works and authors to enable us to use them for purposes of classification. It is clear that highly significant differences have been revealed between the various works included, for there is no possibility that the sort of clustering shown could have occurred by chance.[9] In any case, the results are confirmed by using other methods of cluster. But it is still too early to claim that these results may be ascribed entirely to authorship variation. Other sources of differentiation may well be more important, and it will obviously be necessary to extend the scope of the enquiry to include cases where there is a greater chance of confusion due to the similarity of the works chosen.

Nevertheless, the results are impressive, for a cluster analysis is entirely neutral in its approach to the data, and has no interest in producing one set of groupings rather than another. The clustering is based only on the mathematical content of the variables, the ALETs, BLETs, and CLETs, and does not depend at all on any prior knowledge that certain samples belong to certain works. Yet it has effectively grouped nearly all these samples with the correct parent cluster.

The success rate is not 100%, although precisely how one should estimate success in these circumstances is not at all clear. Obviously it would be preferable if the one stray *Politicus* sample, *Plt.* 10, which has been clustered with the *Oeconomicus* group, were lodged with the main *Politicus* cluster, at least from the point of view of authorship differentiation, and the same applies to the one Aeschines sample, *Tma.* 2, as also to the confusion between Lysias and Thucydides, a confusion which no doubt arises from the custom in ancient historical writing of interspersing narrative with speeches made by those involved in the affairs of the time. The result is that there is a certain element of genre confusion between these two authors, and it would require the use of other techniques of MVA to discriminate between them, or a more selective use of variables.

What importance are we to attach to the fact that some of the samples shown in Figure 6.1 are misaligned? Obviously we cannot draw the conclusion that parts of Thucydides' *History* were written by Lysias, or vice versa,

[9] The level of chance involved is equivalent to that of dealing a pack of cards in any predetermined order, say 10^{-68}, or effectively zero.

or that Xenophon had a hand in the *Politicus*. It is clear, however, that certain sections of these various works do, in some sense, resemble each other, at least in so far as the present variables are concerned. It implies that there is some degree of overlap in the linguistic usage of some authors. We are concerned with establishing how serious a problem this might be and discovering if it is likely to occur with such frequency as to vitiate the whole process of classification.

In fact, as the enquiry proceeds we shall find that greater difficulties of separation do occur between certain authors and certain types of work to the extent that cluster analysis is almost useless as a means of detecting the sort of stylistic affinities in which we are interested, namely those which are distinctly author-based rather than those which relate to some other cause of linguistic similarity.[10] In general, however, the results shown in Figure 6.1 are extremely encouraging and they reveal the sort of level of error which the more sophisticated techniques of multivariate analysis would have no difficulty in correcting. In almost any group of naturally occurring objects the odd abnormal samples are to be expected and their presence does not seriously hamper the process of classification. Although literary samples do not, strictly speaking, belong to this category (of naturally occurring objects), being a sort of secondary production, it is perhaps not straining reality too far to regard them as having many qualities which make them similar to natural objects. The mathematical uncertainty lies in the fact that the distribution of the variables measured is not known.

Of course, merely showing that measurable differences exist between works as revealed by the variables we have chosen to use does not in itself guarantee that we shall be able hereafter to tell the difference between authors, but it is a very important starting-point. For if no such differences had been found, the project would not have merited further study, since the basis from which any progress is to be made is that these variables contain information which is relevant to the measurement of style. Any system of measurement which purports to represent style should be capable of showing differences wherever we know that, on the level of ordinary language usage, such differences exist. Until now it was thought that elementary measurements of the type used here, measurements based entirely on orthography, would merely exhibit random noise and would be no more useful for detecting differences of style than if we had chosen to jumble the complete set of authors in a bag together and then read the words out backwards. Now, however, it appears that some impress of style is operating at this fundamental level of language, and if we can interpret the information adequately we

[10] I make this distinction because the fact that a cluster analysis reveals an affinity between any two samples does have a linguistic relevance, even though the affinity may not concur with our prior judgement. Our knowledge of the works is based on many points of reference other than those relating to purely linguistic criteria.

may find that measurements of this sort contain very much that is distinctive and are far more sensitive to stylistic variation than anything that has been attempted so far in stylometry.

As mentioned above, cluster analysis is an entirely neutral approach—it makes no assumptions about the provenance of the samples or the distribution of the variables. But there are much more powerful techniques available to the statistician. These are based on a more efficient use of the variables and a prior classification of the samples into some group or other. For it could easily be argued that cluster analysis is a rather hit-or-miss affair and there is no reason why we should use 37 variables rather than 367 or only seven. What exactly is the justification for using one set of variables rather than another? Perhaps by using so many we are swamping information under a high level of noise. And this brings us to the further consideration that perhaps we could use the information gathered far more effectively by combining the 37 variables, or a subset of them, in such a way as to maximize the relevant features which they contain, thus enabling us to discriminate with even greater success between different works, or authors, or genres, or according to whatever theme of classification the situation demanded.

It is precisely in this area that multivariate analysis is most valuable, as there are various techniques which have been developed to optimize the information contained in any given set of variables. Thus a discriminant analysis applied to the preceding group of samples (those of Figure 6.1) will produce a discriminant function which can successfully classify all the samples correctly in the work to which they belong. It is necessary initially to specify a suitable classification for each sample based on prior knowledge. This may seem to be rather a superfluous proceeding, simply asking the mathematical model to do for us what historical knowledge has already done far more effectively, for we know accurately which samples belong to the *Republic*, or to any other work, and which do not, but the point is that we do not have a mathematical basis for making the judgement and it is this that discriminant analysis is providing. If there is no mathematical basis for grouping the samples according to their known provenance, the discriminant analysis will effectively tell us so by indicating that many of the samples are misclassified. This does not, of course, imply that Plato wrote Xenophon's works or vice versa, but it does show that the variables we have chosen are not adequate to do the work required of them and we should look elsewhere for information relevant to authorship discrimination. For it may well be that the works in question contain too much variability to allow themselves to be categorized by any persistent quality which remains unchanged throughout all the samples. This, however, is precisely what the investigation is attempting to discover.

Another advantage of discriminant analysis is that it may be used to set up a discriminant function, or a series of such, derived from one set of works,

and these functions may then be applied to any number of works of doubtful or unknown provenance to see how they in turn are classified. In theory, therefore, we could build up a discriminant function for each and every author, based on known and certain attributions, and then apply it to all works of uncertain origin so as to determine their authorship. But in practice there are limitations to the method which will be discussed fully in Chapter 8, which deals with the authenticity of the Platonic canon.

Using discriminant analysis on a group of samples no greater than that of the first example, which has only 55 samples, at the same time employing all 37 variables for the task, is to a certain extent only an academic exercise, since success is almost guaranteed by using such a high ratio of variables to samples. It is equivalent to fitting a regression line to a certain number of points, perfect fit being assured if the number of points (samples) and the number of variables are equal. Even if all the samples are remarkably unlike and the variables contain hardly any information bearing on the provenance of each sample, yet with a sample: variable ratio approaching unity the correct classification would be achieved in the majority of cases. Ideally one should maintain a sample : variable ratio of at least 3 to 1 and in all the cases where it matters I have observed this rule.

The next series of examples uses an increased number of samples so as to illustrate more effectively the difficulties met with in separating works from one another and the way in which the more advanced techniques of MVA deal with the problem. Firstly, a cluster analysis is run on the data to gain a general idea of the affinities which exist between the various works and to obtain a general view of the difficulties which might lie ahead in the attempt to identify authorship characteristics. The tree diagram derived from this cluster analysis is shown in Figure 6.2.

As may be seen by a glance at this tree diagram, the same works are included as for the previous example, but the full number of samples for each work are used and some additional works are also included to add greater variety and to increase the problems of discrimination. The full list of works included is given in Table 6.3.

Despite the immense variety of these works (or perhaps because of it),[11] the cluster analysis gives remarkably good results which indicate that, in the vast majority of cases, all the samples from each individual work show a clear resemblance to each other. It is true that only in the case of Isaeus (*Oration* 3, four samples) do we have 100% success, with all four samples forming one cluster, but for the other works it is usually only one, or at the most two samples which have gone astray. This applies to *Politicus, Hellenica, Against Timarchus, Against Eratosthenes*, and the *Apology. Panegyricus, Archidamus*, and *De Pace* have been confused, presumably a reflection of the distinctive

[11] Sharp distinctions between groups of samples are advantageous in a cluster analysis as it enables the clusters to be formed more definitively.
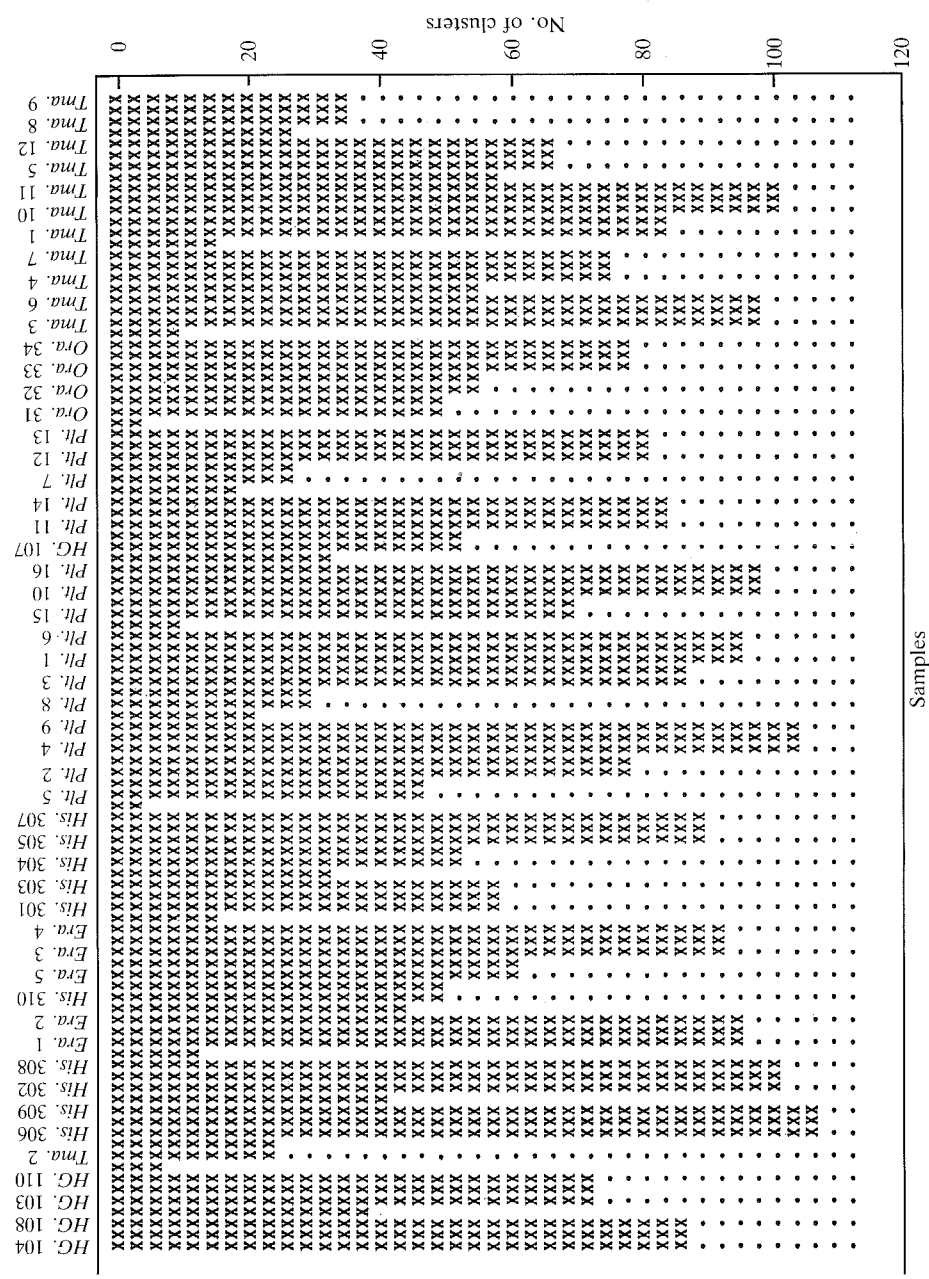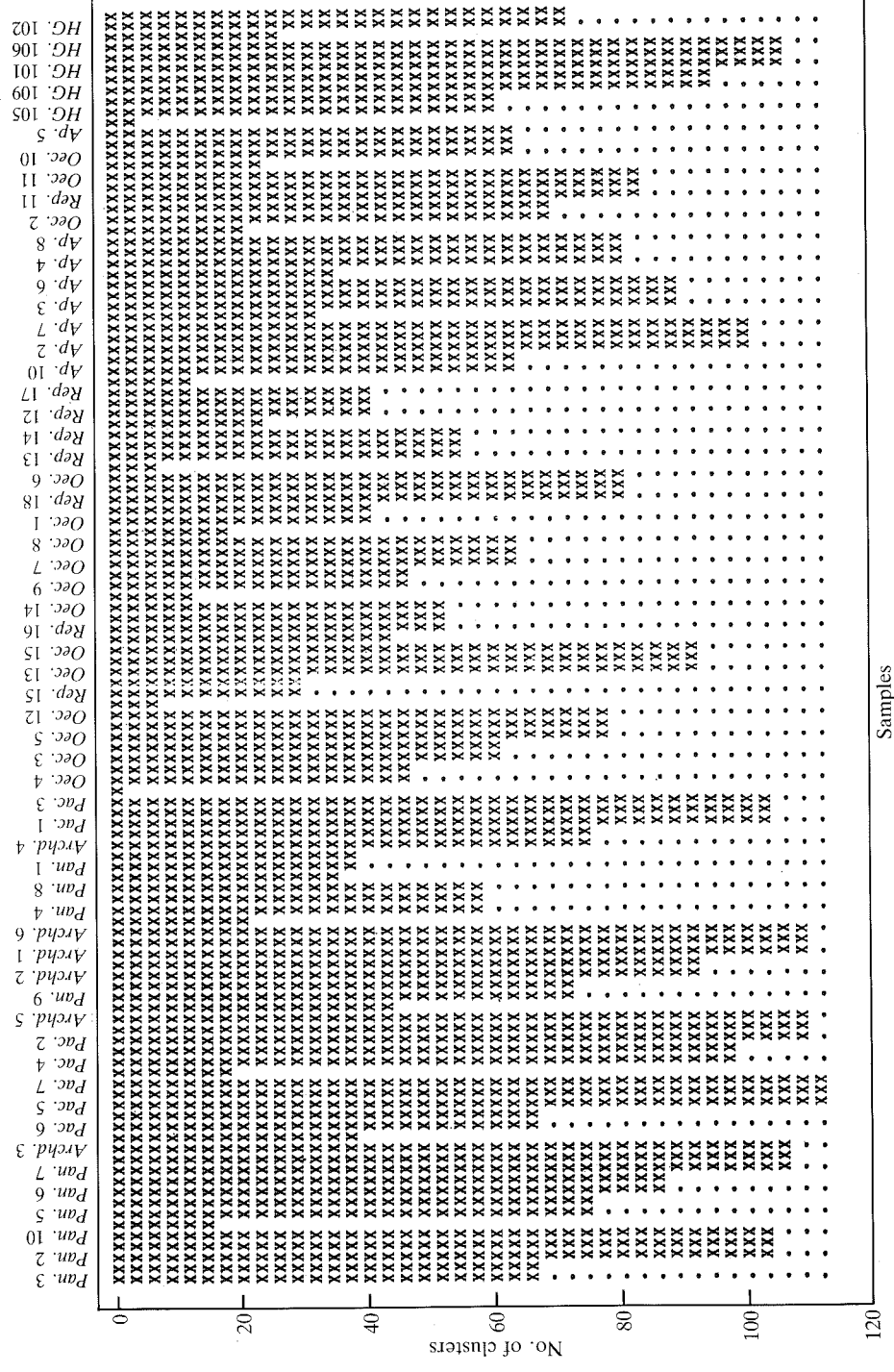
FIG. 6.2 Ward's minimum variance cluster analysis on 111 samples using all 37 variables

TABLE 6.3   *Samples included in Fig. 6.2*

Author	Work	No. of samples
Aeschin.	*Tma.*	12
Is.	*Ora.* 3	4
Isoc.	*Archd.*	6
	Pac.	7
	Pan.	10
Lys.	*Era.*	5
Pl.	*Ap.*	8
	Plt.	16
	Rep. 1	8
Th.	*His.* 3	10
Xen.	*HG*	10
	Oec.	15
TOTAL		111

style of Isocrates. There is also some difficulty in sorting out the respective allegiances of the samples of *Oeconomicus*, the *Republic*, book 1, and the *Apology*. Also, as in the previous example, Thucydides' *History* has been split into two sections by the five samples of Lysias' speech. Nevertheless, the family relationship of most of these samples to the parent group is fairly clear, despite the confusing cross-currents produced by genre and authorship similarities, which perhaps we could have predicted.

Now it is clear that we cannot, from such results, derive any clear picture of authorship, for they do not display a grouping dependent upon authorship characteristics. This is confirmed especially by the separation between *Politicus* and the other two Platonic works and also by the distance which separates *Oeconomicus* and *Hellenica*, both works of Xenophon. At a guess one could speculate that the forces which are operating here are threefold: local, generic, and authorial. The final result is determined by a complex combination of these three forces, all of which contrive in various ways to generate similarities between samples. The local forces are predominant here and cause the overall grouping of samples into the parent work. They relate to the restricted character of each work, that combination of subject matter and linguistic usage favoured by that particular author at the time of writing which gives to each individual work its unique flavour. Secondly, the generic forces are those which stem from the genre of each work and in this case cause the proximity of Isaeus and Aeschines which is apparent in the tree diagram, as well as that of Xenophon in the *Hellenica* to Thucydides and the relationship between *Oeconomicus* and the *Republic*. Finally, we have the

authorial characteristics contributed by the distinctive use of language which we assume to be found in all authors, a personality stamp which acts independently, unaffected by subject matter or genre, and remains (it is hoped) reasonably constant throughout their writing career.

No doubt the above is a fairly simplistic approach to the whole problem of identification of style, and I merely advance it as a broad explanation of the variety to be found in the results obtained from cluster analysis on different combinations of samples, and the consequent difficulty of interpreting those results. Obviously it is not possible to separate these stylistic forces from each other, any more than it is possible to separate style from thought, as if language existed independently and unconnected with the desire to express one's thoughts in a communicable form. All these forces are allied, interdependent, and closely related to each other.

As a general rule it seems that local forces predominate, but this may be because we still have not included a large number of works from a single author. In the case of Isocrates (*Panegyricus*, *Archidamus*, and *De Pace*) it is clear that authorial and generic are stronger than local characteristics, although which of these former two predominates it is impossible to say. We shall also discover later that differentiation between individual works of Plato using cluster analysis is virtually impossible except in the broadest sense of a clear separation between early and late works. But this is to anticipate. It seems clear that cluster analysis is excellent for providing a general picture of where differences and similarities lie, but it cannot do more than it was designed to do and show us, amid all the myriad groupings caused by the variety of linguistic forces operating on each sample, where precisely the authorship affinities lie.

Let us pursue further, however, the task of discriminating between individual works, as this ultimately is the key to any subsequent understanding of authorship discrimination.

The next illustration uses discriminant analysis on the 111 samples taking as variables only the nine BLETs, thus preserving a sample : variable ratio of approximately 12 to 1. We specify in advance that the samples belong to the groups (works) to which prior knowledge allocates them, and the discriminant analysis then gives us the classification results according to the underlying structure of the data.[12] If the variables justify a grouping in the manner required, in this case into parent works, then the discriminant analysis will concur and indicate that the proposed classification is justified by the discriminant function and that most, if not all, of the samples may be

[12] This is a discriminant analysis using the pooled covariance matrix. Use of the 'within' covariance matrix is less suitable because it requires restriction of the number of variables to less than that of the smallest number of samples in any one group. In general, apart from this difficulty, I have found that the use of the 'within' covariance matrix produces a discriminant function which is too sharply defined and its effectiveness beyond the range of the immediate samples which were used to produce it is unpredictable.

classed in the specified groups. If the opposite is the case the results will show that many of the samples have been misclassified.

Using only the BLETs perhaps implies a suggestion that we are placing greater reliance on grammatical features as discriminators, since the preponderance of inflexions in Greek, especially at the ends of words, is likely to influence the BLET readings, and it seems probable that the variance recorded will depend in some way on grammatical constructions and the effect they have on word inflexions. In fact, we find that these nine variables do have considerable discriminant power and only 17 of the 111 samples are listed as being misclassified. The full results are shown in Table 6.4, from which it may be seen that the most serious defect is the classification of four samples of Aeschines with works from various other authors. This represents a failure rate of 33% for this author. However, even at this level of misattribution it is still apparent that there is sufficient evidence to show that *Against Timarchus* has characteristics which distinguish it from all the other works, and that on a majority verdict it is justifiable to regard it as a separate entity. The same argument applies to all the other works, but with increased assurance as the percentage rate of misclassification declines.

But how could we have interpreted the results if they had not been as clear-cut as this? Suppose, for example, that eight out of the 12 Aeschines samples had been classed with Lysias, would we then be justified in assuming that the author of this work was indeed Lysias and not Aeschines as previously thought? Clearly there would be a prima-facie case for declaring that the two works came from the same population, at least in terms of the variables which were used in the analysis, but it would be unwise to declare outright an acceptance of Lysias' authorship without a further study of additional cases in which speeches of a similar nature were thrown in the melting-pot together. In any case, it is highly unlikely that the results would be so one-sided, for if Lysias and Aeschines were so similar in these two speeches it is almost certain that a high proportion of the Lysias samples would be shown as being more properly classified with Aeschines, and we would then be in doubt as to which of the two authors the credit should be given.

There remains also the possibility that the confusion might be caused by coincidence of genre, although here too one would have to exercise caution in ascribing the affinity to that cause rather than to some chance combination of circumstances, since it is clear that genre is not dominant in all cases, but only in those where for other reasons the individual and author-determined characteristics seem to be weak. These problems of mistaken or uncertain authorship will be dealt with fully in Chapter 8, but at this point we may at least observe that the main difficulty is often that of finding similarities, as with a full variable set discrimination at about the 90% level of success always seems to be possible, even if the entire Platonic corpus is included with the other six authors, giving a total of 702 samples from approximately

TABLE 6.4   *Discriminant analysis on 111 samples using all BLETs*

Work	Sample	Classified into work	Posterior probability of membership in work											
			Ap.	Archd.	Era.	HG	His.	Oec.	Ora.	Pac.	Pan.	Plt.	Rep.	Tma.
Plt.	3	Tma.	0.0000	0.0000	0.0667	0.0006	0.0011	0.0000	0.0335	0.0000	0.0000	0.1912	0.0000	0.7069
	6	Ora.	0.0000	0.0000	0.0107	0.0042	0.0003	0.0000	0.6207	0.0000	0.0000	0.2979	0.0000	0.0662
	11	Era.	0.0000	0.0000	0.8130	0.0000	0.0037	0.0000	0.0000	0.0000	0.0000	0.1831	0.0000	0.0002
Tma.	2	Oec.	0.0063	0.0000	0.0000	0.0068	0.0043	0.9165	0.0000	0.0000	0.0000	0.0000	0.0010	0.0650
	3	Ora.	0.0000	0.0000	0.0000	0.0000	0.0000	0.0005	0.6388	0.0000	0.0000	0.0001	0.0001	0.3605
	6	Rep.	0.0115	0.0000	0.0000	0.0000	0.0002	0.0127	0.3029	0.0000	0.0000	0.0000	0.6013	0.0714
	12	Era.	0.0000	0.0000	0.7716	0.0001	0.0000	0.0000	0.0003	0.0000	0.0000	0.0001	0.0000	0.2279
Pac.	1	Archd.	0.0000	0.6088	0.0000	0.0000	0.0000	0.0000	0.0000	0.3297	0.0615	0.0000	0.0000	0.0000
	2	Archd.	0.0000	0.4375	0.0000	0.0000	0.0000	0.0000	0.0000	0.4085	0.1540	0.0000	0.0000	0.0000
Pan.	4	Pac.	0.0000	0.4135	0.0000	0.0000	0.0000	0.0000	0.0000	0.5104	0.0761	0.0000	0.0000	0.0000
	7	Archd.	0.0000	0.6176	0.0000	0.0000	0.0000	0.0000	0.0000	0.3163	0.0660	0.0000	0.0000	0.0000
Archd.	4	Pan.	0.0000	0.1502	0.0000	0.0000	0.0000	0.0000	0.0000	0.1686	0.6812	0.0000	0.0000	0.0000
	6	Pac.	0.0000	0.3495	0.0000	0.0000	0.0000	0.0000	0.0000	0.3689	0.2817	0.0000	0.0000	0.0000
His.	310	Era.	0.0000	0.0000	0.6603	0.0084	0.3312	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Oec.	10	Ap.	0.9101	0.0000	0.0000	0.0000	0.0000	0.0899	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
HG	106	His.	0.0000	0.0000	0.0000	0.3092	0.6907	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	107	His.	0.0001	0.0000	0.1959	0.2489	0.3162	0.0008	0.0009	0.0000	0.0000	0.0337	0.0000	0.2036

45 works. The problem, therefore, is one of reducing the variable set in a way which will not prejudge the results or cause such a high level of overlapping and misclassification of samples as to make interpretation very difficult.

We are therefore entitled to claim with considerable confidence, on the basis of the above results, that even a reduced set of variables, in this case only the 9 BLETs, contain a substantial amount of information relating to the different nature of these 11 works.

It is a comparatively simple matter to increase the level of successful classification by adding further variables. Thus if we use the first 10 ALETs plus all 9 BLETs, a total of 19 variables, the success rate increases to 95.4%. Five samples in all are misclassified, but four of these are from the Isocratean corpus, cases of inter-authorship reclassification which cannot be considered as damaging defects of the method. These four samples are *Archd.* 13, *Pac.* 2 and 3, and *Pan.* 7, all of which are reclassified with alternative speeches from Isocrates. The only other sample which goes astray is *His.* 310, the 10th sample from Thucydides, which is classed as being more akin to Lysias.

Discriminant analysis works by maximizing the *F*-ratio, the ratio of between-group to within-group variance. The original variables are combined in such a way as to achieve this result, a different combination being used to discriminate each work from all the rest. This combination of variables is of the form

$$V = a_1 x_1 + a_2 x_2 + ..., + a_n x_n,$$

where x_n is the *n*th variable, V is the combined variable and a_1, a_2, a_n stand for the coefficients which the analysis generates for each group.[13] V corresponds to the discriminant function, although the true value of the latter can only be expressed in terms of matrix algebra and involves the calculation of the matrix of mean scores for the groups and the inverse of the variance/ covariance matrix. Nevertheless, for each sample what we are in fact left with at the end of the calculation is a single number, the discriminant function, and it is the value of this function for each sample that decides the classification of that sample.

The success of discriminant analysis with these 111 samples shows that sufficient information is available to characterize each work by a mathematical function derived from the variables, and this knowledge justifies the extension of the method subsequently to larger data sets in the hope of detecting differences between authors and of discovering the means of classifying the products of each author in a way which distinguishes them from all others. The technique will figure largely in Chapter 8, where an

[13] In practice some limiting condition must also be applied, otherwise the number of possible combinations which gives *F* its maximum value is infinite. Thus we might specify that the sum of squares of all the coefficients should be one, or $(a_n)^2 = 1$ and this ensures that the value of coefficients for each group is unique.

attempt is made to determine the authenticity of various works in the Platonic canon.

Another technique which I use extensively later is that of canonical discriminant analysis (Candisc), and it is worth looking at it first in this more limited context. With this method the objective of optimizing the use of the variables is directed towards maximizing the value of multiple *R*, the correlation between the combined variable, which is calculated as a combination of all the original variables (ALETs, BLETs, and CLETs), and group membership, which is conceived of as another variable. This may sound rather complex, but broadly speaking it implies that the greater the number of samples which are correctly classified the higher will be the value of *R*. For any one particular combination of samples and variables there will be a unique set of canonical variables which corresponds in each case to the maximum possible value of *R*. The number of canonical variables will be equal to either (1) the number of groups minus one, or (2) the number of original variables, whichever is the lesser.

In a sense Candisc is a data reduction technique, but it also has the aim of maximizing the effectiveness of the original variables in conformity with some pre-specified classification into groups—in the case that we are currently considering it is the classification of samples into the various parent works.

Again, as with discriminant analysis, if there is no basis for the classification, if the readings of the variables are purely fortuitous and do not relate to any characteristics dependent on the author or the work, then this technique will not be able to create fictitious differences out of nothing. This will become obvious by the achievement of a low value for R^2, indicating that the variables do not provide very much of the sort of information that is requested of them, that is, information relating to differences between the works.

Candisc used on the 111 samples of the previous example yields a maximum of 11 canonical variates. I used as the original variables all 19 ALETs plus the 9 BLETs, giving a total of 28 for the Candisc procedure to work on. The result may be presented either as a straightforward output of canonical scores or, as I have chosen, a tree diagram resulting from a cluster analysis on the first seven canonical variates. The latter is more easily assimilated than a vast array of numbers, although essentially it is not adding anything new to the results of the Candisc, but merely presenting them in a more sophisticated format.

The tree diagram thus obtained is shown in Figure 6.3, and we may see from it immediately that nearly all the problems of the preceding cluster analysis of Figure 6.2 have been ironed out.

The *Apology*, Isaeus, Aeschines, the *Politicus*, *Hellenica*, Lysias, and Thucydides are all grouped with 100% success and the only apparent errors
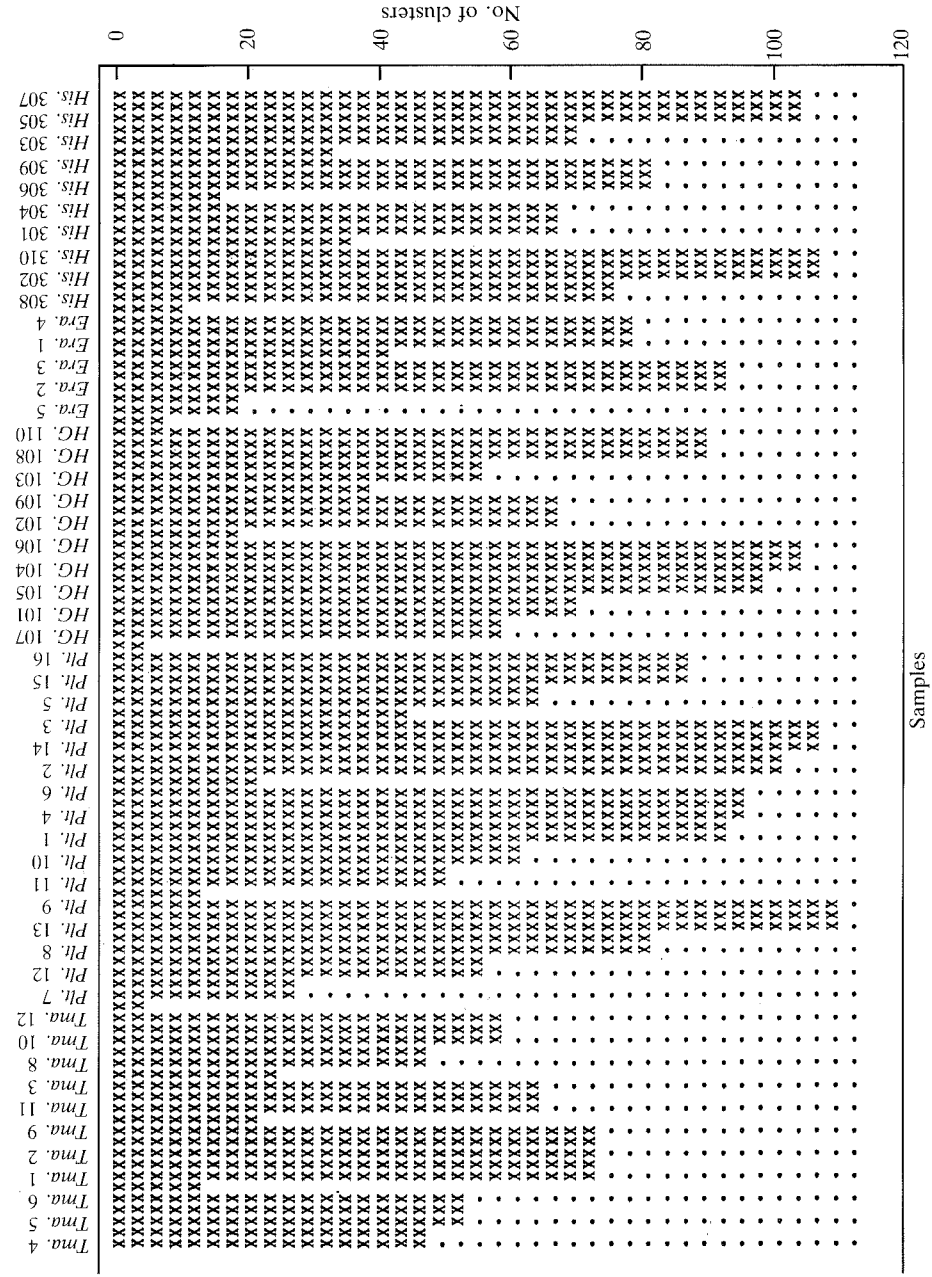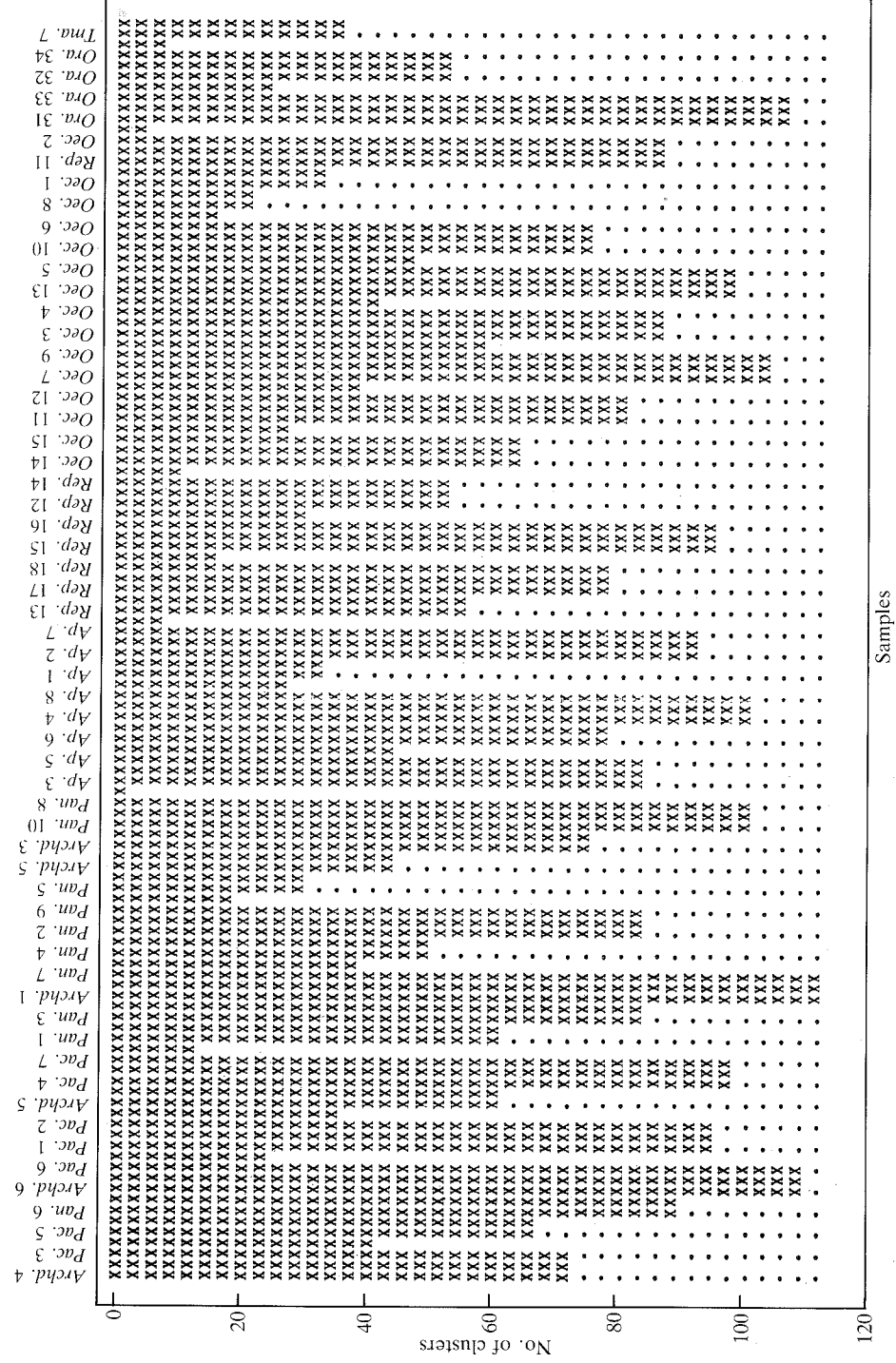
FIG. 6.3   Cluster analysis on 111 samples using seven canonical variables based on ALETS and BLETS

of allocation occur between *Oeconomicus* and book 1 of the *Republic*, and a confusion between the three speeches of Isocrates. This, however, may be improved on by using all 11 Candisc variables, by which means even the Isocratean speeches are separated, although remaining in one large cluster, and only one sample of the entire set is misallocated, *Rep.* 11 from the *Republic*, which is found amongst the *Oeconomicus* cluster.

It is surely valid to speculate on the possible performance of a conventional stylometric approach in a similar situation—whether or not it could hope to distinguish between these various works on the basis of the presence or absence of certain words, or measurements of word length or sentence length? And it would have to do this not by looking at the works as a whole, but by dividing each into as many 1000-word sections as it will make and then showing that the resulting measurements on each section are nearly identical. It seems most unlikely that such results could be achieved by any use of univariate statistics, because, with such a varied range of material, overlapping of scores would be inevitable, especially as each work is divided into relatively small samples. There is certainly no evidence to date that investigations using traditional methods of stylometry are capable of dealing with works at this level of detail. The general effect of univariate methods seems to be one of fragmentation rather than unification, as for every feature that confirms a certain provenance or affinity of a work another half dozen at least will be found which contradict.[14]

Style is such a nebulous concept, so open to different interpretations and so difficult to define precisely, that the only hope of dealing adequately with its complexity seems to lie with multivariate analysis. Stylometry can measure many features which are either directly or indirectly related to style—it does not matter greatly how devious the relationship is, nor is it even necessary in most cases to be able to relate these measurements to conventional concepts of style—but it is only by the use of multivariate techniques that the salient features of these measurements will be brought into full play and we will be able to develop a useful stylometric judgement of the extent to which works and authors differ from each other.

By consulting the relevant statistics derived from MVA it is usually possible to see how effective the initial measurements are as discriminators of style. Thus in the above example of Candisc I reproduce the test statistics for the canonical variables in Table 6.5. These reveal, in the values for canonical

[14] The distinction between multivariate and univariate statistics, although nominally related to the number of features measured, is also dependent on the way the measurements are used. Thus Morton, in his study of the positional *hapax legomena* ('Once: A Test of Authorship Based on Words which are not Repeated in the Sample', *ALLC Journal*, 1(1) (1986), 1–8), uses only this one feature in his analysis of Greek authors, whereas Kenny's New Testament analysis takes 99 features of the Greek language for comparison (*A Stylometric Study of the New Testament* (Oxford, 1986), p. 123). Nevertheless, Kenny's approach is still essentially univariate, for he does not apply any of the combinatorial methods which MVA requires.

R^2, that the original 28 variables are capable of producing canonical variates which account for more than 90% of the group membership variance. Or, to express the matter slightly more simply at the risk of misrepresenting the mathematical realities, more than 90% of the samples may be allocated to the correct work by the use of the first canonical variate, CAN1, so that we have effectively compressed nearly all the information relating to group membership into this one variable.[15]

It is worth bearing in mind when attempting to assess the value of results such as those obtained above that, typically, in the behavioural sciences a figure of 0.5 for R is considered acceptable so that R^2 is often no higher than 0.25. Values of 0.9 for R^2 are comparatively rare because most measurements in the uncertain world of human behaviour and experience are intrinsically flawed—either they cannot be collected adequately, or the conceptual relationship between variables and model is weak, or the complexity of the situation outdistances the theory, or the unpredictability and apparent irrationality of all but the simplest and most elementary aspects of life distorts the basic structure beyond recognition. We would therefore count it as a great stroke of luck to be able to account for as much as 50% of the variance of the dependent variable in any given situation. For, essentially, what we are discussing here is the ability to predict. Could one, for example, predict what students would do well at university by a study of their A-level grades; or, from a knowledge of different soil types, the yield of potatoes per acre; or the future behaviour patterns of men and women by studying their backgrounds, economic, emotional, social, and intellectual; or the economic performance of a country from a knowledge of all the factors that are supposed to contribute to that performance? The answer to all these questions is 'Yes, to a limited degree.' But one would not anticipate much more than, say, 60% success in one's predictions and for the first and last two cases considerably less.

Our hope in this work is to predict authorship using a series of elementary measurements, primitive variables which reveal something of the underlying structure of the language. The success so far in attributing samples correctly suggests that it will not be a great step to make to attempt the task of authorship discrimination. Had the figures been less promising, had the rates for correct classification dropped to 50% or less, then clearly it would have been worth while to return to the original variables and to a consideration of the possibility of improving on the choice and of taking other linguistic features as the basis of measurement. Now, however, with as much as 99% correct classification it would be superfluous to make any further effort to

[15] Canonical R^2 shows the proportion of variance of the group membership variable which is accounted for by the canonical variable specified. A proportion of variance, however, cannot be translated directly into a proportion of samples, so that my simplification does involve a distortion of the mathematical basis of the argument.

TABLE 6.5 *Canonical discriminant analysis on 111 samples using ALETS and BLETS*

111 samples   110 DF total[a]
28 variables   99 DF 'within' classes
12 classes (works)   11 DF 'between' classes

	Canonical correlation	Adjusted canonical correlation	Approximate standard error	Squared canonical correlation	Eigenvalue	Difference	Proportion	Cumulative
1	0.984125	0.978591	0.003003	0.968503	30.7487	23.2954	0.5431	0.5431
2	0.938991	0.912478	0.011279	0.881704	7.4534	1.7660	0.1317	0.6748
3	0.922206	0.892840	0.014258	0.850463	5.6873	0.9128	0.1005	0.7752
4	0.909299	0.890912	0.016512	0.826824	4.7745	2.3532	0.0843	0.8596
5	0.841258	0.780319	0.027868	0.707714	2.4213	0.5274	0.0428	0.9023
6	0.808981	—	0.032947	0.654451	1.8939	0.0972	0.0335	0.9358
7	0.801523	—	0.034092	0.642440	1.7967	0.9741	0.0317	0.9675
8	0.671824	0.588756	0.052312	0.451347	0.8226	0.3065	0.0145	0.9821
9	0.583468	0.488352	0.062887	0.340435	0.5162	0.2244	0.0091	0.9912
10	0.475259	0.337485	0.073810	0.225871	0.2918	0.0846	0.0052	0.9963
11	0.414285	0.332633	0.078982	0.171632	0.2072	—	0.0037	1.0000

Tests of H_0. The canonical correlation in the current row and all that follow are zero

	Likelihood ratio	Approximate F	Numerator DF	Denominator DF	Probability > F
1	0.00000081	7.3122	308	770.981	0.0
2	0.00002567	5.4798	270	712.507	0.0001
3	0.00021702	4.7015	234	651.951	0.0001
4	0.00145126	3.9788	200	589.213	0.0001
5	0.00838028	3.2185	168	524.191	0.0001
6	0.02867153	2.7766	138	456.786	0.0001
7	0.08297374	2.3290	110	386.908	0.0001
8	0.23205520	1.6754	84	314.492	0.0008
9	0.42295476	1.3345	60	239.511	0.0679
10	0.64126324	1.0605	38	162.00	0.3882
11	0.82836784	0.9439	18	82.00	0.5303

Multivariate test statistics and F approximations

Statistic	Value	F	Numerator DF	Denominator DF	Probability > F
Wilks' lambda	8.08608E–07	7.312	308	770.981	0.0
Pillai's trace	6.721384	4.601	308	902.00	0.0001
Hotelling–Lawley trace	56.61362	12.900	308	772.00	0.0
Roy's greatest root	30.74872	90.050	28	82.00	0.0001

Note: F statistic for Roy's greatest root is an upper bound.
[a] DF = degrees of freedom.

extend the scope of the variables, for it is abundantly clear that they are capable of doing all that is required of them. The choice of variables and the way in which they have been used may be taken as having been fully justified by these results, and it now remains to be seen how we may deal with the more complex problems of authenticity and chronology of composition.

To close the chapter I include two examples which show how coincidence of genre or other, probably unknown factors affect the outcome of cluster analyses. In the first case (Figure 6.4) samples from various orations are used. These are taken from the works of five authors. The full list of samples is given in Table 6.6. It is possible to class all these works under the general heading of 'oratory' and, if necessary, to split them into various subdivisions, for they are all in some way influenced by the rhetorical ideas of the time, although in many ways they differ greatly from each other. It is doubtful if much is to be gained from the attempt to classify them in this way, at least for our purposes, for it is evident that, whatever other influences are working to cause them to cluster in the way shown, genre is only partially responsible for the results.

One can accept that the speeches of Isocrates are all of a kind, for he devoted his entire life to rhetoric and its teaching, founding his own school in Athens, which probably pre-dated the foundation of Plato's Academy,[16] and his speeches have a distinctive, much-polished flavour which once tasted can never be forgotten. Hence it comes as no surprise to find that the three

TABLE 6.6   *Samples included in Fig. 6.4*

Author	Work	No. of samples
Aeschin.	*Tma.*	12
Is.	*Ora.* 1	2
	Ora. 2	2
	Ora. 3	4
	Ora. 5	3
Isoc.	*Archd.*	6
	Pac.	7
	Pan.	10
Lys.	*Era.*	5
Pl.	*Ap.*	8
TOTAL		59

[16] J. F. Dobson, *The Greek Orators* (London, 1919), p. 128, R. C. Jebb, *The Attic Orators* (London, 1893), ii. 8. *c*.392 is the date given for the foundation of his school.
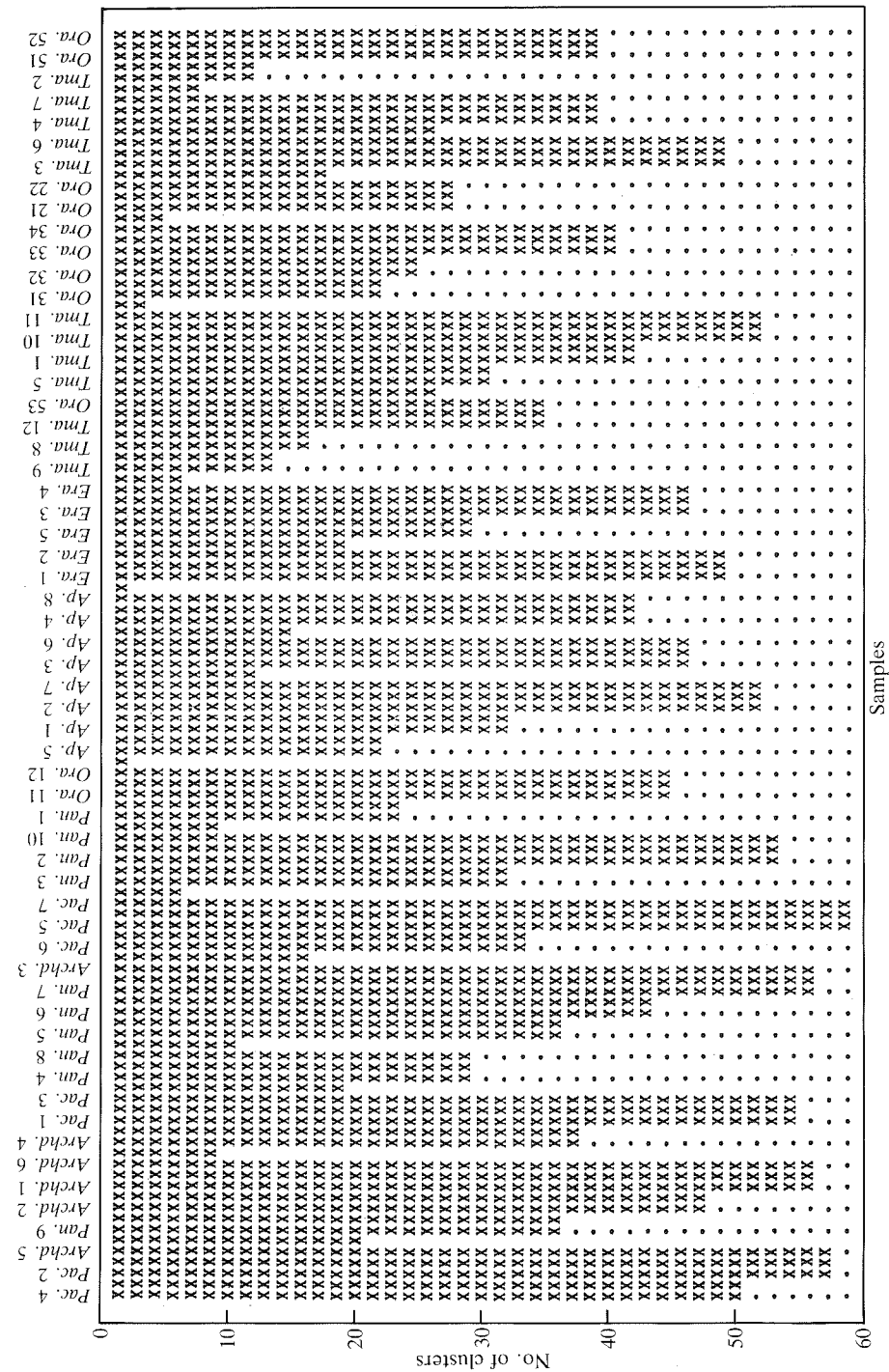

FIG. 6.4   Ward's minimum variance cluster analysis on 59 samples from orations using all 37 variables

speeches of his included here cannot be separated. But it is very difficult to understand why *Oration* 1 of Isaeus should cluster with this group, or indeed why the Isaean samples should lack the unity which is so evident in the case of the *Apology*, Lysias, and Isocrates. Superficially there is nothing in these four speeches by Isaeus that would cause one to anticipate this varied clustering, for individually they seem to have a reasonable coherence (apart from no. 5), yet their affinities with both Isocrates and Aeschines is puzzling. However, there is a tradition that Isaeus was a pupil of Isocrates, as well as being a teacher of Demosthenes, and perhaps *Oration* 1 is evidence of an early Isocratean influence.[17] *Oration* 3 is generally recognized as being different in many ways from all the other speeches, some scholars thinking it to be the earliest of his speeches, others the latest.[18] It does seem to be in a cluster on its own, a cluster more definitive perhaps than any other apart from the one containing the Isocratean works. Its nearest neighbour is, however, another speech by Isaeus, no. 2, so perhaps the affinity shown is a chronological one, although without further enquiry it would be rash to draw hard and fast conclusions from such evidence. I cannot explain why the Aeschines speech should be split in two by the small cluster of Isaeus samples. There is little resemblance between the two authors and one may only speculate that the affinity may be one of contemporaneity, since the Aeschines speech dates to 345 and *Oration* 2 of Isaeus to perhaps less than 10 years earlier. However, Isaeus no. 3 is of uncertain date and no. 5 is too early to justify a grouping with Aeschines on chronological grounds, if indeed its date is correctly set at *c*.389. But at least this latter speech is on the edge of the Aeschines cluster and its relationship may be only a weak one.

But I shall proceed no further with the interpretation of these results, as it is beyond the scope of this enquiry, which aims to justify the methods used and then to apply them to the Platonic corpus in the hope of settling the question of authenticity of certain dubious dialogues and to provide grounds for the chronological ordering of the various works. Suffice it to say that this simple cluster analysis of the orators has revealed much that I doubt any reading of the works could have unearthed and has suggested many questions which might repay further investigation.

The last example, Figure 6.5, uses works by Plato and Xenophon, 76 samples in all. This is made up of Plato's *Apology* (8), *Protagoras* (15), the *Republic*, book 1 (8), and Xenophon's *Memorabilia* (30) and *Oeconomicus* (15). There is a broad division between the works of the two authors, commencing at the point where the *Apology* cluster begins. But apart from this one work, the *Apology*, there is no clear evidence of individuality in any of the other works. Broadly speaking the Platonic samples are grouped on

---

[17] Jebb, *The Attic Orators*, ii. 266.
[18] E. S. Forster, Loeb text, p. 75 supports a later date; on the basis of a study of clausulae rhythms, R. F. Wevers (*Isaeus*) dates it to *c*.389 as one of the earliest speeches of Isaeus.
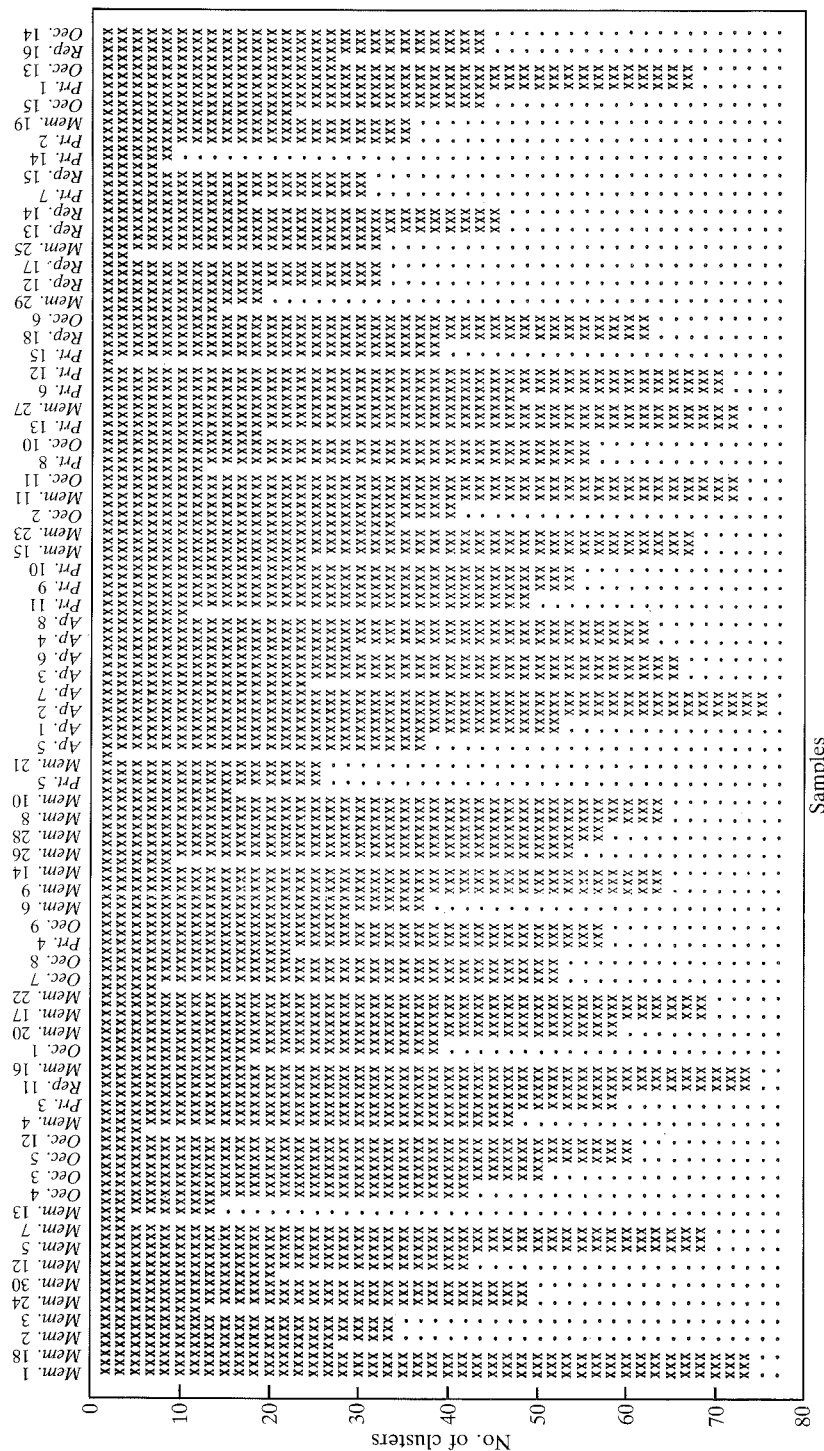
FIG. 6.5 Ward's minimum variance cluster analysis on Plato and Xenophon samples using all 37 variables

the right of the diagram and those of Xenophon on the left, but there is evidently much room for confusion, with 14 of the Xenophon samples appearing with the Plato cluster, or nearly one-third. On the other side the confusion is not so widespread, with only four out of the 31 Plato samples being placed amidst the main *Oeconomicus* and *Memorabilia* cluster.

However, it is clear from such evidence that there will be problems in separating these two authors, and the pronounced differences which were found between various works in most of the previous examples do not occur in all cases.

Before proceeding to a more detailed study of authorship discrimination it is necessary to look more closely at Plato and to give some consideration to the problems of Platonic scholarship which stylometry might be of some assistance in solving. This forms the subject of the following chapter.

7
Plato: The Background

PLATO was born in 427 and died at the age of 80 in 347.[1] His writing career probably spanned more than half a century since it is generally thought that his earliest dialogues date from soon after Socrates' execution in 399. Indeed, if any credence is to be attached to the anecdote which Diogenes Laertius records,[2] when Socrates exclaims in response to a reading of the *Lysis*[3] by Plato, 'Heavens, what a lot of lies the young man tells about me!', then it would be necessary to date this dialogue (and possibly others) to before Socrates' death.[4] Since another tradition records that the *Laws* was left 'in the wax'[5] at his death it does seem likely that the period of writing covered slightly more than the first half of the fourth century.

Of the writings of Plato mentioned in ancient sources, all appear to have survived.[6] At any rate there are no references in later antiquity to works which are no longer extant in our own day, a fact which stands in stark contrast to the fate of the writings of so many other authors from the period. However, the canon does include some dialogues which were thought to be spurious in antiquity and are still considered to be so today. Their titles are *Axiochus, Demodocus, Eryxias, Sisyphus, On Justice, On Virtue*, and the *Definitions*. None of these will be dealt with in this study since they are all too short to fulfil the criterion of 1000-word length required as the basic textual unit for analysis and there would be no point in challenging or confirming received opinion on the basis of inadequate and slender evidence.[7]

This leaves 36 complete dialogues of varying length (counting the 13 letters as one dialogue). Of these the following are judged by many scholars to be of doubtful authenticity: *Alcibiades* 1, *Alcibiades* 2, *Amatores, Clitophon, Epinomis, Hipparchus, Hippias Major, Hippias Minor, Ion, Menexenus, Minos, Theages*. The *Epistles* are a separate case as they do not fall into the class of

---

[1] There is some uncertainty about the exact date of his birth, but it was evidently within two or three years of 427. See W. K. C. Guthrie, *A History of Greek Philosophy*, iv. 10 n. 2. Diogenes Laertius (D.L.), iii. 2–3.

[2] ii. 35.

[3] It is interesting to record that some of the stylometric tests for dating the dialogues result in placing *Ly.* as the earliest (see ch. 9).

[4] This would extend the time-span to include the latter years of the 5th cent.

[5] D.L. iii. 37. ἐν κηρῷ. This presumably means that final corrections had not been made. The original copy was written with a stylus on a board or tablet over which a thin sheet of wax had been spread. The final copy would be written in ink on papyrus.

[6] A. E. Taylor, *Plato: The Man and His Work* (London, 1926), p. 10.

[7] Guthrie, *History*, v. 383 f. and 394–8.

dialogue and raise special problems of their own. As such they will be dealt with separately in a later section. The above list includes all those dialogues over which modern opinion is divided. In the nineteenth century it was fashionable to reject a still greater number, reducing the total of accepted dialogues to a dozen or less. Perhaps this tendency to scepticism in what was after all a supremely rationalistic age should act as a warning to us that all is not as secure as we might believe in our confident ascription of the 24 major works to Plato. At least stylometry might help in determining how close stylistically the main works of the corpus are to each other and enable us to set limits to the subjective element in discussions of authenticity.

Even if all the works of the above list were rejected as spurious the loss might not be thought disastrous for an understanding of Plato. The danger is, however, that unless consistent and reliable standards are used as a basis for acceptance or rejection, the arguments and methods which are advanced against one dialogue, arguments which are usually based on philosophical interpretations and an impressionistic judgement of style, may then be used with equal effect against the more established dialogues of the canon. One might have hoped that after nearly two and a half millennia of exegesis and analysis the main facts about Plato's philosophical thought would have been secured beyond reasonable doubt. However, there is still considerable disagreement on many topics, even fundamental ones, and it is difficult to separate genuine works from spurious in the surrounding uncertainty.

Partly these disagreements stem from the very nature of philosophical speculation, since it is often difficult to formulate a philosophical proposition unambiguously, and Plato, being the leading philosopher of his day, was creating new ideas and continuously working on methods for dealing with existing problems. Partly also the dramatic presentation of these ideas in dialogue form leads to difficulties, because it is no easy matter to determine what Plato's thoughts and conclusions were when they have to be filtered out from the words spoken by a series of characters, historical, fictional, imaginary, who fill the pages of the dialogues. It is rather like asking what were the thoughts of Euripides and Sophocles. Although one can supply an answer in terms of the ideas expressed by a Medea or an Oedipus, yet the connection between that and the innermost thoughts of the authors themselves is often very tenuous.

This is perhaps an overstatement of the case for ambiguity, although when I look at the history of interpretation of a dialogue such as *Parmenides* I am not so sure. Obviously a philosophical work is not to be read in the same spirit as a play, but the dramatic setting does contribute something to the dialogues which involves us in the complexities of human behaviour, and this adds an additional dimension to the philosophical debate which is more akin to artistic experience than to a consideration of the fine detail of philosophical truth.

One could argue that philosophy and the philosophical life embrace the whole of human existence and dialogues such as *Gorgias* and *Phaedo* indicate that Plato believed this to be so. Nevertheless, it is difficult to deal adequately in one work with human life in its entirety and it tends to be split into self-contained parts for ease of discussion. This is no less true of philosophy than of other branches of knowledge. Thus it is customary to discuss the ontological status of the Forms, or Plato's epistemological propositions, as though they could be formulated independently of the characters who expressed them, and in a sense independently of human existence. At any rate modern criticism treats them as such and discussions on topics related to the nature of the Forms and other Platonic concepts can become quite abstract and move a long way from the original dialogue which engendered them.

Yet if Plato had wished to have his ideas formulated in this way he could have done so more simply than by casting them in dialogue form. These dialogues might well be the starting- or finishing-point, or reflective in some way, of discussion which actually took place in the Academy. They express ideas which were currently of interest to Plato and were of contemporary importance in the Academy and elsewhere, wherever philosophical speculation was active. Consequently, one should not necessarily expect from them a coherent set of doctrines, duly laid out in sequence and forming 'one entire and perfect chrysolite', though it would be strange if there were no similarities between them since, apart from the spurious dialogues, they were all written by the same man.[8]

It is clear, however, that the dialogues were intended in some sense as philosophical treatises. The later ones especially are more didactic in tone and often have a declared objective.[9] Many are linked in thought with those of earlier and later dates of composition and there is an overall unity and sequence of development which spans the entire corpus, despite minor contradictions and changes of direction.[10] Plato's most famous pupil, Aristotle, frequently refers to ideas presented by Plato in the dialogues[11] in a way which leads us to believe that they were intended primarily as a means of

[8] It is difficult to say anything about Plato which is not in some way controversial. I feel sure, however, that for most readers Plato evokes a different response from that evoked by Aristotle. This is not due to the subject matter as much as to its presentation, for we accept that Aristotle is trying to give us a contemporary account of the philosophical speculation of the time, but we feel uncertain about the relationship of Plato to the material of the dialogue.

[9] e.g. the task of defining the Sophist in the dialogue of that name. The declared objective often does not correspond with the theme or hidden purpose of the work as perceived by most scholars. Thus an important theme of the above dialogue is thought to be the isolation and analysis of the predicative use of the verb 'to be', as distinct from its use in the absolute sense of existence, a task which had not previously been achieved by anyone.

[10] Presumably only the unitarians would agree wholeheartedly with the claim that there is an abiding unity in Plato's thought. See R. E. Allen (ed.), *Studies in Plato's Metaphysics* (London, 1965), introduction; H. Teloh, *The Development of Plato's Metaphysics* (Pennsylvania, 1981), introduction.

[11] See Guthrie, *History*, iv. 41 and n. 2.

disseminating philosophical thought, despite Plato's disclaimer in the *Phaedrus* and in the *Epistles*.[12]

Nevertheless, it is often difficult in the case of individual dialogues to assess the motives and purposes which led to their composition. While this is especially true of the more abstruse dialogues such as the *Parmenides* and *Cratylus*, even those others which, through the familiarity of frequent reading one assumes to be more open to explanation and thematically extrovert, still arouse deep controversy.[13]

In fact, because so much of the contemporary literature has been lost and because there is a consequent dearth of biographical information, it becomes almost impossible to set any given dialogue in an adequate context, a context which would explain why Plato chose to write it when he did and the ideas, persons, and influences which it was designed to support, contradict, counteract, or examine.

With all these uncertainties it is not surprising that differences of opinion arise as to the interpretation of Platonic doctrines. Scholars are driven to the expedience of assuming that in certain circumstances Socrates is the mouthpiece of Plato and building their account of the latter's thought on that foundation. On such terms *faute de mieux* is Plato's philosophy reconstructed, but obviously with the result that there are gaps and a considerable amount of disagreement. This applies even to the fundamental ideas which Plato introduced, especially to the Theory of Forms, an explanation of reality which most would agree is central to the Platonic philosophy. It attempts to explain the visible, tangible, and ephemeral world of the senses by reference to a more abiding reality which is perceived only by the mind, the world of Forms.

However, it would not be appropriate to attempt here a descriptive analysis of Plato's philosophy. I am concerned mainly with the relevance of stylometry to the problem of authenticity and the possibility of establishing a chronological sequence for the dialogues.

The importance which individuals might attach to the outcome of such enquiries depends to a large extent on the position which they already hold with regard to the works of dubious authenticity. Thus if one were a firm believer in the genuineness of the *Epistles* it would be gratifying to discover that stylometry could confirm one's faith, or frustrating to find that this belief had been undermined. In addition, many interpretations of the works are based on an assumed sequence of composition and presumably it would

[12] *Phdr.* 276 c–d; *Eps.* 2, 314 c, 7, 341 b–e. Authorship of the epistles is disputed.

[13] I refer to works such as *Prt.*, *Grg.*, and *Smp.*, where the need to refute certain philosophical ideas or to introduce new ones is scarcely sufficient to explain the full panoply of dramatic presentation with its glittering cast and the sparkling interplay of wit and intellectual fervour. Something seems to be lacking in our knowledge of the contemporary scene and its impact on Plato's mind.

be helpful if this sequence, whatever it might be, could be established by independent means.

Alternatively, one might hold that the *Epistles* and the dialogues mentioned earlier as being of dubious origin are of little worth, and that acceptance or rejection need not alter greatly anyone's conception of Plato. Such an argument is perhaps possible to sustain for the dialogues, although even here one would hesitate to dismiss too lightly *Clitophon*, *Epinomis*, and *Hippias Major* as being of no account. But in the case of the *Epistles*, especially *Epistle* 7, we do run into greater difficulties, because the content here shapes the way in which we interpret the Theory of Forms and affects our understanding of Plato's attitude to political activities. Was he entirely the disinterested onlooker or did he believe perhaps that the philosopher could play an active role in shaping human societies and did he himself attempt to put his theories to the proof? Is the long, analytic passage of *Epistle* 7 (342 a–344 c) to be taken as a genuine statement of the scope of the Theory of Forms and how we might interpret it, or is it the work of some pupil doing his best to expound the Platonic doctrines?

Generally speaking, rejection of the *Epistles* goes alongside a narrow interpretation of the Forms,[14] since if they (the *Epistles*) are accepted as genuine, it is difficult to reconcile this with a possible abandonment by Plato in later life of the Theory of Forms, or a restriction of its application to only a few limited classes or concepts.

The question of chronology also is not one to be treated too lightly, for it is germane if one sees in Plato a development of ideas, rather than a static reiteration of a philosophy which he formulated in his earlier years and subsequently did little to change. Most commentators do in fact accept that there is considerable change of content as well as of emphasis and it is therefore quite important to get the sequence right, especially if we can discover a method which has some claim to objectivity and does not rely on a subjective assessment of philosophical content or linguistic data.

I will therefore set the following tasks as an objective for the stylometric investigation:

1. to establish the probability of authenticity of the *Epistles*;
2. to examine, similarly, the three dialogues *Alcibiades* 1, *Epinomis*, and *Hippias Major* and assess the likelihood of their being genuine;
3. similarly for the minor dialogues, *Alcibiades* 2, *Amatores*, *Clitophon*, *Hipparchus*, *Hippias Minor*, *Ion*, *Menexenus*, *Minos*, and *Theages*;
4. to investigate the general relationship of all the dialogues to each other so as to determine the extent of divergences which exist between them;
5. to establish an approximate chronology for all the dialogues.

[14] See Guthrie, *History*, v. 399–401, and especially 402 n. 1, for a survey of recent opinion, based partly on computerized studies.

We will now look at each of these five objectives in turn, but it is important to bear in mind that all the investigations will be interlinked, and information gleaned from one set of results need not be kept in a separate compartment, never to be used except when applied in the original context. Rather, since the methods and techniques are essentially new, we must continually make reference across to results already obtained, and suggestions of approach derived from any part of the enquiry which may seem to be relevant to the current investigation will be freely applied.

The Epistles

These are 13 in number, being ostensibly by Plato, and written mostly towards the end of his life. The most important by far is the seventh, not only for the biographical information which it contains but also for the philosophical digression,[15] which is unique among the writings in that it contains a direct description of what purports to be the essentials of his later philosophy undistorted by transmission through the mouth of a fictional character or by the dramatic situation. To Guthrie it is 'nothing less than a short apologia for his whole life and thought'.[16]

To be sure on the question of authenticity, as far as human affairs ever can be assured, would certainly be a great help to scholars, not so much from the point of view of historical accuracy[17] as that of philosophical exegesis. For this epistle contains information relevant to an understanding of Plato's philosophical system as he himself perceived it at a late period in his life (he was approximately 75), and those[18] who maintain that the later writings show a considerable modification of his treatment of the Theory of Forms, if not a complete abandonment of it, must somehow deal with the evidence which they find here to the contrary, either by deciding that the letter is spurious or by reconciling what is contained therein to their own interpretation of the later philosophy.

There seems to be a positive and inescapable restatement of the Theory of Forms in terms of the essence or being of an object, the thing itself, as distinct from the knowledge which we gain of particular examples of that object or the three related attributes of quality by which that knowledge is achieved (342 A–D). Plato explains this by taking the example of the circle. Particular instances (or instantiations, to use the frequently encountered philosophical

---

[15] 342 A–344 B.   [16] *History*, v. 403.

[17] The biographical details could be accurate even if the *Epistles* are not by Plato. Certainly it is not impossible that someone nearly contemporary could have inserted the details of the Syracusan ventures into the *Epistles*, or even concocted them. We have no means of checking this independently, since it is clear that the later *Lives* which are preserved, those by D.L. and Apuleius (2nd cent. AD) and those by Olympiodorus and an anonymous author (6th cent. AD), seem to derive from traditions which rely heavily on the *Epistles*.

[18] The so-called Revisionists. See Teloh, *The Development of Plato's Metaphysics*, pp. 10 f. and 209. Also R. E. Allen, *Studies in Plato's Metaphysics*, introduction.

term) of circular objects may occur variously, by being drawn by the geometer, for example, or by being turned by the engineer on a lathe. These objects are characterized by having a name, a definition, and an image. From these three attributes a fourth is derived, namely the knowledge which we have of each individual object. It is the 'essence' or 'being' of each object (the circle itself, αὐτὸς ὁ κύκλος, in the example) which constitutes the ultimate reality of things in the universe, the Forms, as they are generally known, the fifth item in Plato's sequence of things which are associated with each object. All the first four are related to the fifth, the essence of the thing, which is in itself distinct from them and does not suffer, as they do, coming into being, or decay, or the possession of contradictory qualities: '... the circle itself, to which all the other qualities are related, suffers none of these experiences, being different from the others' (342 C); '... but the circle itself, as we declare, contains in itself neither more nor less of the nature which is opposite to it' (343 A).

The first four attributes, name, definition, instantiation (or image or example), and knowledge, are related to the quality of the object, whereas the fifth is concerned with, or is itself the essence or being of that object,[19] and it is of that which the soul especially seeks to have knowledge. In this quest it is continually thwarted by the four attributes of quality which obtrude themselves as if they were the true essence of being which is the object of the soul's search. Thus the soul becomes confused and the true perception of reality is easily confuted by the changing and shifting appearance which the four quality attributes present of phenomena in the universe.

As to the types of object for which Forms exist Plato gives the following statement: 'The same is true also of the straight and spherical form,[20] and of colour and of the good, the beautiful and the just, and of all bodies either artificially produced or arising naturally, water, fire and all such things, and of all living creatures and dispositions in souls, and of all experiences and actions' (342 D). This is as definite and direct a statement of the scope and application of the Theory of Forms to the sensible world as can be found anywhere in the dialogues. Moreover, the transcendental nature of these Forms is also emphasized, for the soul seeks to know primarily the nature of things and by acquiring knowledge of the first four qualitative realities, by testing them dialectically in a friendly atmosphere, by sifting them continuously and examining them 'there bursts out the light of intelligence and understanding regarding each object' (344 B).

Irrespective of how one chooses to interpret all this it is obviously important to have some criterion, as objective as possible, by which to judge the authenticity of the epistle. Nowadays majority opinion favours the

---

[19] The τί as opposed to the ποῖον τί. 342 E, 343 B.

[20] i.e. the same explanations as were given in the case of the circle.

acceptance of the *Epistles* (*Epistle* 1 excepted) as genuine. Previously, in the nineteenth century, they were almost universally rejected, so that the wheel has come full circle. What is more surprising, perhaps, is that some convinced unitarians found it necessary to reject the whole collection as forgeries,[21] even though it would seem to support the contention of an underlying consistency of philosophical ideas throughout Plato's life. Partly this is due to what appears to be a deliberate attempt to obfuscate his own ideas and principles which Plato indulges in in the second epistle (312 D ff.), to the personal revelations of the thirteenth which are difficult to reconcile with their concept of a character adapted only to the rarefied atmosphere of philosophical debate, and to the autobiographical tone of the seventh, which they distrust.

There is no point however in any further rehearsal of the arguments *pro* and *contra* these epistles.[22] They are based mainly on exegetical uncertainties and personal preferences, not on any indubitable historical fact or incontrovertible philosophical principle. The assertions of stylistic merit, or lack of it, are just as wild and varied in this field as in that of any other of the dubious dialogues.

We shall be examining the five epistles, 2, 3, 7, 8, and 13, which alone are of sufficient length to allow this sort of investigation, and looking solely at the stylometric evidence for inclusion in the genuine corpus of Platonic works. The letters ignored are (with approximate word length of each shown in brackets), 1 (300), 4 (400), 5 (300), 9 (180), 10 (70), 11 (270), and 12 (80). Perhaps the longer ones would just provide enough stylometric evidence to base a judgement on, but it would be departing from the principles enunciated in the early part of this study, that all samples are to be 1000 words in length. These eight letters are in any case of minor importance, and most critics would probably be content to form an opinion of their origin by reliance on what may be deduced from the more weighty epistles of the series and the plausibility of content of each taken individually.

The stylometric tests applied to the five main epistles will attempt to determine precisely how close they are in style to the remainder of the Platonic corpus using criteria which are specially selected to emphasize differences of authorship. In addition, the placing of these epistles relative to the other dialogues will be important, as they may in most cases be dated absolutely (assuming that they are genuine), and they should therefore group themselves approximately with dialogues which are conjecturally of the same date. In particular, the three late epistles, 3, 7, and 8, must be close to the *Laws*, which is universally reckoned to be a late work.[23] The two earlier ones,

---

[21] See Guthrie, *History*, v. 399–401 for a discussion of current opinion on the topic.

[22] J. Harward, *The Platonic Epistles* (Cambridge, 1932), pp. 86 ff. and 188–9 deal in detail with the stylistic features of the *Epistles*. See also G. R. Morrow, *Plato's Epistles* (Indianapolis and New York, 1962).

[23] It is generally thought to be one of Plato's final works, if not indeed the last. Ryle suggests in *Plato's Progress* that it was substantially completed by the time of Plato's second visit to Sicily (361–360) but few agree with this interpretation.

---

2 and 13, are more problematic, because we do not have any satisfactory reference point against which to place them. 13 is the earliest of the Sicilian letters and must be dated to *c*.366.[24] *Epistle* 2 belongs either to 364 or 360, depending on which Olympic games are the subject of the reference.[25] The problem is that none of the dialogues may be confidently dated to that time, for most internal references give at best a *terminus a quo* for the date of composition and the available clues are usually the subject of contention. As Field observes,[26] 'one of the things that we should be most glad to have would be a single reliable date for the composition of any of the dialogues. Yet it appears likely that not even among the writings of his immediate successors was such information to be found.'

We cannot be sure, therefore, where these two letters should occur in the corpus, although we may conjecture that the *Theaetetus* will be one of the dialogues nearest to them, as that in all probability dates from *c*.369.[27] This would make *Epistle* 13 pretty close, but no doubt differences of genre will play their part in complicating the problem.

The details, however, will be discussed more fully when we come to analyse the stylometric evidence for authenticity of the epistles and the chronology of the dialogues.

Alcibiades 1, Epinomis, and Hippias Major

These are the three dialogues of moderate importance over which opinion is still divided.[28] Guthrie, for example, ignores *Alcibiades* 1,[29] rejects *Epinomis*,[30] and accepts *Hippias Major*.[31] *Alcibiades* 1 is probably the earliest of the three, while *Epinomis* is the last, being a sequel to the *Laws* (i.e. if the three are genuine). There are excellent examples of Socrates' 'aporetic' approach in *Alcibiades* 1, where he is shown reducing the young Alcibiades to total perplexity through his probing questions (e.g. 127 D). The choice of characters for the dialogue probably arose from the anti-Socratic sentiment[32] which stemmed partly from Socrates' known acquaintance and affection for Alcibiades, a man who had done great damage to Athens in the Peloponnesian War, and Socrates would have been held responsible, in the popular imagination, for the evil developments in Alcibiades' character. At any rate, such seems to be the gist of the accusation[33] against which Xenophon defends

---

[24] Harward, *The Platonic Epistles*, p. 230.

[25] 310 D. Harward, *The Platonic Epistles*, p. 166.

[26] G. C. Field, *Plato and His Contemporaries* (London, 1930), p. 2.

[27] Guthrie, *History*, v. 361. He gives 369 to 367 as the most likely date. It could be later, since as a tribute to Theaetetus it need not necessarily be exactly contemporary with his death. 'Of good men the memory is always green'.

[28] The list is somewhat arbitrary. These three dialogues are rather longer than those in the next section, and the philosophical content is of greater moment.

[29] Guthrie, *History*, v. 387.   [30] Ibid. v. 385.   [31] Ibid. iv. 175.

[32] Field, *Plato and His Contemporaries*, pp. 149—f.

[33] The accusation was made by Polycrates, who published an anti-Socratic treatise in the form of a speech for the prosecution at Socrates' trial. Ibid. 136.

Socrates in the *Memorabilia*.[34] Perhaps Plato's contribution to the liberation of Socrates' memory from the taint of this particular charge may have been this dialogue.[35]

However, its authenticity is not crucial to an understanding of Plato, although it would be interesting to know if this was one of his earlier works or a later forgery. It is an excellent introduction to Plato's ethical and political philosophy and perhaps for this reason the slur of doubt is cast upon it,[36] since it is not thought probable that Plato would have constructed such a bald and occasionally lifeless summary.

The *Hippias Major* or *Greater Hippias* on the contrary is thought by some to be too lively and its humour too gross and crude.[37] Apart from this it does contain material which depicts Socrates' concept of universals,[38] the abstract ideas such as beauty, justice, and goodness, which were the forerunners of Plato's Forms. For this reason attempts are often made to fit it into some scheme which shows the development of Plato's philosophy and its authenticity or its chronological position in the canon may, in such schemes, be crucial. Consequently, quite a lot of attention has been focused on it, although majority opinion seems now to favour acceptance of it as genuine. Ideas similar to those found in the *Hippias Major* are also met with in other earlier dialogues, such as *Euthyphro*, and we would expect that *Hippias Major* would be of a similar date.

Finally, the *Epinomis*.[39] As mentioned previously, this appears to be a sequel to the *Laws*.[40] On internal evidence the case may be argued either way for or against authenticity, as is true for so many of the dialogues. Suspicion is aroused against it chiefly because Diogenes Leartius mentions that it was thought by some to be the work of Philippus of Opus.[41] With a slight straining of the sense of the passage it could also be taken to mean that Philippus copied out the work as he did also for the *Laws*. In any case one does not have to take D.L.'s comments as incontrovertible for the simple

---

[34] I. ii. 12–28.

[35] Anti-Socratic sentiment was evidently enough to secure Socrates' condemnation at this trial, but the extent to which it blasted his memory subsequently is difficult to assess. The literature of the hostile tradition (which has not survived), including Polycrates' pamphlet, probably arose in response to the apologetic and adulatory productions of Socrates' disciples, such as those of Antisthenes. At what stage Plato and Xenophon joined the fray is not known. See Field, *Plato and His Contemporaries*, p. 136.

[36] Taylor, *Plato*, pp. 12 f., 522–6.

[37] Guthrie, *History*, iv. 175 and notes for summary of pros and cons. C. H. Kahn also rejects it ('Did Plato Write Socratic Dialogues?', *CQ*, NS 31(1) (1981), 305).

[38] Guthrie, *History*, iv. 188–91; Teloh, *The Development of Plato's Metaphysics*, pp. 67 ff. for interpretations.

[39] Taylor, *Plato*, pp. 497 ff.; Guthrie, *History*, v. 385, and references.

[40] Taylor was a staunch defender of this work: 'If the *Epinomis* is spurious, we must deny the authenticity of the most important pronouncement of the philosophy of arithmetic to be found in the whole Platonic corpus' (*Plato*, p. 14). Guthrie, however, remains unconvinced (*History*, v. 385).

[41] D.L. iii. 37.

---

reason that he often records divergent traditions and scholars have no choice but to pick and choose between the evidence.[42]

We shall be concerned with the stylometric evaluation, attempting to show that the proximity or separation of a given work from the main corpus either confirms or precludes Platonic authorship. We anticipate also that each dialogue should group approximately with those adjudged to be contemporaneous, making due allowances for possible genre affinities, which may distort the results.

The fact that so little may be proved by traditional methods regarding authenticity renders the stylometric approach of great value because it does not depend on any philosophical allegiance and betrays no prior partisanship for one author over another. In that sense it may be said to be objective, and impartiality is further guaranteed by the choice of variables, which, as indicated in Chapter 2, could hardly be claimed to favour any one author or style. It is the underlying characteristics of language which are being measured, a sort of wave motion which varies from sample to sample, but, nevertheless, under certain well-defined conditions shows a similar pattern.

Of course, if the results are entirely unexpected, and we find, for example, that Plato could not have written the *Laws* or the *Republic*, then it will be necessary to re-examine the basis of our judgements. But where there is reasonable accord between expectation and outcome then it seems to be quite legitimate to accept the verdict of stylometry in cases where traditional methods have failed.

Alcibiades 2, Amatores,[43] Clitophon, Hipparchus, Hippias Minor, Ion, Menexenus, Minos, Theages

These are all comparatively slight in length and content, and for these two reasons, as much as any other, previous generations have athetized them. *Ion, Hippias Minor*, and *Menexenus*[44] are nowadays usually reinstated, the two latter because they are referred to by Aristotle,[45] the former because no good reason may be found for rejecting it. Information on the current status of the others may be found most conveniently in Guthrie,[46] who, however, does not attach much importance to the subject of authenticity in these cases. 'The question of their authenticity is of no great importance for students of Plato'. He himself only deals with the following seven as being possibly spurious: *Alcibiades* 2, *Amatores, Clitophon, Epinomis, Hipparchus, Minos, Theages*, and he does not classify *Alcibiades* 1. Consequently, the problem

---

[42] Most, for example, reject the story of Plato being sold into slavery (iii. 19–20) and the claim, derived from Aristoxenus, that most of the *Republic* is contained in Protagoras' contradictions (iii. 37–8).

[43] Also referred to as the *Lovers* or the *Rivals*.

[44] Guthrie, *History*, iv. 199 (Ion), 191 (*Hp.Mi.*), 312 (*Mx.*).

[45] *Rhetoric*, 1367b8 and 1415b30 (*Mx.*), *Metaphysics*, 1025a6 (*Hp.Mi.*).

[46] Guthrie, *History*, iv. 41.

need not loom large in his mind, for there is nothing in any of them which it would be a tragic loss to discover was not written by Plato, nor anything either which could cause such offence to the memory of Plato that it is vital for it to be excised from the canon. However, the wholesale rejection of large chunks of the corpus in the nineteenth century should cause us to guard against complacency in this matter.

Grote alone of Victorian scholars defended the authenticity of the whole canon,[47] and there is perhaps more willingness nowadays to accept his verdict, at least until stylometric evidence proves otherwise.

Finally, lest it be considered that the whole subject of authenticity has been treated too lightly, I should add that more information dealing both with the problems associated with individual dialogues and the methods to be used will be presented in the relevant sections of the following chapter.

Interrelationships between the dialogues

This will be of crucial importance. It is essential to know how each dialogue relates to others in its vicinity and to those more distant. It would be useless to prove, for example, that there was no significant difference between *Epinomis* and the *Laws* if the same could also be proved for any number of Platonic dialogues.[48] Or that significant differences existed between *Hipparchus* or *Theages* and the *Republic* if similar differences were also found between the latter and many other of the established dialogues of the canon.

Hence we shall require some sort of quantitative measure of similarity (or difference) so that an overall picture may be obtained of the entire corpus and any extremes of style may then be assessed against this background, rather than look at each dubious dialogue in turn and individually, and then attempt to decide for or against authenticity in a total vacuum.

This part of the work will overlap with the stylometric assessment of chronology, a topic dealt with in the following section.

Chronology of composition of the dialogues

It would be helpful if some of the dialogues gave reliable evidence of their dates of composition, as these could act as fixed reference points by which to date the remainder. Unfortunately, clues of this nature are not abundant and the few datable references are usually attended by uncertainty.[49] Perhaps the most reliable of those usually cited is in the introductory section of the *Theaetetus*. Theaetetus himself is being carried home suffering from wounds sustained on the battlefield at Corinth (142 A–B). The battle in question is

[47] G. Grote, *Plato and the Other Companions of Socrates*, 3 vols. (London, 1875), pp. 206–11 and 452.

[48] Or even works by Xenophon and others which will be included for comparison.

[49] See Field, *Plato and His Contemporaries*, ch. 5 for a summary of these datable events mentioned in the dialogues. Also Ross, *Plato's Theory of Ideas* (Oxford, 1951), ch. 1; Guthrie, *History*, iv. 52.

probably that of 369 and the whole dialogue may well be a tribute to the memory of Theaetetus who, it is thought, died subsequently of his wounds. An earlier expedition against Corinth took place about 395, but this seems to be altogether too early a date for the composition of this dialogue, which deals with problems more typical of the later works. 369 would therefore be approximate for the *Theaetetus* and it is unlikely that it would have been many years after, for Plato returned to Sicily in 365. (I assume that the story of the Sicilian visits is true, though it need not affect greatly the dating of the *Theaetetus*.)

Another dialogue which contains a historical reference is the *Symposium*, where Aristophanes speaks of the 'dispersal of the Arcadians by the Spartans' (193 A). This is thought to refer to an event in 385, but since the dramatic date of the dialogue is *c*.416 and its supposed narration took place *c*.400, the mention here of a historical incident of 385 is evidently an anachronism. The dialogues were written, however, for a contemporary audience, who would have appreciated, we presume, a topical allusion, and it seems to be possible that the *Symposium* could have been written close to that date.

A mention in the *Laws* (683 B) of a conquest of Locri by Syracuse is usually taken to refer to an action by Dionysius of about 356, nine years before Plato's death, and this accords reasonably well with the supposed date of composition of the work, which is universally accepted as being one of his last.

The *Menexenus* is supposed to date itself precisely, since it takes the eulogy of Athenian history down to 387, the date of the Peace of Antalcidas, even though Socrates, who is reporting the funeral oration, died in 399. The anachronism is by no means a slur on the dialogue's authenticity since Plato is often liberal in such matters. This date, though, (387) presumably only sets an upper limit for the year of composition, for the apparent theme of the untruthfulness and fickleness of rhetoric is frequently encountered in Plato and would not have been out of place at almost any period of his writing career.[50]

Finally, the *Gorgias* is often dated to around 387,[51] when Plato was approximately 40 and first visited Sicily, since its tone of bitter disillusionment with politics corresponds with the passage in the seventh epistle[52] describing his revulsion against the intrigues of the time and his judgement that no improvement in the human situation was possible until philosophers became kings or kings philosophers. However, the connection is rather tenuous, especially if one remembers that *Epistle* 7 was written 35 years after

[50] e.g. *Grg., Phdr., Mx., passim.*

[51] E. R. Dodds, *Plato, Gorgias* (Oxford, 1959), pp. 18–30.

[52] 324 C–326 B. Especially the statement 'This was the view I held when I came to Italy and Sicily at the time of my first visit' (i.e. 387).

the first Sicilian visit, and one would not expect absolute chronological accuracy to apply to recollection of states of mind, whatever the events which brought them into existence may have been.[53]

This more or less brings to an end the list of absolute dates from which the order of composition and dating of the dialogues is surmised. Some further information is deducible, regarding their relative placing, from cross-references within them. Thus *Critias* follows the *Timaeus*, *Politicus* succeeds the *Sophist*, and the *Epinomis* (if it is genuine) is a sequel to the *Laws*. In addition, the *Timaeus* refers back to the *Republic*, purporting to be a continuation of that discussion, but most critics think that there is a considerable gap in the dates of creation of the two. This is a topic about which there is considerable controversy, especially since the publication of an article by G. E. L. Owen[54] challenging the traditional late dating of the Timaeus.

Most critics also claim to recognize references to the *Parmenides* in the *Theaetetus* (183 E) and *Sophist* (217 C), and the inference is that the former must pre-date the latter two dialogues. The *Theaetetus* also rejects the use of reported speech, a fact which plausibly confirms that it post-dates the *Parmenides*, which uses in its first section an especially cumbersome form of narrative (A told B, who told C, who told me, who am now telling you), as also does the *Symposium* and several presumably earlier dialogues.[55] The *Laws* also post-dates the *Republic*, as we gather from a reference in Aristotle,[56] but this is not surprising as tradition dates it as one of the last, if not the last, of Plato's works.

All this information combined gives an approximate sequence for some of the dialogues. This is shown diagramatically in Figure 7.1. It represents only a rough guide to what might be expected from a chronological listing. None of the dates has absolute accuracy and the sequence as shown is not incontrovertible. *Lysis* has been dated prior to 399 owing to the (probably apocryphal) story in Diogenes Laertius (see the beginning of this chapter). *Phaedo*, the *Apology*, and *Crito* confirm by their subject matter that they must be subsequent to Socrates' death, and by inference the same may be said of *Euthyphro*, for its dramatic date is shortly before Socrates' trial. How soon after his death is not known, but common opinion tends to put the *Apology* and *Crito* within five years of that event.

The relative position of all the late dialogues is uncertain, except where

---

[53] Kahn also relies heavily on this dating ('Did Plato Write Socratic Dialogues?', p. 307).

[54] 'The Place of the *Timaeus* in Plato's dialogues', in R. E. Allen (ed.), *Studies in Plato's Metaphysics* (first pub. 1953; London, 1965), p. 313.

[55] Perhaps I have exaggerated. In *Prm.* the sequence is Pythodorus—Antiphon—Cephalus—Listener; in *Smp.* a friend converses with Apollodorus who heard the story from Aristodemus. Strictly speaking the sequence is Socrates *et al.*—Aristodemus—Apollodorus—Friend—Reader. It is interesting that many early dialogues are conversations recorded directly without the interposition of any narrator (*Hp.Ma., Grg., La., Men.*).

[56] *Pol.* 1264^b26.

	Ly?
399	Death of Socrates
	Ap., Cri., Euthph., Ph.
385	
	Grg., Mx., Smp.

369	*Th.*
355	*Phlb.*
	Prm.
353	*Sph.*
	Plt.
347	Death of plato

FIG. 7.1 Approximate chronology for the composition of some Platonic dialogues. Arrows indicate post-dating of one dialogue by another.

indicated by an arrow. Thus we do not know in what relation the *Timaeus–Critias* group stand to the *Sophist–Politicus* pair and to place the latter two after the *Republic* is a decision based on several generations of scholarly research which has established fairly conclusively a late group of dialogues. The diagram represents a simplification of the possible combinations and presupposes that (1) the date quoted for the historical reference is nearly identical with the date of composition; and (2) each dialogue was written fairly quickly, was not revised subsequently, and its date of composition did not overlap that of the others.

However, it is quite possible that the longer dialogues took several years to write, especially the *Republic* and the *Laws*, and their completion may have been delayed while other dialogues of less moment were attended to.[57] We have to assume that this was not the case, simply to restrict the problem to manageable dimensions. Therefore, one should not impose an unwarranted

---

[57] Guthrie, *History*, iv. 52 f.

neatness on any solution proffered as a means of concealing the underlying uncertainties.

As far as (1) above is concerned, it is worth noting that all the dates mentioned previously may be dislodged by argument. As Pope observed of the religious controversies of the late seventeenth century: 'I found myself a Papist and a Protestant by turns, according to the last book I read. I am afraid most seekers are in the same case, and when they stop, they are not so properly converted as outwitted'.[58] The conjectural dates may be taken only as a rough guide, and, if other evidence suggests that they are unreliable, we must be even more chary of using them.

It is strange that Guthrie regards the use of historical citations and cross-references in the dialogues as the most objective method for dating them.[59] In fact, the greatest strides in producing an approximate order of composition have been made by the use of stylometric and linguistic tests, a method which was initiated by Campbell[60] in the mid-nineteenth century. It was he who first showed that the *Politicus* and *Sophist* must be late dialogues and introduced the idea of an affinity of style which could be measured by quantitative assessment of the minutiae of language. The work was continued by Lutoslawski,[61] Ritter,[62] Blass,[63] and others. Brandwood[64] surveyed all the evidence up to 1958, the year of his thesis, and himself did further analysis of clausulae rhythms so as to order more accurately the later dialogues.

The current position is that three successive groups of dialogues are generally accepted, or possibly four. I repeat in Table 7.1 the lists given by Guthrie[65] and Skemp,[66] the former using Cornford's conclusions,[67] the latter derived from Brandwood.

The orders of composition proposed by earlier scholars, von Arnim, Lutoslawski, Raeder, Ritter, and Wilamowitz, are given by Ross in *Plato's Theory of Ideas* (p. 2). He also offers a tentative chronology of 'those of the earlier dialogues which throw light on the theory of Ideas, and of all the later works', a list which I also reproduce.[68] This final list is more definitive than the other two in that it gives a sequence to the earlier dialogues, whereas both Guthrie and Brandwood refrain from pronouncing a verdict on dialogues which pre-date the *Republic* or the middle period of Plato's writing.[69] The most exhaustive stylometric research in this field has been done by Brand-

---

[58] Alexander Pope, letter to the Bishop of Rochester, 20 Nov. 1717, in J. Aitken (ed.), *English Letters of the XVIIIth Century* (London, 1946).
[59] Guthrie, *History*, iv. 42, 52.
[60] L. Campbell, *Sophistes and Politicus of Plato* (Oxford, 1867), introduction.
[61] Lutoslawski, *The Origin and Growth of Plato's Logic*.
[62] C. Ritter, *Platon*, 2 vols. (Munich, 1910).
[63] See J. B. Skemp, *Plato* (Greece and Rome: New Surveys in the Classics, x; Oxford, 1976), p. 13.
[64] 'The Dating of Plato's Works by the Stylistic Method'.
[65] *History*, iv. 50.   [66] *Plato*.   [67] *Cambridge Ancient History*, vi. 311 ff.
[68] *Plato's Theory of Ideas*, p. 10.   [69] Guthrie, *History*, iv. 50; Skemp, *Plato*, p. 14.

TABLE 7.1 *Chronology for the dialogues, derived from Guthrie, Brandwood, and Ross*

Guthrie's chronology

Early		Middle		Late	
Ap.	Hp.Mi.	Men.	Phdr.	Prm.	Ti.
Cri.	Hp.Ma.	Phd.	Euthd.	Tht.	Cri.
La.	Prt.	Rep.	Mx.	Sph.	Phlb.
Chrm.	Grg.	Smp.	Cra.	Plt.	Lg.
Euthphr.	Ion				

Brandwood's chronology

Group 1*		Group 2*		Group 3	Group 4	
Ap.	Hp.Mi.	Cra.	Mx.	Rep.	Ti.	Phlb.
Chrm.	Ion	Euthd.	Men.	Prm.	Cri.	Lg.
Cri.	La.	Grg.	Phd.	Tht.	Sph.	Epin.
Euthphr.	Prt.	Hp.Ma.	Smp.	Phdr.	Plt.	Ep.
		Ly.				

Ross's chronology

Birth of Plato, 429–427	Second visit to Sicily, 367–366
Chrm.	Sph.
La.	Plt.
Euthphr.	Third visit to Sicily, 361–360
Hp.Ma.	Ti.
Men.	Cri.
First visit to Sicily, 389–388	Phlb.
Cra. (?)	Ep. 7, 353–352
Smp., 385 or later	Lg.
Phd.	Death of Plato, 348–347
Rep.	
Phdr.	
Prm.	
Tht., 369 or later	

* In alphabetical order only.

wood,[70] and his conclusion is that no evidence exists for dating the early dialogues on the grounds of stylometric variation.[71] To see how fluid the situation is, one only needs to consult the article by C. H. Kahn referred to in the previous note, an article in which he proposes pushing back the *Gorgias* to a relatively earlier date and setting some of the dialogues of definition in the later period. To quote his own words: 'My heresy consists in removing

---

[70] 'The Dating of Plato's Works by the Stylistic Method'.
[71] Kahn, 'Did Plato Write Socratic Dialogues?', p. 306 n. 3.

the *Protagoras* and the four dialogues of definition (*La., Ch., Ly., Euthyphro*) from their usual place before the *Gorgias* in the 390s, and relocating them after the *Gorgias*, in the middle and late 380s' (p. 310).

He proposes to accept only four dialogues as being earlier than the *Gorgias*, namely the *Apology, Crito, Ion*, and *Hippias Minor* (he rejects the *Hippias Major* as unauthentic), in contrast to Guthrie who, following many other scholars and the weight of tradition, presumes that the *Gorgias* represents a departure from the purely Socratic and aporetic dialogues and becomes more the mouthpiece for Plato's own philosophical ideas, foreshadowing the great constructive works of the middle period.

I should also add that not all scholars pin their faith on the stylometric method of dating, a fact which should surprise no one as, despite some successes, it can hardly claim to be a proven method. The objection raised by Mackenzie in her article 'Putting the *Cratylus* in its Place'[72] is perhaps typical.

This is not, of course, to concede that stylometric considerations are or can be decisive in the dating of Platonic dialogues. After all, here we have a highly literate author, who may well achieve some of his effect by the deliberate echoing of the style of an earlier work. We cannot, that is, make any definite or plausible claims about where his development is unconscious and where he employs conscious allusions. It follows from this that the stylometric tests, which purport to examine unconscious development, beg the entire question of the method of Platonic composition.[73]

It would not be appropriate here to attempt a defence of stylometry, since the methods employed and the results achieved in this study may be allowed to speak for themselves.

My own position in tackling the problem of chronology will be to adopt a *tabula rasa* approach, making no other assumption than that of supposing the dialogues to follow some sequence or other.[74] This is not intended as a signal of distrust of what has already been achieved, but is simply because the

---

[72] M. M. Mackenzie, *CQ* 36(1) (1986), p. 150 n. 67.

[73] I do not really understand what is implied by this last sentence. It is true that some of the stylometric tests which have been used are not necessarily unconscious, for example clausulae rhythms and hiatus avoidance, but this hardly applies to many other of the characteristics which have been measured, such as use of particles and connectives and other frequently occurring words. While an author might be semi-consciously aware of his/her habits in such matters, it is not likely that anyone could keep the entire pattern of these linguistic features before the mind's eye when engaged in a piece of creative writing.

[74] When using all 37 variables, or a random selection of them, to produce a sequence, then no prior assumption about the relative positions of the dialogues is made. However, when selecting the best set of variables for emphasizing temporal changes of style, it is necessary to specify that one dialogue post-dates another. Thus I will most frequently assume that *Lg.* post-dates *Rep.*, a fairly innocuous assumption (and one for which there is plenty of evidence) and not as drastic as taking *Lg.* as being the final work of Plato's life. In any case, all that is really required is the hypothesis that any one dialogue differs in style from any other for reasons which may be ascribed to temporal stylistic variation. If one makes the mistake of assuming that dialogue A is later than dialogue B, when in fact it precedes it, the consequence would be that the sequence would be inverted, but the order would still be correct.

---

methods to be used do allow this freedom. This will become clear in the following chapters. Once a sequence has been established it should be possible to fix some firm dates to it by reference to dialogues of which the date is known (see above), or, preferably, by seeing how the sequence relates to the *Epistles*. These are a far more reliable source of accurate dating (assuming they may be shown to be genuine) than the dialogues themselves, since they refer to precise historical events and the contextual evidence usually dates them to within one or two years. *Epistle 7*, for example, refers to the murder of Dion, which occurred in 354, and it cannot have been written much later than that date, since the situation in Syracuse changed so rapidly that Plato's advice would become irrelevant in the space of a few months and be overtaken by circumstances.

The authenticity of the *Epistles* is dealt with fully in the following chapter, prior to the chapter on chronology, since it is obvious that we cannot rely on the evidence that the *Epistles* give of their dates of composition if they all turn out to be forgeries.

For a full discussion of the dating of the *Epistles* the best sources are Harward[75] and Morrow.[76] The following list of probable dates for the five epistles used in this study is taken from the former:

  2   360 or 364
  3   355
  7   353
  8   353
 13   366

Unfortunately, this only gives a span of 13 years, so there will remain many problems associated with the dating of the earlier dialogues. However, it will be a useful starting-point, and to presuppose in any case that it would be possible to produce an absolute date for each dialogue would be foolish in the extreme. The main objective is to make the best use of the material available to see what it will offer by way of solution to these undoubtedly complex problems.

It would be unrealistic to pretend that anyone who attempts to assess the authenticity of works in the Platonic canon could be free from prejudice. The traditional methods used to determine whether or not a thing is genuine are psychologically complex and depend on many chance details of character and experience in the person making the judgement. The problem extends over the whole range of arts and literature wherever doubts of interpretation or suspicions of provenance exist. To take an example from the humbler arts, Louis L. Lipski comments on the difficulty of judging the authenticity of early English delftware:

---

[75] *The Platonic Epistles.*   [76] *Plato's* Epistles.

There is a widespread belief that a collector, after years spent in handling antique ceramics, acquires an intuition which enables him, while examining a particular piece, to declare that it is not 'right'. This is a fallacy. The truth is that every piece which a collector examines leaves in his subconscious mind a number of items of information which, when correlated with details of other examples which are stored in his mind, create for him a mental picture of what the object in question should look like. In compiling this section of the book [on pieces of doubtful authenticity] the author has aimed not to rely on intuition or on a general impression when forming a judgement about whether a piece is 'right' or 'wrong', but to examine the individual characteristics of each example, compare them with those of other similar pieces which he has seen, and then arrive at the clearest and most logical assessment of what is 'wrong' with a piece and why it is not genuine.[77]

But despite such claims of clarity and logicality it is evident that much depends on human fallibility. Such comments apply no less to the problem of determining what is spurious among literary works, for each judgement relies on a stored mass of information about the works of the author in question, and one's idea of what that author's style is really like is a sort of distillation or crystallization of all those multifarious facts. This tends to weight opinion in favour of works which do not depart very widely from the norm, as if an author could only continue to write by repeating what he/she had already said in well-authenticated works. But it is the unpredictable which is most difficult to attribute satisfactorily, for, if style in language was of such a consistent character as has occasionally been implied, all the problems of authentication would have been solved long ago.

I am myself aware of having been swayed for or against a particular dialogue by a scathing or laudatory article, even before reading the dialogue itself, whereby one might hope to form a more balanced judgement. More experienced scholars are probably less easily influenced, but one cannot be sure that the basis for their decisions does not extend back to some seminal and formative period of their early career, a time when judgements are often hasty and ill considered.

Stylometry can offer a way out of these difficulties, because it is free of initial preference for any one interpretation and has no vested interest in maintaining that one should prevail over another—its reputation is not at stake—but the ultimate decision on whether or not one should accept the results of stylometric investigation must rest with the human agent, the expert who knows a good deal about the author being studied.

I have attempted to present the results which my own approach to stylometry has elicited in as fair and unbiased a way as possible, aware of the fact that others may disagree with my interpretations. I do not claim that what I shall show in the next two chapters represents the final word of stylometry on the subject of the Platonic dialogues, for what this science

[77] L. L. Lipski and M. Archer, *Dated English Delftware* (London, 1984).

might achieve in such fields of enquiry has scarcely begun to be formulated. What I do claim is to show that there is clear mathematical evidence for our belief in the differences between the dialogues and the differences between authors, and to offer a sound basis for the measurement of these differences and for determining the interrelationships between the dialogues, however much interpretations and emphasis may be changed by subsequent analysis and commentary.

The notion that 'much of his [Plato's] philosophy arises from a reflection on realities which are the same in all ages'[78] is one which I have found increasingly difficult to retain. For it seems to depend on an assumption that there exists an unalterable residuum of human experience from which one may strip the cultural and sociological influences, to be left with the human psyche pure and uncorrupted. But the consciousness of an age and that of an individual within it is shaped by so much that neither writing can transmit nor archaeology recover. Realities are perceived differently by different societies although imagination and knowledge may in part bridge the gap between them. Plato's age and our own are separated by an immense divide all the more serious because in many respects we do not even know that it is there. Yet few could claim to know what it was like to be alive in fifth-century Athens, since so much of the ancient world is gone and lost for ever, and we wander in a sort of semi-darkness, unaware of the gaping chasms and the mountain ranges that lie all around.

In a sense, therefore, the mechanical task (if such it is) of determining the authenticity and chronology of the dialogues by computer analysis is refreshingly simple. It need not be affected by personal preference nor need one be daunted by the scope of Plato's writing, his political, ethical, epistemological, and metaphysical theories, and the wealth of comment that they have generated—so much so that the student is liable to feel swamped and to experience that dizziness which Callicles' box on the ears would have given to Socrates.[79] In the circumstances it is almost pleasant to detach oneself from controversy, limiting oneself to a simple task and presenting the results in as objective a manner as possible, saying to all who show interest, 'This is what stylometry tells us about the Platonic dialogues. You may disagree, or offer your own interpretation, or present other evidence, but you ignore the results at your peril.'

For all lovers of Plato there should be an interest in what follows, and I hope that anyone who simply enjoys seeing an age-old problem studied in a different context and with new methods will not be disappointed.

[78] Field, *Plato and His Contemporaries*, p. 2.   [79] *Grg.* 486 A–C, 527 A.

8
The Authenticity of the Platonic Dialogues

THE task of classifying samples according to some identifying characteristics which are related to authorship is one which is most easily accomplished by discriminant analysis. It is simply an extension of the method introduced in Chapter 6, where the samples were classified according to the work to which they belonged, and all misclassifications were recorded. In this case the group category is that of authorship and all the samples are allocated initially to the supposed author in each case. The discriminant analysis is then run and all those samples for which the value of the discriminant function does not justify a classification with the named author are listed as being incorrectly classified. In this way we can obtain a good idea of the relative homogeneity of the works of different authors, the adequacy of the discriminant function for isolating anomalous samples, and the extent to which authors differ from one another.

Thus if we take the entire set of 702 samples of this study, consisting of the 493 samples of the Platonic canon and the remaining 209 from the six other authors, placing them in groups according to the nominal author in each case, and then run the discriminant analysis with all 37 variables, the list of misclassifications shown in Table 8.1 is obtained. It must be stressed that the placing of samples from the doubtful works of the Platonic corpus within the Plato group does not necessarily beg the question of authorship. The number of samples involved is small in comparison with the main mass of genuine samples, perhaps not more than 50 out of a total of 493, or approximately 12%, and this should not be enough to cause a large enough alteration of the discriminant function to make it discriminate in favour of material which differs considerably from the genuine samples.

On the other hand, it is necessary to have a clear idea of the theoretical implications of discriminant analysis. We are testing to see whether the samples which we have allocated to each group could belong to the population from which the group was drawn. Thus the Plato group is supposed to be representative of the notional[1] population of Platonic

[1] Notional because the entire corpus of Platonic writings is included, so that the population cannot consist of anything other than this—no other works of Plato are known. We may assume, if we wish, that this set of 493 samples represents a random selection of the vast number of samples that Plato could have written in that style had he desired to do so.

writings and of the characteristics which it exhibits. Are there any samples which seem to fall outside this category and to belong rather to the population of writings by any one of the other six authors?

The answer to this question is the list of misclassified samples in Table 8.1, but it is clear that we cannot simply translate this list into a bold statement of authorship identification, as it would involve us in the absurdity of supposing that Xenophon had a hand in writing parts of at least 10 Platonic dialogues, not to mention Thucydides' *History*, and that Aeschines and Lysias were also responsible for other parts of Plato's works.

The results, therefore, may be interpreted partly as a statement of the extent to which language fails to follow the predicted statistical pattern, either because the variables do not have the multivariate normal distribution which is assumed for them, or because, in a limited number of cases, there is no difference corresponding to that which we assume to exist due to the difference of authorship, at least in terms of the variables which we have chosen to measure.

If we look at the results in more detail we find that the two orators, Isaeus and Isocrates, are least likely to be confused with any of the other five authors. They both have a 100% success rate of classification and no samples from other authors are allocated to them. Aeschines and Lysias come next in order of clarity of discrimination, since all their own work is correctly ascribed to them, but each has managed to acquire three samples from other authors. Then follows Thucydides, who loses three of his 49 samples to Xenophon and Lysias but has five others incorrectly ascribed to him (four Xenophon, one Lysias). Finally, Plato and Xenophon, between whom there is a greater level of confusion, although even for these two authors the rate of successful classification stands at 93.71% and 80% respectively, remarkably high figures given the intractability of the material and the unpredictability of the authors' linguistic habits.

Some of the failures of classification are almost predictable. Thus we would expect there to be some difficulty in differentiating between some samples of Thucydides and Xenophon's *Hellenica*, not because the two authors are obviously close in style, but because of genre attraction and the fact that both authors present speeches reputedly delivered by historical characters, often dealing with similar themes.[2] In addition, Xenophon, one must suppose, was attempting to imitate Thucydides' *History* when he wrote the *Hellenica*, for it is a continuation of book 8 of that work, which breaks off abruptly in mid-sentence, and Xenophon takes up the story at the point where Thucydides' work ceases. It is interesting, therefore, to see that the first three samples of *Hellenica* are classed with Thucydides. This does raise the spectre of the possibility that imitative writing can fool the computer, or the

[2] In *HG*, bk. 1 the only speeches to be found are at vi. 5, vi. 8–11, and vii. 16–33.

TABLE 8.1 Samples listed as misclassified in discriminant analysis on all seven authors, using all variables

Author	Work	Sample	Classified into author	Posterior probability of membership in author						
				Aeschin.	Is.	Isoc.	Lys.	Pl.	Th.	Xen.
Pl.	Alc. 1	6	Xen.	0.0000	0.000	0.0000	0.0000	0.0626	0.0064	0.9310
	Alc. 2	1	Xen.	0.0000	0.0000	0.0000	0.1067	0.0875	0.0001	0.8056
	Amat.	2	Xen.	0.0000	0.0000	0.0000	0.0000	0.0027	0.0132	0.9841
	Criti.	4	Xen.	0.0000	0.0000	0.0000	0.0000	0.3794	0.0632	0.5574
	Ep. 7	8	Xen.	0.0000	0.0000	0.0000	0.0019	0.4778	0.0017	0.5186
	Euthd.	2	Xen.	0.0001	0.0000	0.0000	0.0000	0.3015	0.0131	0.6853
		4	Xen.	0.0000	0.0000	0.0000	0.0000	0.4945	0.0000	0.5054
		8	Aeschin.	0.9985	0.0000	0.0000	0.0000	0.0009	0.0000	0.0006
	Hipparch. 1		Xen.	0.0026	0.0001	0.0000	0.0005	0.2059	0.0001	0.7909
	Ion	2	Xen.	0.0696	0.0000	0.0000	0.0005	0.1945	0.0695	0.6658
	Lg.	602	Aeschin.	0.7803	0.0000	0.0000	0.0003	0.2139	0.0000	0.0056
		1102	Xen.	0.0000	0.0000	0.0000	0.0000	0.1307	0.0010	0.8684
	Hp.Mi.	2	Xen.	0.0001	0.0000	0.0000	0.0000	0.0766	0.0000	0.9233
	Mx.	1	Xen.	0.0000	0.0000	0.0000	0.0020	0.4802	0.0038	0.5140
		3	Th.	0.0000	0.0000	0.0000	0.0035	0.0001	0.9946	0.0018
		4	Lys.	0.1300	0.0173	0.0000	0.4167	0.3791	0.0001	0.0568
	Phdr.	2	Xen.	0.0000	0.0000	0.0031	0.3684	0.0199	0.0936	0.5150
	Plt.	10	Xen.	0.0033	0.0000	0.0000	0.0017	0.4039	0.0024	0.5888
		16	Xen.	0.0000	0.0000	0.0000	0.0001	0.4995	0.0000	0.5004
	Prm.	1	Xen.	0.0000	0.0000	0.0000	0.0000	0.0362	0.0005	0.9634
	Prt.	4	Xen.	0.0000	0.0000	0.0000	0.0012	0.3328	0.0029	0.6631
		9	Xen.	0.0194	0.0000	0.0000	0.0003	0.2837	0.1279	0.5687
Th.	Rep.	11	Xen.	0.0003	0.0000	0.0000	0.0000	0.0542	0.0082	0.9373
		14	Xen.	0.0061	0.0000	0.0000	0.0000	0.2190	0.0003	0.7746
		15	Xen.	0.0000	0.0000	0.0000	0.0005	0.1574	0.0001	0.8419
		62	Xen.	0.0000	0.0000	0.0000	0.0001	0.2520	0.0793	0.6686
		78	Xen.	0.0000	0.0000	0.0000	0.0000	0.2966	0.0431	0.6603
		102	Aeschin.	0.6840	0.0000	0.0000	0.0081	0.1052	0.0002	0.2026
	Smp.	5	Xen.	0.0000	0.0000	0.0000	0.0000	0.3475	0.0005	0.6520
	Tht.	8	Xen.	0.0000	0.0000	0.0000	0.0001	0.2964	0.0002	0.7032
	Ti.	7	Xen.	0.0000	0.0000	0.0000	0.0000	0.0685	0.0034	0.9282
	His.	306	Xen.	0.0000	0.0000	0.0000	0.0000	0.0002	0.3814	0.6185
		310	Lys.	0.0000	0.0000	0.0000	0.8096	0.0002	0.1665	0.0238
		413	Xen.	0.0000	0.0000	0.0000	0.0000	0.0009	0.0277	0.9714
Xen.	HG	101	Th.	0.0000	0.0000	0.0000	0.0000	0.0003	0.9495	0.0502
		102	Th.	0.0000	0.0000	0.0000	0.0000	0.0000	0.5418	0.4582
		103	Th.	0.0000	0.0000	0.0000	0.0000	0.0000	0.7251	0.2749
		110	Th.	0.0000	0.0000	0.0000	0.0000	0.0001	0.6973	0.3026
	Mem.	1	Pl.	0.0000	0.0000	0.0000	0.1420	0.5880	0.0001	0.2699
		2	Lys.	0.0003	0.0000	0.0000	0.8198	0.1356	0.0005	0.0437
		3	Pl.	0.0026	0.0000	0.0000	0.0015	0.5678	0.0170	0.4112
		7	Pl.	0.0000	0.0000	0.0000	0.0000	0.9874	0.0000	0.0126
		27	Pl.	0.0000	0.0000	0.0000	0.0000	0.8098	0.0000	0.1902
		30	Pl.	0.0002	0.0000	0.0000	0.0001	0.5682	0.0111	0.4205
	Oec.	6	Pl.	0.0000	0.0000	0.0000	0.0000	0.8286	0.0000	0.1714
		10	Pl.	0.0000	0.0000	0.0000	0.0000	0.9891	0.0000	0.0109
		15	Pl.	0.0000	0.0000	0.0000	0.0000	0.9422	0.0000	0.0578

statistical method, although I suspect that the problem is more likely to be one of genre confusion, for two of the Thucydidean samples are classed with Xenophon, yet there is no question of the former trying to imitate the latter.

We have to consider the possibility that the discriminant function is being asked to achieve too much in the case of some of these authors. This is likely to be true especially where the demands of genre pull an author in more than one direction and establish affinities with several neighbours, so that it becomes difficult to derive a discriminant function which will exclude stray samples from other authors, should they bear some resemblance to this author, yet which is also sufficiently wide to accept all the samples which do indeed belong to him, and which probably exhibit considerable variety. In the circumstances, it seems surprising that the method works at all, given the innate variability to be found in so many of these authors, for even at the level of individual works it is often difficult to find consistency of style. Dialogues such as the *Protagoras* and the *Gorgias* embrace a huge range of styles, changing from swift, witty *badinage* to intense philosophical speculation, from pastiche of literary criticism to a colourful description of a gathering of Sophists, and from myths of creation explaining the birth of society to the eschatalogical myths which Socrates uses to justify or expand his vision of moral perfection. And that is only to mention two dialogues. If one widens the list to include the *Republic, Parmenides, Cratylus, Phaedo,* and the *Symposium,* it seems impossible that common ground could be found between them, at least anything which is capable of linguistic definition. There is then the further gap to be accommodated between these and the later works, the *Laws, Sophist, Politicus, Philebus, Timaeus,* and *Critias,* all of which differ so much in linguistic structure from the previous group that, apart from certain affinities of subject, one could be forgiven for assuming them to be by a different author.

Equally with Xenophon, at the level of detail which 1000-word samples imply, it would be difficult to find any persistent characteristic which could be used to illustrate the consistency of his writing, and it is highly probable that, if we were not blessed with information that told us otherwise, we would ascribe the authorship of the *Hellenica* to someone other than the author of the *Memorabilia* and *Oeconomicus.*

The problem is, therefore, not to establish whether discriminant analysis is capable of separating authors, since the above example shows that this can be achieved with a high rate of success, but of determining what is implied in the cases of failure, when samples are shown as being misclassified, and of deciding how to use the method to help in solving questions of dubious authorship. Between Plato and Xenophon it is in fact possible to achieve perfect discrimination by using the 'within' covariance matrix, rather than the pooled matrix of the previous example (Table 8.1).[3] All 493 of the Plato

---

[3] See *SAS User's Guide,* pp. 381 ff. 100% is achieved by taking the 493 samples of Plato and

samples are classified with Plato and all 65 Xenophon samples are also correctly placed. But this is probably more of a tribute to the strength of the mathematical model, especially as some of the Plato group may not have been written by Plato. However, I do not propose to enter into a discussion of the technicalities of the distinction between these two methods of discriminant analysis, other than to mention that the latter, using the 'within' covariance matrix, is thought to be a better approach to the problem of discriminating between groups, as it uses the variance to be found within each group as the basis for the discriminant function, rather than that of the entire set of samples, and the former is likely to be more representative of a group's (or author's) individuality than the latter.

However, the fact that we can achieve perfect discrimination between Plato and Xenophon does not in itself prove that all the Platonic corpus is genuine, or indeed that the Xenophon samples are genuine either. What it does show is the mathematical feasibility of the method, and that there is sufficient information contained in the variables to enable us to detect a significant difference between each group, in this case the groups being formed by their classification according to the nominal authorship of each sample. Were there no differences between the variable readings for the samples of each group, the analysis could not construct them out of nothing and form an imaginary separation between the authors, but the absence of any real difference would be indicated by a large number of samples labelled as misclassified.

In fact, the strength of discriminant analysis as a technique may be demonstrated even further by using it to separate the samples into individual works. Even if we use all 702 samples, attempting to classify them correctly into the 52 parent works, the number of samples which the analysis shows us as being, on statistical grounds, misclassified, is only 99, giving an overall success rate for correct classification of samples into their parent works of greater than 85%.[4]

The fact that discrimination between such a wide range of works, some of

---

the 65 of Xenophon. Other authors could be included, but there is a technical limitation imposed on the minimum number of samples allowable for each group. This must not be less than the total number of variables used in the analysis, in this case 37. Consequently, Aeschines, Isaeus, and Lysias cannot be included, or if they are, surplus variables will be docked to bring the number down to the level of the total number of samples of that author in each case. The point to remember is that for *each* author a discriminant function is calculated which separates that author from all the remaining six in the analysis, hence one has as many discriminant functions as there are authors. Where the pooled matrix is used it is the total number of samples which is the limiting parameter on the number of variables to be employed. For the 'within' matrix it is the number of samples in any one group.

[4] In fact since five works (*Eps.* 2, 3, 8, and 13, and *Clit.*) consist of only one sample they cannot be included in the discriminant analysis. This is because no *F*-ratio can be calculated for a single sample, and such works will be labelled as misclassified samples. A truer figure for the rate of misclassification is therefore 94 out of 702 samples, or 13.4%, giving a success rate for correct classification of 86.6%.

which are shown by cluster analysis to be remarkably similar (e.g. the Isocratean speeches shown in Figure 6.3), is, nevertheless, possible, should give us pause in placing too much faith in the results of a discriminant analysis which separates authors, for it is clear that the mathematical process which can do this is extremely powerful and may not reflect what we assume to be an underlying reality of an individuality of style. It is extremely difficult to translate the complexities of a discriminant function into an easily recognizable linguistic difference, to show that the authors are separated for such and such a reason, for the function in each case consists of all 37 variables, each multiplied by a specially calculated coefficient, which ranges typically from -15 to $+145$ in the cases of authorship discrimination described above, plus a constant which is calculated as

$$-0.5\,\bar{X}j'\,S\bar{X}j.$$

Here $\bar{X}j$ is the matrix of mean scores for group j and S is the pooled variance/covariance matrix. This constant has a value of approximately -6800 in the typical example using seven authors, such as in the example shown in Table 8.1. We cannot assume, therefore, that the discriminant function is tractable in the sense that it will give us some understanding of the reasons why these authors or works differ from each other, for the mathematical process is operating at such a level of complexity that an understanding of the linguistic relevance of the figures, the so-called reification of the data, is virtually impossible.

However, such is not the objective of discriminant analysis, which merely seeks to maximize the correct classification of the samples, by optimizing the *F*-ratio. Factor analysis would be a more appropriate method to use if our main concern were to illustrate the relationship between the variables and the separate categories of authorship or work. But in passing I should mention that, despite the complexity of the discriminant function, it is clear that greater importance attaches to the BLETs because the coefficients applied to them exceed those applied to all the other variables by a factor of about five to one. We may infer from this that a large part of the differences between the authors is measurable in terms of inflexional variations.

But this is a digression from the main enquiry, which is an investigation of the possibility of determining authenticity. In Chapter 5, which introduced the techniques of multivariate analysis, I drew the analogy of classification into botanical species, suggesting that literary detection presented similar problems. In retrospect this seems to be an over-simplification, for in the case of botanical or biological species there is an underlying residuum of consistent characteristics traceable ultimately to the structure of the DNA molecule for that species, which determines the inherited characteristics, whereas for an author there is no such physical reality to rely on. Of course, Plato, whether he is 30 or 80, is still in some sense the same man, but the mind

is subject to continuous evolution and what he may have written as a young man is likely to differ stylistically from his later works. We cannot blandly assume that there will be a consistency of style and that certain measurable features will remain unchanged throughout an author's career, for we do not have sufficient knowledge of the psychological processes which are affected by ageing and the impact that this might have on the use of language.

Nevertheless, it is clear from the two examples of authorship discrimination given above, using the pooled or 'within' covariance matrix, that some constant numerical quantity may be derived from the variables which enables us to make reasonably accurate judgements about provenance, even though discriminant analysis used in this way may not be the best approach to solving the problem of authenticity. The limitations of discriminant analysis may be illustrated by artificially constructing a worst-case scenario in which only two works of the two chosen authors are known with certainty to have been written by them, let us say the *Republic* and *Memorabilia*. These two works may then be used as paradigms from which two discriminant functions are calculated, one for each author. These two functions are then applied to all the remaining Platonic works and to the *Oeconomicus* and *Hellenica* of Xenophon, to see how such works are classified, whether correctly or to the wrong author. The results are not nearly as good as for confirmatory discriminant analysis, although they achieve a reasonable degree of success, giving 59% correct classification of the Plato samples and 86% for Xenophon. Using different choices of variables or different works as paradigms it is possible to improve on these figures, increasing Plato's success rate to 74%, but generally speaking an improvement in accuracy for one author's classification is counterbalanced by a decline in the other's.[5]

Of course, a success rate of 74% with samples can mean a much higher rate of correct attribution when applied to an entire work, for then we need only to accept a simple majority verdict. Thus if 15 of the *Gorgias* samples are classified with Plato and only nine with Xenophon we are justified in accepting a Platonic origin for the work, even though the characteristics which have been used to form the discriminant function are not so compelling as to place the verdict beyond any reasonable doubt. We may find that most of the Platonic works are, on this basis, correctly classified, but the presence of so many stray samples is clearly disturbing. It is as difficult to

---

[5] I investigated the effect of using many different combinations of paradigms and variables, using also the 'within' and pooled matrix options alternately. Use of the former (the 'within' covariance matrix) usually gave poor results for Xenophon, especially when the number of variables used was not much less than the number of samples. Thus with *Rep.* (80) and *Mem.* (30) as paradigms, and using the ALETs and BLETs and the 'within' option, all 413 of the Platonic test samples, i.e. works other than *Rep.*, are correctly classified as belonging to the same population as *Rep.* With the 35 samples of *HG* and *Oec.* there is 100% failure, and all are shown as being more akin to *Rep.* than to *Mem.*

explain why the Xenophon samples should stray into the Plato camp as it is to account for the movement in the opposite direction.

Nevertheless, I think that there are several points which ought to be mentioned which will perhaps increase our understanding of the partial failure of discriminant analysis in this context and will help to show why the approach is almost bound to have only limited success. In the first place, we have to consider that, looked at in a traditional way, the problem would not be easy to solve, for almost any human agent would be perplexed, faced with the task of deciding which of the 40 non-attributed works should be assigned to Plato and which to Xenophon on the basis only of a knowledge of the *Republic* and *Memorabilia*. For it must be done on a sample-by-sample basis, not by looking at the works as a whole, but by taking each 1000-word sample individually and making a separate decision in each case, for the computer is not provided with any indication that the samples are linked in any way, and to make the task comparable in human terms we need to impose the same conditions in both cases.

It is difficult, perhaps impossible, to think oneself out of present knowledge and to impose some sort of pristine state in which the only information one has relating to these two authors is that derived from a knowledge of the two paradigmatic works, the *Memorabilia* and the *Republic*. Yet if such were the case, would it really be possible to predict that the author of the *Memorabilia* also wrote the *Hellenica*, or that the Plato who wrote the *Republic* also wrote the *Laws*? The choice must be limited to these two authors alone, and a definite decision in favour of one or the other made for each sample, for the aim of discriminant analysis is not to give a general description of each of the various works, but to allocate the samples definitively according to the author which each one most resembles. I suspect that most experts, either literary, linguistic, or philosophical, would have only limited success.

By looking at the problem in this way it does help us to acquire a better understanding of the limitations of discriminant analysis. In the example we quoted, where the *Memorabilia* and the *Republic* are used as paradigms, the analysis is not really posing the general question as to which works most nearly resemble in style that of either of these two authors, but two distinct uses are being made of the data. The first one corresponds to the question 'What is it that most effectively discriminates between the *Republic* and *Memorabilia*?', and it is precisely in the formation of an answer to this question that the weakness of the approach becomes apparent, for the discriminant function which separates these two works may be totally useless as a general discriminator of the styles of the two authors. It may well be that the local effects which are found only in these two works are highlighted, and that these characteristics are not to be found at all in the remaining works for which we are seeking some attribution, so that allocation of these samples

will just be a hit-and-miss affair. For the second question which the analysis effectively asks is 'Which of the two works do the unattributed samples most closely resemble, on the basis only of the discriminant functions calculated for the *Memorabilia* and the *Republic*?' Yet if those functions are limited in their application and emphasize characteristics which are locally valid for those two works only (a bias which is almost inevitable, since discriminant analysis seeks to maximize the differences between the two works, and these may spring from subject matter or vocabulary as much as from any other general linguistic usage), then it is inevitable that serious errors of allocation will occur.

I have laboured the above example, at the risk of wearying the reader, in response to my own perplexity on discovering that one could not simply take a single work from the output of an author and expect it to produce a species definition of that author's style. What in effect is produced is a species definition of that particular work and even that is restricted to its ability to differentiate that work from whatever else might have been included on that occasion in the analysis. For a literary work does not have the stability of character which we find occurring in species within the natural world, so that each discriminant function is local and related to context and may have only slender connections with authorship characteristics.

In fact, it is doubtful whether MVA could really be of much assistance in deciding authorship if the nearly complete ignorance which I have supposed concerning these two authors were to prevail, so that only the *Memorabilia* and the *Republic* were attributed with certainty. All it could do would be to chart the similarities between various works, but it is evident that we could not rely on it to solve the question of the authenticity of works such as *Theages* and *Hipparchus* when even the provenance of the major dialogues would be the subject of some doubt.

Fortunately, it is not necessary to grapple with a problem of such magnitude, for the main bulk of the dialogues is securely attributed and it is only a dozen or so shorter works which are the subject of contention. But we still have to face the difficulty that if any of the Platonic works are spurious the putative author is not represented in the discriminant analysis. If *Hippias Major*, for example, were not written by Plato, but by some otherwise unknown author, there are no extant examples of that author's work against which to make a comparison. For although it is reasonable to suppose that it will differ in many important respects from the genuine Platonic corpus, if indeed it is a forgery, yet it does not follow that it will be classed as a result with any one of the other six authors whom we have included in the analysis. Whatever may have been its origin there are not many commentators who would claim that it was written by Aeschines or Isaeus, Isocrates, Lysias, Thucydides, or Xenophon. This applies also to all of the potentially spurious dialogues, so that the discriminant analysis as I have presented it in the first

example is loaded in favour of a Platonic authorship of the uncertain works. This must be so because, in the effort to pass off a work as being by Plato, the forger must be credited with some success in making it resemble a genuine Platonic dialogue, at least more so than it resembles something by Xenophon or the others. So that in the environment of these seven authors we should expect that the majority of spurious samples which purport to be by Plato should in fact be classified with his works, rather than with one of the other six authors, unless by some curious stroke of fortune any of these works was indeed written by one of them.

This is more or less what does happen, for only 10 samples from the doubtful dialogues are found to be classified with an alternative author. I give the full list in Table 8.2. The works which I take to be the object of some suspicion are the following: *Alcibiades* 1, *Alcibiades* 2, *Amatores*, *Clitophon*, *Epinomis*, *Epistles* 2, 3, 7, 8, and 13, *Hipparchus*, *Hippias Major*, *Hippias Minor*, *Ion*, *Menexenus*, *Minos*, and *Theages*. These give a total of 60 samples of which 10 are found to lodge with another author (mostly with Xenophon). This is a fairly high percentage, 16.7% to be precise, but it is still below the 20% rate of misclassification of the Xenophon samples, and not enough to allow us to condemn outright any one of the works represented in the list. The overall rate of misclassification for the Plato samples is 6.3%, considerably below that for the potentially spurious dialogues, and it sinks to 5.1% if we include only the assured dialogues of the canon. On the other hand, if we eliminate the *Menexenus* from the list of misclassified samples, a dialogue which is going to be subject to genre estrangement in any environment, we are left with only seven misclassified samples from a total of 56, which

TABLE 8.2   *Samples from the doubtful works of the Platonic corpus classified with an alternative author*

Work	Sample	Classified with
Alc. 1	6	Xen.
Alc. 2	1	Xen.
Amat.	2	Xen.
Ep. 7	8	Xen.
Hipparch.	1	Xen.
Hp.Mi.	2	Xen.
Ion	2	Xen.
Mx.	1	Xen.
Mx.	3	Th.
Mx.	4	Lys.

reduces the percentage rate to 12.5%, a figure not too far removed from the Platonic average.

Generally speaking, it would not appear to be safe to derive any conclusions about the authenticity of these works from the above results, because they are too patchy and they seem to stem more from the inherent difficulty of defining adequately the styles of Plato and Xenophon in such a way as to accommodate all their potential vagaries without at the same time widening the discriminant function to such an extent as to embrace far more samples of the opposing author than native samples from their own works.[6] If we decided that over *Amatores* and *Hipparchus* the clouds of suspicion must hang, because 50% of their samples (they have only two samples each) are classified with Xenophon, then we would also have to look seriously at the status of *Euthydemus* and book 1 of the *Republic*, both of which have three misclassified samples. Since the original number of samples is 10 in the *Euthydemus* and eight in the *Republic*, book 1, the figure is unsatisfactorily high.

For *Euthydemus* the attack perhaps is not too damaging because one of the misclassified samples has been allocated to Aeschines, the other two to Xenophon, and for these latter two the posterior probability figures are only marginally against Plato as author. However, for the *Republic*, book 1, which has only eight samples, the matter is more serious as three of these eight are lodged fairly and squarely with Xenophon. These are shown as *Rep.* 11, 14, and 15 in Table 8.1. Yet, whatever the merits of the argument in favour of giving book 1 of the *Republic* an earlier date than the remaining nine books,[7] it was never intended to imply that the authorship was in doubt. We must ascribe this misclassification to some anomalous qualities in these three samples which lift them away from what is typically Platonic. In the case of *Rep.*11 the fact that it is an introductory section is a partial explanation, since these are often awkwardly characterized in relation to the remaining work, but for the other two we would have to look for some other cause. Probably their abnormality is something to do with the fact that these two samples contain the Thrasymachus episode, but one cannot be more specific without a much more detailed investigation involving these samples alone, a task which cannot be undertaken here.

Finally, I must not leave consideration of the list of misclassified samples of Table 8.1 without mentioning the one sample of Plato which might have some claim to a true classification with an alternative author. This is the second sample of *Phaedrus*, which consists of the speech which Phaedrus

[6] I am using poetic licence here to explain the underlying difficulty which discriminant analysis is confronting, namely that there is a good deal of overlap between the two authors. There is no possibility that the discriminant functions can choose any one of a number of potential values, and that they skip around until the best one is discovered, for there is only one value available in each case, the one which maximizes the *F*-ratio.

[7] Guthrie, *History*, iv. 437.

claims to have acquired from Lysias that very morning (230 E). There are several possibilities to consider:

1. that it is a genuine work by Lysias;
2. that it is based on a work by Lysias but altered by Plato;
3. that it is entirely the work of Plato, imitative of Lysias;
4. that it is entirely by Plato and not intended to be imitative;
5. that it is by some other person unknown.

We find that it is classified with Xenophon, with some indication of affinity to Lysias, the respective posterior probabilities being 0.368 and 0.515. This gives us grounds for concluding that it differs substantially from the main mass of Platonic material, but not necessarily that it is not by him. And, since we do not know which of the five possibilities listed above is true, we must still remain undecided as to whether or not a deliberate attempt to imitate can successfully deceive the computer or the stylometrician. With other examples of Discrim using a subset of samples or variables and either classifying by author or document, this particular sample, *Phdr.* 2, is classified variously with Lysias, or Xenophon, or the *Memorabilia*, or even with Isocrates, but hardly ever with Plato. We may conclude that it is odd and distinctly atypical of Plato, but it would be unwise to make any more definitive statement about its origin based on this evidence.

To sum up and draw together the loose ends of what has so far been discussed: we have attempted to study the problem of authenticity by using discriminant analysis with a bald classification into the known or proposed author, assuming that all the possibly spurious works of the Platonic canon should be listed under Platonic authorship. This has proved to be unsatisfactory mainly for two reasons, the most important being that the problem is more open-ended than such a neat classification implies, as the 17 dialogues and epistles which are open to suspicion might be the work of 17 individual authors. The discriminant analysis makes no provision for this, but only allows for classification with one of the seven authors of the study, all seven of whom may very well use styles which do not match in any way that of a possible intruder into the Platonic corpus, whoever it might be, or whatever work it might be that is suspect. Although it is likely that the discriminant analysis will reveal the oddities among the samples in any of the seven authors, we cannot be sure that this will always occur, especially as the possibly spurious works have been included with Plato and there will be some adjustment of the discriminant function to accommodate them.

If we imagine the samples from each author to be clustering in space, the effect of the discriminant function is to interpose a plane surface[8] which

---

[8] Since we are dealing with more than three dimensions, the correct terms to use in this context are hyperspace and hyperplane. I have used common terms here to make the analogy more obvious.

isolates the samples of one author from all others. In the process it is probable that outliers do not get included, but their classification with a foreign author can be rather unpredictable, for it is difficult to assess what are the most important features which the analysis emphasizes as contributing to the differences between the authors.

The fact that the majority of the samples, over 90%, are correctly classified, does, however, imply that there is some basis for assuming authorship characteristics to exist and to have been successfully measured by the 37 variables.

The second difficulty which has prevented us from using the results of the analysis as a guide to authenticity is the obvious overlap between Plato and Xenophon. Had Plato written nothing until the age of 60, or had Xenophon not written the *Hellenica*, I suspect that the problem would not have loomed so large. But it is evidently difficult to find common ground between the *Hellenica* and the two other works of Xenophon included, the *Memorabilia* and *Oeconomicus*. The result is that Xenophon seems to keep a foot in both camps, being close to Thucydides and to Plato, and the discriminant function is less effective in such a situation. All three of the works included fare equally badly in terms of correct classification, with 20% of the samples in each case being allocated in error either to Thucydides or Plato, with one stray sample going to Lysias. Compared with a 6% rate of error for Plato and less for the other authors, the problem, looked at from this angle, seems to be one of adequately discriminating Xenophon from the other authors, rather than that of identifying the Platonic samples.

However, my concern here is to discover more about the Platonic works, and the enquiry must be orientated towards that end. I will therefore leave the examination of these preliminary results and proceed now to look more closely at some other ways of using discriminant analysis which have so far only been briefly mentioned or not yet introduced.

One way of overcoming the distorting effect of ascribing all of the Platonic samples to one author, a proceeding which seems to beg the question of authorship, is to ascribe to each of the dubious works its own author. This I have done in what follows and I have named the authors conventionally as Author 1, Author 2, etc., and the computer printout lists them as such. In Table 8.3 I give a full list of the works for which each author is supposed to have been responsible.

I have not included *Ion*, *Menexenus*, or *Clitophon*, partly through oversight, partly because what I am attempting here is illustrative rather than definitive, and partly because I believe them to be genuine. I am not intending to draw any firm conclusions from what emerges using these classifications, but merely to show what sort of results are obtained and how they may be used in assessing authenticity.

In the first example of discriminant analysis quoted it was pointed out how

TABLE 8.3   *Ascription of dubious works from the Platonic corpus to neutral 'Authors'*

Author	Work	No. of samples
1	*Thg.*	3
2	*Hipparch.*	2
3	*Amat.*	2
4	*Min.*	2
5	*Hp.Mi.*	4
6	*Alc.* 1	9
7	*Alc.* 2	4
8	*Ep.* 7	8
9	*Ep.* 3	1
10	*Ep.* 8	1
11	*Ep.* 2	1
12	*Ep.* 13	1
13	*Epin.*	6
14	*Hp.Ma.*	8
TOTAL		52

the imposition of a single author on the Platonic samples could distort the results. But in the above case there is also the possibility of distortion, for, just as there is no certainty that Plato wrote all of the 493 samples ascribed to him, it is equally uncertain that 14 separate authors wrote those works which I have listed above. Some may have been by Plato and others not, but we are not in a position to know. Therefore, to pretend that these works differ severally from each other and from Plato and that discrimination between them is possible may be entirely unrealistic and produce strange results in the classification process.

Another alteration that I have made in an attempt to overcome the confusion between Plato and Xenophon experienced in the previous example is to split both of these authors into two parts. The division for Xenophon is based only on the observed genre difference between *Hellenica* and the other two works, a difference which is sufficient to create difficulty in the accurate classification of Xenophon under one heading. I have therefore ascribed the *Memorabilia* and *Oeconomicus* to an author called Xenophon 1 and *Hellenica* to Xenophon 2.

For Plato the division into Plato 1 and Plato 2 is based on observations which are more proper to the following chapter which deals with chronology. Here I must anticipate to reveal that the most striking difference to be found between the Platonic works is that which cluster analysis shows to exist

between early and late works, or, more accurately, between late works and those which are variously described as early, intermediate, or middle. The late works, which the chronological tests of the following chapter confirm, are given below in alphabetical order:

Clitophon
Critias
Laws
Philebus
Politicus
Sophist
Timaeus.

Clitophon is the only work of this list which need cause any surprise, for all the others are well established as late dialogues. But in any case it is a single sample and being only one among many it will not cause any hiccup in the analysis. At this point I will not attempt to justify its inclusion with this set, but defer discussion of its status until later in the chapter. Had it been treated as the work of an additional author its most likely fate would have been to have been classed with this group anyway, since as a single sample it cannot be treated independently as a group, and in the discriminant analyses which classify all samples into their parent works it is usually classed with the *Laws*.

All the above works are ascribed to the author Plato 2. The *Epinomis*, and *Epistles* 3, 7, and 8, which if genuine should belong to this group, have already been separately dealt with under Authors 8, 9, 10, and 13.

All the remaining works of Plato which do not occur in the above list or the list of 14 authors of Table 8.3 are classified under the heading Plato 1.

The purpose of this division is to eliminate some of the difficulties of classification errors between Plato and Xenophon and also to find out in a more general sense what is the relationship between the main parts of the Platonic corpus and the possibly spurious works which have been separately listed under notional authors.

When we run a discriminant analysis on this arrangement of authors using all 37 variables some very interesting cross-classifications are found to occur. But before giving the list of samples thus found to be misclassified there is one small technical point to be disposed of. This is the fact that discriminant analysis can only produce a discriminant function for a group which contains two samples or more.[9] A single sample is not a group, so in the case of the four categories containing the *Epistles* 2, 3, 8, and 13, each labelled under a separate author, no discriminant function can be calculated and they cannot be said to exist in a separate group at all. The result is that each sample in such a category will be listed as misclassified and entered into the group

---

[9] The technical reason is that the F-ratio, which Discrim maximizes, is the ratio of within- to between-group variance. If there is only one sample there can be no within-group variance.

(author) to which the posterior probability shows it to be closest. This gives us an idea of its level of resemblance to the other authors, but, in terms of the other groups, these samples are at a disadvantage, because the process cannot work in the opposite direction, and no samples from any source will be classified with these epistles, so that we cannot form any idea of how well they interlock with the remaining works of Plato.

This will become more apparent as we study the list of misclassifications which I give in Table 8.4. Where appropriate, in the final column, I have given the name of the work into which the sample is classified, rather than the name of the author, in the cases where we are dealing with Authors 1 to 14, who are notionally responsible for only one work each.

It is perhaps rather difficult to take in all this extensive information, and I will attempt to summarize what appears to me to be important. Firstly, in case it is not already evident I should repeat that this is a list of those samples which *statistically* are considered to be incorrectly placed in the work to which they traditionally belong. They are classified instead with the work (or author) to which they are found to be nearest. Thus the *Hellenica* samples 111 and 112 at the bottom of the list are found to be closer to Lysias and Thucydides respectively and therefore are not classed with the *Hellenica* (to which we know they belong) but with these other two authors instead.

The point of the analysis is that if our 14 pseudo-authors are very different from Plato there should not be much of this sort of cross-classification, but if they are similar, many of the Plato samples will be lodged with them.

There is a reduction in the total number of cross-classifications occurring between Plato 1 and 2 and Xenophon 1, the possibility that there would be any confusion between Plato and Xenophon 2, the *Hellenica* group, being fairly remote. In fact the linkage is clearly only between Plato 1 and Xenophon 1, as a glance at the list of misclassified samples will show. Comparing these with the previous set of samples shown in Table 8.1, when a simple seven-author classification was used, we can see that the number of misallocated samples from the *Memorabilia* and *Oeconomicus* has fallen from six and three respectively (total nine) to four and one (total five) and that only two of these latter samples are classed with Plato 1: *Mem.* 3 and *Oec.* 10. In the reverse direction the traffic of samples from the genuine Platonic corpus into the Xenophon category is also approximately halved. Formerly 20 samples from Plato, from works other than those listed under Authors 1 to 14, were reclassified with Xenophon, or 18 if one excludes those from the *Menexenus* and *Ion*. This number falls to 13 in Table 8.4, or 12 if one excludes the one sample from *Ion*.

For the most part it is the same samples which are involved, but those from the Platonic corpus which no longer feature on the Xenophon side are the following: sample 4 from *Critias*, 1102 from the *Laws*, 10 and 16 from the *Politicus*, 9 from *Protagoras*, 78 from the *Republic*, 8 from the *Theaetetus*,

TABLE 8.4   *Author misclassifications using 37 variables*

Author	Work	Sample	Classified into author/work
6	*Alc.* 1	8	Pl. 1
8	*Ep.* 7	4	Pl. 2
9	*Ep.* 3	1	*Ep.* 7
10	*Ep.* 8	1	*Ep.* 7
11	*Ep.* 2	1	*Thg.*
12	*Ep.* 13	1	*Alc.* 1
14	*Hp.Ma.*	2	*Alc.* 1
Pl. 1	*Ap.*	2	*Alc.* 2
		4	*Alc.* 2
	Chrm.	2	*Alc.* 1
	Cra.	2	Pl. 2
		4	*Min.*
		10	*Hp.Ma.*
		14	*Hp.Ma.*
	Euthd.	2	Xen. 1
		4	Xen. 1
		8	*Hp.Mi.*
		9	*Hipparch.*
	Euthphr.	1	*Thg.*
		2	*Thg.*
		4	*Alc.* 1
		5	*Alc.* 1
	Grg.	1	*Hp.Ma.*
		9	*Hp.Ma.*
		13	*Alc.* 1
		15	*Alc.* 1
		16	*Alc.* 1
		24	*Hp.Mi.*
	Ion	2	Xen. 1
		3	*Hp.Mi.*
	La.	2	*Alc.* 1
		4	*Hp.Ma.*
		7	*Hp.Ma.*
	Ly.	3	*Hp.Mi.*
		4	*Alc.* 1
	Men.	2	*Hp.Ma.*
		7	*Alc.* 2
		8	*Alc.* 1
	Mx.	2	*Alc.* 2
		3	Th.
		4	Lys.

TABLE 8.4 *continued*

Author	Work	Sample	Classified into author/work
	Phd.	13	Xen. 1
		15	*Alc.* 1
		17	*Hp.Mi.*
	Phdr.	1	*Alc.* 1
		2	Xen. 1
		8	*Ep.* 7
		14	*Alc.* 1
	Prm.	1	Xen. 1
		11	*Alc.* 1
	Prt.	4	Xen. 1
		7	*Amat.*
		13	*Alc.* 1
		14	*Hp.Ma.*
	Rep.	11	Xen. 1
		13	*Alc.* 1
		14	Xen. 1
		15	Xen. 1
		17	*Alc.* 1
		18	*Alc.* 1
		31	*Alc.* 1
		32	*Alc.* 1
		33	*Alc.* 1
		53	*Alc.* 1
		62	Xen. 1
		102	Xen. 1
		107	*Alc.* 1
		108	*Ep.* 7
	Smp.	5	Xen. 1
	Tht.	14	*Hipparch.*
Pl. 2	*Lg.*	105	*Ep.* 7
		201	*Epin.*
		305	*Epin.*
		504	*Ep.* 7
		907	*Epin.*
		1006	*Ep.* 7
		1207	*Ep.* 7
	Phlb.	1	Pl. 1
		9	*Epin.*
		10	Pl. 1
	Sph.	14	*Epin.*
	Ti.	9	*Ep.* 7
		11	*Ep.* 7
		12	*Epin.*

TABLE 8.4 *continued*

Author	Work	Sample	Classified into author/work
Th.	*History*	301	Lys.
		413	Xen. 1
Xen. 1	*Mem.*	1	*Alc.* 2
		2	Lys.
		3	Pl. 1
		29	*Hp.Mi.*
	Oec.	10	Pl. 1
Xen. 2	*HG*	111	Lys.
		112	Th.

and 7 from the *Timaeus*, a net loss of eight from the 18 of the previous list. It appears that our efforts to reduce the number of errors of classification between Plato and Xenophon have met with some success, but not quite as markedly as we might have hoped. The possibility of confusion still exists between the two authors, and to a lesser extent between other authors, although still affecting only a small percentage of the total number of samples involved.

The addition of 14 other authors has complicated matters still further, but herein lies the main interest of the whole exercise, because the cross-classification between these putative authors and Plato reveals a great deal about the similarities which draw them together, suggesting that in several cases the two authors are identical. It is not so much that the samples from these disputed works fail to classify with the parent work, for that only occurs in three instances (*Hp.Ma.* 2, *Alc.* 1 8, and *Ep.* 7 4), but that samples from Plato in considerable numbers are found to cross the divide and are allocated to one of the 10 works listed under separate authors. (As explained previously, *Epistles* 2, 3, 8, and 13 cannot be included in the analysis as they consist of only one sample each.) While one might expect that an imitative writer would score some success in achieving similarity to the author emulated, there is no reason to suppose that the traffic should be two-way and that the original author should have been found in turn to have imitated the imitator. For that is what appears to have happened, since although these 10 works show considerable individuality, as evinced by the fact that only the three samples quoted are listed as misclassified out of a total of 48, yet 61 of the Plato samples are found to be classed variously with them.

The most interesting cases of misclassification relate to the dialogues *Alcibiades* 1, *Epinomis*, and *Hippias Major*, and to *Epistle* 7. I give in Table 8.5 a list of all the samples involved under the heading of the appropriate dialogue. Results for Xenophon 1 are also given for comparison.

TABLE 8.5   *Samples from the Platonic corpus misclassified into four*
doubtful dialogues and Xenophon 1

Alc. 1
   Chrm. 2, *Euthphr.* 4, 5, *Grg.* 3, 13, 15, 16, *La.* 2, *Ly.* 4, *Men.* 8, *Phd.* 15, *Phdr.* 1,
   Prm. 11, *Prt.* 8, 13, *Rep.* 13, 17, 18, 31, 32, 33, 53, 107
Epin.
   Lg. 201, 305, 907, *Phd.* 109, *Sph.* 14, *Ti.* 12
Ep. 7
   Lg. 105, 504, 1006, 1207, *Phdr.* 8, *Rep.* 108, *Ti.* 9, 11
Hp.Ma.
   Cra. 10, 14, *Grg.* 1, 9, *La.* 4, 7, *Men.* 2, *Prt.* 14
Xen. 1
   Euthd. 2, 4, *Ion* 2, *Phd.* 13, *Phdr.* 2, *Prm.* 1, *Prt.* 4, *Rep.* 11, 14, 15, 62, 102, *Smp.*
   5

It could be argued that this proves nothing about the similarity of these four allegedly spurious dialogues to the genuine Plato, since so many Plato samples are found to resemble Xenophon. *Alcibiades* 1 certainly scores more highly in terms of numbers attracted to itself, having 24 genuine Plato samples classed with it, more than the figure for any other author and almost double the number recorded for Xenophon. But this could be just a chance effect, and who is to say that it has a particular significance?

On the other hand, I am inclined to think that the analysis is weighted in favour of Xenophon, for the Xenophon 1 group contains two works, the *Memorabilia* and *Oeconomicus*, works which are not remarkably close to each other. The total number of samples for the group is 45, and the consequent difficulty of classifying all these samples under one head results in a certain amount of laxity in the discriminant function for this group, in the sense that it is less capable of cutting off outliers from other groups; whereas for the short dialogues such as *Hipparchus* and *Minos* of only two samples each the discriminant function will be much more exclusive and not allow many outsiders into the fold. In fact, there is a definite correlation between the number of samples in the original group and the number of alien samples admitted to that group, as may be seen readily from the figures in Table 8.6.

The correlation coefficient for the two sets of figures for the first nine dialogues given in Table 8.6 is 0.95. Obviously *Alcibiades* 1 and Xenophon 1 do not follow the pattern and that is why I have excluded them.

But with a correlation as high as this it is clear that we cannot draw any conclusions from the number of outside samples alone which are attributed to each of the 14 authors, for the cause may be no more profound than the increasing difficulty of providing an adequate discriminant function as the number of samples in the group rises. So that we cannot say that the *Hippias*

TABLE 8.6   *No. of alien samples classified with*
dubious dialogues

Work	No. of samples	
	In parent work	Aliens
Alc. 2	4	5
Amat.	2	1
Ep. 7	8	8
Epin.	6	6
Hipparch.	2	2
Hp.Ma.	8	8
Hp.Mi.	4	5
Min.	2	0
Thg.	3	2
Alc. 1	9	24
Xen. 1	45	14

Major is more like the genuine Plato than *Theages*, since it has eight of the genuine Plato samples allocated to it as against only two for the latter, and similarly for the other six dialogues, since this is no more than one would expect when the correlation between the two quantities is as high as has been shown.

On the other hand, there is perhaps a certain amount of circularity in this argument, since we are using the figures from the table to calculate the correlation coefficient and then using this coefficient as an argument against the validity of the very same figures. The correlation would not be so high if we were to include the results for *Alcibiades* 1 and Xenophon 1, but there must inevitably be a lack of confidence in any conclusions that we might derive concerning authenticity for any one of the nine works other than those two, even though at first sight they looked rather promising.

This is not to say that nothing can be learned from this misclassification. For the fact that *Alcibiades* 1 does not seem to follow the general pattern but has a much higher proportion of samples from Plato attributed to it than we would expect from the performance of the other doubtful works does at least tell us that, in the race to be considered as a genuine dialogue, it cannot be looked upon as an outsider. If any of these dialogues are to be elevated to the status of genuine works then this one must surely be one of the prime candidates. It cannot be claimed of it that its style is totally unlike that of Plato, for in many cases—24 to be exact if we refer only to this analysis—the stylometric judgement is that the listed samples are more like this work than Plato is like himself.

In addition, it is clear that the rate of misclassification of samples with Xenophon is lower than it should be if it were to follow the same pattern as the 10 uncertain works, for one would expect as many as 45 stray samples from Plato to be classified in this group, yet only 14 are found. This does suggest that there could be a more substantial difference between Plato and Xenophon (or Plato 1 and Xenophon 1) than between Plato and the other putative authors.

It is also noticeable that nearly all the stray samples from Plato 2 are lodged with either *Epistle* 7 or the *Epinomis*, apart from two which are placed with Plato 1. This does tell us something about the character of these two works, for it is precisely in such company that we would expect to find them if they were genuine, so that it already appears that any claims that these two works are grossly unlike Plato or attempt to imitate his style with pitiful incompetence do not fit the facts. It is evident that there is a great deal in common between these two works and the style of Plato's later writing.

For the remaining minor dialogues I give in Table 8.7 a list of the alien samples which are classed with each.

I cannot account for the presence of two samples of the *Apology* with *Alcibiades* 2, as there is no obvious resemblance between the two dialogues. If, as has been claimed,[10] *Alcibiades* 2 is apocryphal, it is strange that it should show any resemblances to the genuine Plato, for one would expect a more emphatic rejection than these figures seem to imply. However, the status of this and the other dubious dialogues will be discussed in greater detail later in the chapter.

TABLE 8.7   *Alien samples classified with* Alcibiades *2,*
Amatores, Hipparchus, Hippias Minor, Minos, Theages

Alc. 2
   Ap. 2, 4, *Mem.* 1*, *Men.* 7, *Mx.* 2
Amat.
   Prt. 7
Hipparch.
   Euthd. 9, *Tht.* 14
Hp.Mi.
   Euthd. 8, *Grg.* 24, *Ly.* 3, *Mem.* 29*, *Phd.* 17
Min.
   Cra. 4
Thg.
   Euthphr. 1, 2

---

* Xen.

[10] Guthrie, *History*, v. 387.

The four single-sample epistles are attributed as follows:

Epistle 2 to *Theages*
Epistle 3 to *Epistle* 7
Epistle 8 to *Epistle* 7
Epistle 13 to *Alcibiades* 1

The attribution of *Epistles* 3 and 8 to the author of *Epistle* 7 does confirm the similarity of style of these three works, whether or not one is prepared to accept them as genuine. For 2 and 13 there is greater scope for doubt, because apart from the uncertainty of authorship of *Theages* and *Alcibiades* 1 it is probable that they are too early to justify an affinity of style with these epistles, as the latter would date from *c*.360, whereas all early dialogues could hardly be later than the middle 380s. However, the similarity may spring from some unexpected features of the epistolary style bringing these two works into contact with characteristics of the early dialogues, and the resemblance may not spring from chronological affinities. Some shades of suspicion, nevertheless, must attach to these two epistles, and the problem is discussed more fully when more evidence has accumulated to enable us to make a more balanced judgement.

I return now to the problem of improving the separation between Plato and Xenophon and thus having a better foundation for deciding the question of authenticity in individual cases. It is possible to select a subset of variables which is directed towards this end, rather than just using the 37 variables *en masse* or any one group of them indiscriminately. The subset is chosen by using a stepwise discriminant analysis on a restricted group of samples. In this case I have chosen to use the Stepdisc on all early and middle works of Plato, excluding any over which the shadow of doubt hangs, plus the *Memorabilia* and *Oeconomicus* of Xenophon. The problem is to select those variables which contribute most to the discrimination between these two authors as defined by the works attributed to them. The Stepdisc will select those variables which contribute most to the maximization of the *F*-ratio at each step of the calculation, adding one variable each time, or taking one away if it no longer makes a significant contribution. One can then use these variables in a subsequent discriminant analysis. I have used the first 10 such variables in a discriminant analysis on all 702 samples with the authors defined as in the previous case, that is by creating 14 additional authors and by splitting both Plato and Xenophon into two parts.

As is to be expected the number of cross-classifications is vastly increased, since with a subset of variables, only 10 in place of the original 37, the discriminant analysis cannot hope to work so well. With the concentration on differentiation between Plato and Xenophon and the consequent neglect of the other five authors, even those who were formerly distinctive and not to be confused with others in the analysis, Isocrates and Isaeus, are thrown into the melting-pot and there is a general mêlée and confusion between all 23

Thg.
  Alc. 1 2,[b] *Cri.* 3, *Euthphr.* 1, 2, *Euthd.* 1, 2, 4, 6, *Grg.* 6, 23, *Lg.* 103, 205, 801, *Men.* 1, *Mx.* 1, *Oec.* 7,[c] *Ora.* 34,[c] *Phd.* 3, 7, 13, *Phlb.* 5, 6, 7, *Prm.* 10, *Prt.* 1, *Rep.* 104, *Smp.* 2, 6, 13, *Sph.* 7, 8, *Tht.* 2, 8, 20, *Tma.* 6[c]

Hipparch.
  Cra. 5, 10, *Euthphr.* 3, *Mem.*[c] 7, 23, 29, *Ora.* 92,[c] *Prt.* 7

Amat.
  Alc. 2 1,[b] *Chrm.* 3, *Grg.* 13, *HG* 112,[c] *His.* 511,[c] *Ly.* 4, *Mem.* 28[c], *Ora.* 82,[c] *Phd.* 2, *Rep.* 13, 26, 28, 39, 62, 72

Min.
  Cra. 3, 4, 6, 15, *Grg.* 19, *Lg.* 404

Hp.Mi.
  Cra. 12, *Grg.* 12, 24, *Ion* 3, *Lg.* 702, *Oec.* 5,[c] *Phd.* 9, 17, *Phdr.* 9, *Phlb.* 1, 14, *Plt.* 8, 15, *Prm.* 2, *Rep.* 35, 75, *Sph.* 4, *Tht.* 18, *Ti.* 10

Alc. 1
  Ap. 1, *Cra.* 1, 8, *Cri.* 2, *Euthphr.* 5, *Grg.* 1, 3, 8, 14, 16, 21, *Hp.Ma.* 7,[b] *La.* 3, 6, 7, *Lg.* 302, 711, 802, *Ly.* 5, *Men.* 6, *Phdr.* 14, *Prm.* 8, 12, 13, *Prt.* 15, *Rep.* 25, 27, 32, 33, 34, 51, 52, 53, 56, 57, 58, 82, 83, 88, 101, 106, *Smp.* 11, 12, *Sph.* 2, 3, 12, *Tht.* 12, 15

Alc. 2
  Ap. 3, 4, 6, *Cri.* 4, *Euthd.* 10, *Grg.* 11, *His.* 308,[c] *Ion* 2, *Lg.* 101, 303, 307, 903, *Mem.*[c] 4, 13, 16, 20, 30, *Men.* 7, 8, 9, *Prt.* 12, *Rep.* 12, 14, 31, 55, *Sph.* 1, *Tht.* 9

Ep. 7
  Ap. 5, *Clit.* 101, *Criti.* 4, *Ep.* 3 1,[b] *Ep.* 8 1,[b] *Ep.* 13 1,[b] *Era.* 1,[c] *His.* 301,[c] *La.* 1, *Lg.* 105, 304, 403, 503, 505, 707, 905, 909, 1006, 1007, 1201, *Mem.* 9,[c] *Men.* 4, *Oec.* 2,[c] *Phd.* 19, *Phdr.* 1, 4, 8, *Plt.* 1, *Rep.* 81, *Tht.* 3, *Ti.* 2, 3, 5

Epin.
  Euthd. 9, *Lg.* 201, 502, 602, 703, 806, 907, 1204, 1205, *Oec.*[c] 6, 8, *Phdr.* 6, 7, *Rep.* 15, 38, *Ti.* 6, 7, 8, 12, 14, 15, 16, 18, *Tma.* 2[c]

Hp.Ma.
  Alc. 1[b] 5, 9, *Amat.* 1,[b] *Cra.* 16, *Grg.* 4, 7, 9, 15, 20, *La.* 5, *Ly.* 3, *Men.* 3, *Phd.* 14, *Phdr.* 10, *Phlb.* 13, *Prm.* 4, 11, *Prt.* 10, *Rep.* 42, 43, 45, 96, *Smp.* 8, 10, 14, *Tht.* 11

Pl. 1
  Alc. 1 1, 8, *Lg.* 106, 305, 402, 805, 1107, *Phlb.* 4, 10, *Sph.* 9, 13

Pl. 2
  Ep. 2 1, *Ep.* 7 4, *Euthd.* 7, *Hp.Mi.* 1, *Ly.* 1, *Mx.* 2, 4, *Oec.*[c] 10, 12, *Ora.* 83,[c] *Phd.* 18, *Prm.* 3, *Prt.* 2, 5, 13, *Rep.* 36, 37, 46, 77, 108, *Smp.* 4, *Tht.* 16, 17

Xen. 1
  Cra. 2, *Epin.* 1,[c] *Era.* 2,[c] *His.* 413,[c] *Lg.* 204, 1207, *Ora.* 52,[c] *Phlb.* 11, 12, 15, *Plt.* 13, *Prm.* 1, *Smp.* 5, *Ti.* 20

Xen. 2
  Criti. 3, *Mem.* 27, *Mx.* 3

[a] In the case of Authors 1–14, the title of the work is given to simplify references.
[b] Misclassifications between any two of the 14 putative authors.
[c] Samples known from historical evidence to belong to an author other than the one with whom the sample is classified (e.g. we know *Tma.* 6 to belong to Aeschin., and not to the author of *Thg.*, whoever that might have been), other than cross-classifications from Pl. to Xen. 1 and 2.

authors (i.e. Authors 1–14, Plato 1 and 2, Xenophon 1 and 2, Aeschines, Isaeus, Isocrates, Lysias, and Thucydides).

I shall concentrate on the misclassifications occurring between Authors 1–14, Plato 1 and 2, and Xenophon 1 and 2. These are listed in Table 8.8.

In addition, there are another 67 misclassifications involving samples classified with one of the five authors, Aeschines, Isaeus, Isocrates, Lysias, and Thucydides. These I give in Table 8.9 in a summarized form.

If we add this subtotal of misclassified samples to the previous list of 300 or so it will be seen that over half of the total number of samples have been reallocated. This exercise may seem to be rather pointless in terms of discriminating between authors, but the main interest of it lies in highlighting the similarities, if any, which are presumed to exist.

The only author who comes out of this operation more or less unscathed is Isocrates. To him is allocated only one stray sample from *Phaedrus*, sample 2, the Lysias speech as it is called. Apart from this there are no other misplacements into the Isocratean canon, and he himself does not suffer any of his own samples to be located with other authors. This does indicate remarkable consistency, especially when so much confusion is found in the surrounding authors and the selection of speeches included spans a considerable portion of his career, with *Panathenaicus* right at the end of his life, and *Panegyricus* having been written some 40 years earlier.

However, the disarray which attends the other authors is rather bewildering, and it obviously requires caution in the attempt to interpret its

TABLE 8.9 *No. of misclassifications into Aeschines, Isaeus, Isocrates, Lysias, and Thucydides*

	Aeschin.	Is.	Isoc.	Lys.	Th.
Aeschin.					1
Author 1		1			
Author 4		1			
Author 8					1
Is.	3				
Lys.	1				
Pl. 1	10	10	1	1	
Pl. 2	6	3		7	
Th.	5			5	
Xen. 1	2			2	2
Xen. 2				3	1
TOTAL	27	15	1	18	5
GRAND TOTAL:	67				

significance. Firstly, it is important to note that the reason for the preponder-
ance of the misplaced Platonic samples is due to the large number of samples
from Plato which are included, outbalancing the other authors by a factor of
about eight to one. Of this larger number of samples a somewhat higher
proportion are found to be misclassified than for other authors. The figures
for rates of misclassification are shown in Table 8.10. Arranged in descending
order of correctness this becomes, as a percentage,

Isocrates	100
Aeschines	75
Xenophon 2	75
Thucydides	69.4
Isaeus	65.2
Xenophon 1	44.4
Plato 2	42.3
Lysias	40
Plato 1	30.6

Between the final four authors there is not a vast difference, except perhaps
for Plato 1 who is nearly 10 points behind the next least successfully
discriminated author.

The question which really needs to be considered is the way in which these
misclassifications are distributed and how important it is that, for example,
Hipparchus has only been allocated four samples from Plato 1, whereas
Amatores has 10. For such difficulties I can only offer a rule-of-thumb
interpretation, as the way in which I am using discriminant analysis is
scarcely orthodox and there are no theoretical guide-lines of which I am
aware for dealing with the problem.

TABLE 8.10   *Misclassified samples, by author*

Author	No. of samples	% Misclassified
Aeschin.	12	25
Is.	23	34.8
Isoc.	56	0
Lys.	5	60
Pl. 1	278	69.4
Pl. 2	163	57.7
Th.	49	30.6
Xen. 1	45	55.6
Xen. 2	20	25

Of the authors with whom Plato 1 and 2 are now confused, Aeschines and
Isaeus have the largest number of misclassified samples (ignoring Authors 1–
14), followed by Xenophon 1 and Lysias. The exact figures are as shown in
Table 8.11.

In this analysis there no longer seems to be the same problem of
correlation of the number of misclassified samples with the size of the work
(or author) accepting them, for, if we use these figures and combine them
with those for the other putative authors, ignoring those which relate to
Xenophon 1 and *Alcibiades* 1, as we did in the previous example, the
correlation coefficient between the two quantities is found to be -0.24, a
figure which shows that there is little connection between the two. We can,
therefore, rely more readily on the absolute number of samples transferred
from Plato to the putative author as a guide to the probability of a work's
authenticity.

Erring on the generous side we could take 10 as the typical figure for the
number of samples of Plato (either Plato 1 or Plato 2 individually) which
might be involved in any cross-authorship classification. This is above the
average figure for the authors shown in Table 8.11. Below this level some sort
of authorship division may be assumed to exist, since such a division would
act as a barrier to the transference of samples, whereas in the absence of this
barrier large numbers of samples might be expected to cross from one to the
other. An identity of authorship would seem to require that at least as many
samples as are observed to cross the divide between Plato 1 or Plato 2 and
any one of the authors whom we know to be distinct should be allocated to
the competing author.

Alternatively, one could require that at least as many samples as are
transferred within the Plato 1 and Plato 2 groups (since these are one author)
should be found to have been allocated from either one of these groups to the
disputed work. From Table 8.8 it may be seen that nine samples are
transferred from Plato 2 to Plato 1 and 17 in the reverse direction, giving an
average of 13, and this could perhaps be the threshold figure for assessing
whether a work is or is not of Platonic origin. But this is perhaps not such a
good guide as the previously suggested alternative, since the choice of

TABLE 8.11   *No. of Platonic samples classified with other authors*

	Aeschin.	Is.	Xen. 1	Lys.	Average
Pl. 1	10	10	3	2	6.25
Pl. 2	6	3	7	7	6.0
TOTAL	16	13	10	9	12.0

variables was determined by the decision to emphasize the differences between works of Plato 1 and Xenophon 1, the later works of Plato being left out of the calculation. Nevertheless, it is clearly better to err on the side of caution rather than to ascribe a work to Plato on slender evidence.

From the list of misclassified samples of Table 8.8 it appears that the prime candidates for works of Platonic authorship are those shown in Table 8.12. The figures for *Alcibiades* 2 are high enough to warrant its inclusion in the works which have a good chance of being genuine, but the most disturbing feature of the misclassifications relating to this dialogue is the presence of five samples from the *Memorabilia*. This is a far higher proportion of the Xenophon 1 group than the 16 samples of Plato attributed to this dialogue are of the Plato 1 group. It appears to be no closer to Plato than it is to Xenophon, but perhaps occupies an intermediate place in stylometric terms between the two authors. On the other hand, if it is a later pastiche and uses language quite unlike that employed by Plato, as has been claimed, it is surprising that this is not more apparent in its isolation from the surroundings, just as Isocrates has been isolated by his distinctive style.

Amatores is possibly a borderline case with 10 samples from Plato 1 ascribed to it, and since six of these are from the *Republic*, the central work of Plato's life, one is disposed to be more in favour of accepting it. Nevertheless,

TABLE 8.12 *Prime candidates for Platonic authorship demonstrated by no. of samples from Plato 1 and 2*

Work	Pl. 1	Pl. 2	Total
Alc. 1	41	6	47
Thg.	23	8	31
Ep. 7	9	17	26
Hp.Ma.	24	0	24
Epin.	5	16	21
Alc. 2	16	5	21
Hp.Mi.	11	7	18
Amat.	10	0	10
Min.	5	1	6
Hipparch.	4	0	4

Note: the table is arranged in descending order of no. of samples from Pl. transferred into each dialogue. Works in the top group are likely to be genuine, according to these figures, although subsequent evidence suggests that *Hp.Mi.* is suspect.

the number of Platonic samples it has attracted is low, no greater than Aeschines, Isaeus, or Xenophon, and that is perhaps an adequate reason for rejecting it.

As for the other five works, the evidence of similarity to the Platonic dialogues is, if not overwhelming, quite impressive. *Alcibiades* 1 is especially notable in having attracted the largest number of Platonic samples of any work, including 16 from the *Republic*, or 20% of the whole. This is extraordinary for a work which has been considered to be entirely imitative. It is also noteworthy in not having trapped any foreign samples from Xenophon (or from any other author except possibly from *Hippias Major*), so that the inference of an entirely Platonic character for this work seems to be justifiable.

Theages also is fairly convincing, with such a large number of samples attributed to it, although it is itself only a short work, and on this evidence it appears to be genuine. The distribution of samples between early and late works is somewhat disturbing, however, for one would not anticipate that it would have much in common with works contemporary with the *Laws*.

No such objection may be made against *Hippias Major*, however, for all of the 24 samples attributed to it have come from the Plato 1 group, a confirmation that its style is closest to that of Plato's earlier period. *Hippias Minor* is a borderline case, since it does not differ emphatically from those that follow, or indeed from Aeschines, Isaeus, and Xenophon 1 in terms of the samples attracted to it—only 18 compared with Aeschines' 16—and we would require additional evidence either way before being confident about taking a decision.

As for *Epinomis* and *Epistle* 7 it is clear that they both have strong affinities with the group of later Platonic works, and, this being so, one would expect their overall score to be slightly low, since the Plato 2 group is smaller than the earlier Plato 1 group. It is to the later period that these two works should belong and the evidence so far definitely points in that direction.

Of course, it could be argued that all that has been shown is that some authors are more successful imitators than others. In which case it would imply that the authors of *Alcibiades* 1, *Theages*, *Epistle* 7, *Epinomis*, and *Hippias Major* are more like Plato than Plato is like himself, for there is a greater volume of traffic in the form of misclassified samples between these five authors and Plato than there is between Plato 1 and Plato 2. The question really hinges on how we choose to define likeness, and how we measure it, and, finally, on how close two things have to be stylistically before they can be pronounced to be by the same author. These questions are not easy to answer—indeed, I do not believe that there is an unequivocal answer to any of them. I shall examine in more detail the question of proximity shortly. But for the present a few more examples of discriminant analysis may be helpful.

As stated earlier it is possible to separate all the 48 works (52 if one counts the single-sample epistles) into groups, with an 80% rate of success of classification, even allowing for such similar groups as the Isocratean orations, where one would expect discrimination between individual speeches to be difficult. It is therefore not entirely surprising that one can achieve discrimination between authors, even though the notional grouping of samples under the heading of a single author may not be in accord with the reality. There may be some mistakes in the traditional allocations. The evidence of cross-classifications between the works might be of some assistance, therefore, in indicating hidden relationships which are based on authorship groupings.

Using all 702 samples with the 37 variables of the full set and running a discriminant analysis to classify all samples with the parent work, as stated, about 20% of the samples go astray. Some of these are from Isocratean works, six samples in all, which are shifted from one Isocrates speech to another. However, I shall deal here with results which are of more direct interest to us, giving the cross-classifications between the Platonic works, including the Spuria, and the Xenophon works only. These cross-classifications are given in Table 8.13.

Examples such as these give us some information about the relationships between certain works, but they do not tell us much about authorship except in the extreme cases, such as Isocrates, where the cross-classifications (six in all) are entirely confined to the one author, involving movement of samples within the Isocratean corpus; or in a more general sense by indicating the innate variability of some authors, with diverse characteristics arising in many works, making the problem of classification troublesome and the consequent probability of error fairly high. If no knowledge were available about the authors, but we had only the works themselves to use as our source of information, it is doubtful if we could arrive at an accurate assessment of who was responsible for which work, because the system of cross-classification between works is fairly complex and probably in many cases is only partially affected by authorship differences.

It is important to realize that, when the discriminant analysis is asked to classify samples according to the parent work, authorship characteristics are no longer relevant, indeed they may be a positive hindrance to the task of separation of one work from another. This is because significant differences must be found for each work individually, and the fact that many works were written by one author could imply that substantial differences do not exist, except perhaps to a limited extent in subject matter. The fact that the process is, nevertheless, quite successful testifies to the mathematical strength of discriminant analysis, but whether or not this will help to identify authors is rather doubtful.

The fact that 52 separate works by at least seven different authors can be

TABLE 8.13   *Cross-classifications between the Platonic corpus and Xenophon works*

Doubtful dialogues

Alc. 1	*Ep.* 13 1, *Grg.* 15, *Rep.* 18
Alc. 2	*Mem.* 1
Ep. 7	*Ep.* 3 1, *Lg.* 1005, 1006, *Rep.* 108
Epin.	*Lg.* 305
Hp.Mi.	*Grg.* 24

Amat., Hipparch., Hp.Ma., Ion, Min., Mx., Thg.: no samples

Xenophon works

Mem.	*Prm.* 1, *Phdr.* 2, *Prt.* 4, *Rep.* 31

Oec.: no samples

Remaining Plato works

Ap.	*Smp.* 12, *Tht.* 1
Chrm.	*Ep.* 2 1, *Men.* 5, *Prm.* 3, *Prt.* 7, *Rep.* 24
Cra.	*Phd.* 18
Euthd.	*Ly.* 2, *Rep.* 14
Grg.	*Hp.Ma.* 2, *Mem.* 7
La.	*Men.* 2
Lg.	*Clit.* 1, *Ep.* 8 1, *Sph.* 11, *Ti.* 1, 2
Ly.	*Rep.* 17, 27
Men.	*Grg.* 16
Phd.	*Alc.* 1 8, *Euthd.* 3, *Phdr.* 6, *Rep.* 15, 16, 46, 91, 96, 106, *Smp.* 10
Plt.	*Lg.* 708, *Sph.* 3
Prt.	*Grg.* 23, *Phd.* 13, *Rep.* 11, 102, *Smp.* 2
Rep.	*Chrm.* 1, *Phdr.* 13
Smp.	*Phd.* 1, 11, 19, *Prm.* 2, *Prt.* 11, *Rep.* 17, *Tht.* 3
Tht.	*Men.* 6, *Phdr.* 5, *Phd.* 6, *Prt.* 15, *Rep.* 32, 33
Ti.	*Plt.* 5, 16

Criti., Cri., Euthphr., Prm.: no samples

reasonably effectively classified with a success rate of about 80% shows that there is a large amount of information in the variables, and the most relevant part of this for discrimination between works is selected by the mathematical process of calculating a discriminant function for each work. In theory the requirement of discriminating between any two works is of equal importance, and the computer cannot tell that we attach greater weight to differentiating between the *Republic* and *Memorabilia* than we do to separating *Phaedo* from the former. From our point of view, for the purposes of the present enquiry, it would not matter if all the main undisputed Platonic works formed a hazy mass, impossible to split into component parts, with no foreign contamination from outside authors. We could then easily distinguish the outliers and say which of the potentially spurious works were

indeed false, and which belonged to the main corpus. To a certain extent this can be achieved by initially using the discriminant analysis with the classification according to authorship specified at the outset, and each sample being allocated to the author whom knowledge or common tradition indicated. This approach was found to be defective because it did not allow enough outlets for the potentially spurious works and by implication it defined normality according to the large mass of genuine and mainstream works, whereas in all probability the works of uncertain origin, even if genuine, would be peripheral and not have the characteristics defined by the central samples.

The present approach is more concerned with establishing the relationships between all these works, and seeing which ones most resemble each other, but in doing so the authorship differentials are jettisoned. The search for differences between works which may be stylistically very similar may cause the discriminant function to emphasize variables which drown the differences between authors. In such circumstances it is often found that dialogues by Plato, for example, are closer to the *Memorabilia* or the *Oeconomicus* of Xenophon, because certain features have been emphasized which bring them into proximity, or because the individuality which belongs to a work due to the particular style of the author who wrote it is only one of the sources of variance within that work and may be weak in comparison with the other forces which have shaped its language. A general set of variables, especially if it is large, will measure the resultant of all these forces and give us an overall picture of the relationships between all the works included, but this picture will probably conflict with our preconceived judgement of the way in which they should relate, since this judgement is based for the most part on a knowledge of authorship.

The above cross-classifications are not very instructive, other than as confirmation of some of the affinities which have already been noted, between *Epistle* 7 and the *Laws* for example, and between *Alcibiades* 1 and some samples from the main works. The four single-sample epistles are classified as follows:

Epistle 2	*Charmides*
Epistle 3	*Epistle* 7
Epistle 8	*Laws*
Epistle 13	*Alcibiades* 1.

This information will be useful later.

The cross-classifications between Xenophon and Plato seem to have been reduced still further, with only a core of four samples from Plato now classed with the *Memorabilia*, from the *Parmenides*, *Phaedrus*, *Protagoras*, and the *Republic*. Three of these are the familiar samples which have caused trouble in the past, the opening section of the *Parmenides*, the 'Lysias' speech of

Phaedrus, and the Prometheus myth of *Protagoras*. The fourth sample, *Rep.* 31 is a relative newcomer, as it does not appear in the first list of misclassified samples when the division was made into seven authors. It is the first part of book 3 of the *Republic* and deals with immorality in the writings of the poets. It contains several quotations and for some reason it appears to be closer to the *Memorabilia* than to the main body of the *Republic*.

We cannot derive any information about authenticity from the fact that the *Amatores*, *Hippias Major*, *Hipparchus*, *Ion*, *Menexenus*, *Minos*, and *Theages* have no external samples attributed to them, because the same is true of *Critias*, *Crito*, *Euthyphro*, and *Parmenides*, all genuine dialogues.

Perhaps we could learn more about the interconnections of the various dialogues by artificially making the task of discrimination more difficult, thus increasing the number of cross-classifications. This is done quite simply by reducing the number of variables. Using only the BLET variables possibly gives a more purely linguistic assessment of the relationships between the dialogues, since these are the variables which relate directly to word endings, and, Greek being a highly inflected language, grammatical and syntactical preferences are likely to be emphasized by these measurements. However, one should not assume that such measurements are necessarily a more adequate guide to authorship differentials than anything else which we might choose to measure, since it is an open question how effectively linguistic habits might be measured by such observations. This subject is discussed more fully in Chapter 10. For the present we are more concerned with the results than the theory, and the results, in terms of cross-classifications between works, are given in Table 8.14 for a discriminant analysis which classifies samples according to the parent work using the BLET variables only. I have presented the results in summary form because the huge number of misclassified samples makes compression desirable.

In this table a Plato sample is defined as any sample which belongs to the traditional Platonic corpus, and I have not separated samples from the potentially spurious dialogues under a separate heading. They account for only 16 out of the 285 samples of Plato which are redistributed, a relatively harmless addition. Consequently, some of the figures given for the Plato transference to individual dialogues will be slightly inflated, especially for those who consider that all these dialogues are spurious.

The dialogues which are found to have most alien samples from the Plato group classified with them are given in Table 8.15. We could find in this table table evidence for accepting *Epistle* 7, the *Amatores*, *Epinomis*, and *Menexenus* as genuine, but the whole thing is rather confusing because several dialogues which belong to the main stream of Platonic writing do not feature in the list, notably the *Republic* and *Gorgias*, with five and three external samples respectively classified with them. However, one cannot conclude from this that these dialogues are unlike anything else that Plato ever wrote,

TABLE 8.14   *Misclassification into doubtful works resulting from discriminant analysis on all samples using* BLETs *only**

Alc. 1	4 Pl., 1 *Aeschin.*
Alc. 2	6 Pl., 1 *Mem.*
Amat.	14 Pl., 1 *Oec.*
Ep. 7	17 Pl., 1 *Clit.*, 1 *HG*, 2 *Mem.*
Epin.	12 Pl.
Hipparch.	8 Pl., 1 *HG*
Hp.Ma.	7 Pl. 1 *Oec.*
Hp.Mi.	7 Pl., 2 *Mem.*
Ion	8 Pl., 1 *Oec.*
Min.	1 Pl.
Mx.	9 Pl., 3 *Mem.*
Th.	4 Pl., 1 *Mem.*
Mem.	8 Pl., 1 *Oec.*
Oec.	4 Pl., 1 *Aeschin.*, 1 *Mem.*
Ap.	4 Pl.
Cri.	4 Pl., 1 *Mem.*, 1 *Oec.*
Criti.	4 Pl.
Grg.	3 Pl., 1 *Mem.*
La.	9 Pl., 2 *Oec.*
Men.	5 Pl., 1 *Mem.*, 1 *Oec.*
Phdr.	10 Pl., 1 Is.
Plt.	15 Pl.
Rep.	5 Pl., 1 *Oec.*
Smp.	6 Pl., 1 *Mem.*
Ti.	5 Pl.

* For comparison the figures for Xen. and for a few of the Platonic dialogues are also given.

TABLE 8.15   *Misclassifications from discriminant analysis using* BLETs *only, showing dialogues accepting the most Platonic samples*

Work	No. of samples	Work	No. of samples
Chrm.	21	*La.*	9
Ep. 7	18	*Mx.*	9
Plt.	15	*Prt.*	9
Amat.	14	*Sph.*	9
Phd.	13	*Ion*	8
Epin.	12	*Hipparch.*	8
Phlb.	11	*Mem.*	8
Phdr.	10	*Hp.Ma.*	7
Euthd.	9	*Hp.Mi.*	7

for the simple reason that more of the samples from both of them are classified with other dialogues. Thus the *Gorgias* loses 22 out of its 24 samples to other dialogues, a percentage loss of 91.7%, and the *Republic* loses 69 out of 80, or 86.3%. The only other work to exceed these percentages is *Protagoras*, which has all 15 of its samples classified with other dialogues. This may be an indication of the variability of these works, implying that it is impossible to define a discriminant function which will classify adequately a high proportion of the samples from them into the parent work.

By contrast *Charmides*, which has the highest number of external samples allocated to it (21), loses only three of its seven samples to other dialogues, and the *Apology* and *Epistle* 7 only one each (from eight). These latter two results may be due to genre discrimination, since the *Apology* is not a dialogue, being a speech purported to have been given by Socrates in his defence against the charge of impiety, and *Epistle* 7 is nominally in epistolary form although evidently intended for a wider audience than those directly addressed by it. In genre it does differ considerably from dialogue, so that it is all the more surprising to find *Epistle* 7 so high on the list of works containing samples transferred from other dialogues of the Platonic corpus. Despite the barrier of genre many individual samples, especially from the *Laws*, are found to be closer to this epistle than they are to the dialogue from which they originated.

The fact that the *Amatores* and *Epinomis* are also high on the list disposes one to think of them in favourable terms as candidates for Platonic authorship. Nevertheless, the results are perplexing in other ways, for what is one to make of the low position of *Alcibiades* 1 and *Theages*, which each have only four transferred samples, whereas in the previous example, with the 14-Author classification and a specially selected discriminatory set of variables, they were at the head of the list of works attracting transferred samples from the corpus of genuine works (Table 8.8)?

I am inclined to accept the conclusion that the use of this set of variables, the BLETs, although they do have a specious linguistic importance, is no more relevant to the task of author discrimination than any other set of randomly selected variables. Although it is useful to have information about the interrelationships of all these works one must be cautious in attempting to translate such information into an authorship classification scheme.

The general difficulty of arriving at any firm conclusion by this method is perhaps best illustrated by an examination of the results for Isaeus. All 23 samples from this author are listed as though originating from one work, so that in a sense it is an authorship classification, even though the samples are taken from nine different speeches. We find that only 10 out of these 23 samples survive inclusion in this category, the rest being allocated variously. Thus the two samples of *Oration* 1 are attributed to the *Antidosis* of Isocrates, two samples of *Oration* 2 to Aeschines, and *Oration* 9 to *Hipparchus*, the last

being one of the dubious dialogues attributed to Plato. Single samples from other orations (3, 5, 6, 8, and 11) are classed with *Phædrus*, Aeschines, and Lysias. While it is easy to see that the genre affinities of the oratorical works are bound to create some confusion, and the connection between *Oration* 1 of Isaeus and the Isocratean corpus has already been pointed out (Chapter 6), the link between *Oration* 9 and *Hipparchus* is difficult to explain, unless one wishes to maintain an identity of authorship, a proposition which, though theoretically possible, I would not wish to defend too stoutly.[11]

However, the method does succeed in revealing a diversity of linkages which would be difficult to find by any other means, except perhaps by neighbour analysis, which obtains its results on a simple calculation of the distance of each sample from the various groups. Discriminant analysis, however, introduces an additional element of optimization of the separation of all the works from each other, and its strength lies in this approach rather than in the mere fact of separation, which could be attempted without any enhancement of the raw data.[12]

The conflicting results obtained with different uses of the variables does suggest that if it is authorship discrimination which is paramount then some attempt must be made to select variables according to some criterion of authorship, otherwise the variance arising from other sources will mask the effects we are seeking, or random noise will drown them.

But it does appear that some more information is to be gleaned from a study of those works which are closest to each other, because if it may be shown that some of the suspected works are actually far closer in style to works such as the *Republic* and *Laws* than any other of the accepted genuine works of the canon, then such evidence, taken in conjunction with what has already been discovered, could easily tip the balance in favour of authenticity.

Mahanalobis distances between all 52 works are given in Table 8.16. These are based on mean scores calculated from the readings for all samples in each work. As explained in Chapter 5, Mahanalobis distances take account of the correlations between the variables, and in this way compensate for redun-

---

[11] *Hipparch.* is dismissed by most scholars as being spurious, although I ultimately leave the question of its authenticity unsolved. I suppose that it is not impossible for it to have been written by Isaeus, since he was contemporary with the Academy, but it is difficult to envisage why he should want to do so. Generally speaking the Isaeus speeches stick together well as a group. It is only at this level, using only a few variables which are probably not so relevant to the characteristics of Isaean style, that the classification breaks down. For the authenticity of *Hipparch.* see Guthrie, *History*, v. 389.

[12] A similar scatter of samples is found using neighbour analysis, but I have found it less useful as a technique. There is a distinct correlation between the number of alien samples which a work will have attributed to it and the length of that work, a correlation which makes interpretation difficult. In addition, it does not use any enhancement method to maximize the differences between works, so that individual characteristics are far less pronounced and the scatter of samples in all directions leads to great confusion.

dancy of information. These Mahanalobis distances may be obtained as a by-product of discriminant analysis or Candisc. Measurements for the single-sample works *Clitophon* and *Epistles* 2, 3, 8, and 13 are also included. I propose to use these figures to look at all the works individually and to decide which of them are closest to one another and which most distant, attempting to assess at the same time the relevance of such information both in a general sense and in relation to the individual dialogue. One word of caution is necessary, however, for a high score in itself should not be taken as an absolute guarantee of a vast distance between two works, without some reference to the associated probability figures, for often, even though the measurement may appear to be superficially high, there is a distinct possibility that it could have occurred by chance. This is especially true of the shorter works, for, if we consider that these are samples drawn from a large, homogeneous population, then there is a far greater likelihood that two individual samples taken at random will differ by a large amount than that two groups of 20 samples each, both drawn at random from the same population, will differ to that extent. Thus a distance of 3.78 between the *Politicus* and *Sophist* (16 and 15 samples) is significant, whereas that of 8.31 between *Minos* and *Meno* (2 and 9 samples) is not. Nevertheless, I shall attach some importance to a string of large differences between any one dialogue and most of the other Platonic works, even though individually the differences may not be significant.

Lack of space prevents the inclusion of the full list of probabilities associated with these Mahanalobis distances.

I now proceed to give individual assessments of each work, or group of works, according to the accumulated evidence, taking into account also the historical background. Works by authors other than Plato are also considered, although not in such great detail as the Platonic dialogues. The problem of authentication of works by any one author cannot be dealt with in isolation but must be related also to the evidence of proximity or homogeneity to be found in other authors and to the evidence of kinship between them.

Isocrates

All five of the Isocratean speeches, *Antidosis*, *Archidamus*, *De Pace*, *Panathenaicus*, and *Panegyricus* illustrate well a grouping caused by common authorship. The homogeneity of the Isocratean works has already been well demonstrated by the previous examples of discriminant analysis. Here the differences between the speeches range from 2.93 (*Archidamus* + *De Pace*) to 4.58 (*Archidamus* + *Antidosis*), whereas for all the other works a distance lower than 7.0 between them and Isocrates is nowhere to be found. The average separation between one of these five speeches and any other non-Isocratean work is of the order of 10, Isocrates being distinctive in this

TABLE 8.16 *Mahanalobis distances between all 52 works derived from canonical discriminant analysis using 37 variables*

Document	Alc. 1	Alc. 2	Amat.	Ant.	Ap.	Archd.	Chrm.	Clit.	Cra.
Alc. 1.		5.4598	6.1920	10.4619	5.1212	12.0400	4.5384	8.5023	5.3229
Alc. 2	5.4598		6.5840	8.9786	5.0517	10.3591	6.5356	8.6556	5.6787
Amat.	6.1920	6.5840		10.5425	7.1829	12.1207	6.0470	9.1849	7.3859
Ant.	10.4619	8.9786	10.5425		10.2914	4.5837	10.9186	8.7027	11.1140
Ap.	5.1212	5.0517	7.1829	10.2914		11.5667	6.4653	8.6529	6.2700
Archd.	12.0400	10.3591	12.1207	4.5837	11.5667		12.0916	9.9157	12.2668
Chrm.	4.5384	6.5356	6.0470	10.9186	6.4653	12.0916		8.1142	6.4278
Clit.	8.5023	8.6556	9.1849	8.7027	8.6529	9.9157	8.1142	.	8.7005
Cra.	5.3229	5.6787	7.3859	11.1140	6.2700	12.2668	6.4278	8.7005	.
Cri.	4.5960	5.6907	7.1919	10.7148	5.7403	11.9536	5.6094	7.8823	6.1198
Criti.	9.3824	8.6348	9.9551	9.4058	10.1881	9.7224	9.9882	7.6523	8.3770
Ep. 2	6.4891	7.9963	7.2894	11.6104	7.3340	12.5584	5.6856	9.4404	8.2888
Ep. 3	8.3255	9.0663	9.6923	9.7666	9.8677	10.1676	9.3988	8.6955	9.4531
Ep. 7	6.7597	6.4225	7.7981	8.4696	6.9638	9.5065	7.2426	6.2772	7.6279
Ep. 8	9.3640	8.5244	9.1389	8.9940	9.8058	9.8409	9.5926	7.9091	9.5599
Ep. 13	5.3204	7.1190	8.2269	11.3771	7.1107	12.3142	6.8454	9.0526	7.2020
Epin.	8.1144	7.2326	8.5067	9.1179	8.3595	9.5800	8.1748	7.3507	7.4545
Era.	7.8870	5.9609	8.8959	7.2193	7.6512	7.9008	8.6935	7.6672	8.5458
Euthd.	4.7942	6.4873	6.2100	10.2494	5.3875	11.4525	4.6770	7.6422	6.5055
Euthphr.	4.5858	7.2696	8.4391	12.1579	6.8252	13.6430	6.7354	10.7038	6.6907
Grg.	3.5180	5.4988	5.5835	10.0057	5.2181	11.9893	4.5779	8.2456	5.2755
H.G.	8.0584	7.3631	8.8771	9.9073	8.6534	10.1095	9.0421	8.3310	8.3115
Hipparch.	6.6146	7.2070	7.3432	10.2973	7.6844	12.2873	7.0670	9.3784	6.8645
His.	7.6755	6.6472	7.6298	9.1722	7.8851	9.5936	8.4599	7.7634	7.2831
Hp.Ma.	4.8976	6.6533	7.7152	11.4134	6.3050	12.5918	6.0498	9.3040	4.6667
Hp.Mi.	5.1982	6.4001	7.1650	11.0635	6.6505	12.3591	5.7238	9.0834	6.1982
	Alc. 1	Alc. 2	Amat.	Ant.	Ap.	Archd.	Chrm.	Clit.	Cra.
Ion	5.3689	7.1343	7.0547	10.8141	7.0048	12.5744	6.2429	9.3122	6.7959
La.	4.2668	5.7803	7.1628	10.0919	4.4672	11.2984	6.1734	8.6718	5.3756
Lg.	7.1233	6.9278	7.7280	7.7355	7.6354	8.7016	7.7437	6.2014	7.0702
Ly.	5.0379	6.6614	6.9565	11.3879	5.8836	12.8770	5.3813	9.5610	7.0355
Mem.	5.0245	5.0698	5.2282	9.3922	5.7746	10.6619	5.8186	7.5167	5.8403
Men.	3.6677	4.7983	6.4479	10.6324	4.5969	12.1230	4.8575	8.3549	5.2020
Min.	7.6343	7.3758	9.0378	12.2543	8.1785	13.4735	9.0681	11.1547	5.9583
Mx.	7.4297	5.8261	8.0916	8.7661	6.2932	9.0308	7.4770	7.5908	7.0933
Oec.	5.4632	6.5294	5.7590	10.2086	5.9895	11.6820	5.3445	7.9655	6.1150
Ora.	8.7207	7.4983	9.5884	7.8335	7.7876	9.0541	8.6185	7.9684	8.9214
Pac.	11.8801	10.0269	12.1158	3.6449	11.6693	2.9303	12.1482	9.8036	12.2088
Pan.	11.6783	9.9215	12.1087	4.1228	11.6572	3.3772	11.8890	9.2106	11.9118
Panth.	11.1793	9.8601	11.6258	3.0319	11.1657	3.9262	11.5700	9.4117	11.7942
Phd.	4.1939	5.9975	6.3831	10.9926	5.4502	11.9790	4.1458	8.0179	4.7180
Phdr.	3.8201	4.9283	5.9742	9.4659	5.1001	10.6666	4.3520	7.0190	5.2252
Phlb.	5.8999	5.8295	7.7959	7.9019	6.5234	8.8579	6.7771	6.9104	6.0458
Plt.	7.9399	7.2103	8.2411	7.6576	8.7082	8.7155	8.8252	7.5066	7.3323
Prm.	7.2680	8.7514	9.0741	13.0228	8.2610	13.7727	7.5341	10.6710	7.7783
Prt.	4.3925	5.4369	6.0550	9.8490	4.8395	11.1453	4.7488	7.5914	4.6591
Rep.	3.8503	5.4202	5.5645	9.3522	5.1812	10.6699	3.9147	7.2524	5.7842
Smp.	3.8871	5.9320	6.1303	10.1826	4.9024	11.2253	4.4668	7.6259	5.1679
Sph.	7.0738	6.8761	7.7278	8.5732	7.8390	9.6816	7.5347	7.2966	6.3867
Thg.	4.6587	6.0484	6.6069	11.0927	5.4819	12.5260	5.8035	9.8660	6.6770
Tht.	4.1345	5.2873	6.3922	10.6582	4.6607	11.4817	4.3857	7.5277	4.7270
Ti.	8.6666	8.1213	9.3752	9.2962	9.3505	9.6962	8.7420	6.9612	7.5941
Tma.	7.6574	7.7577	8.2112	7.5931	7.6698	8.9189	8.1658	7.0504	8.1928
	Alc. 1	Alc. 2	Amat.	Ant.	Ap.	Archd.	Chrm.	Clit.	Cra.

TABLE 8.16 continued

Document	Cri.	Criti.	Ep. 2	Ep. 3	Ep. 7	Ep. 8	Ep. 13	Epin.	Era.
Alc. 1.	4.5960	9.3824	6.4891	8.3255	6.7597	9.3640	5.3204	8.1144	7.8870
Alc. 2	5.6907	8.6348	7.9963	9.0663	6.4225	8.5244	7.1190	7.2326	5.9609
Amat.	7.1919	9.9551	7.2894	9.6923	7.7981	9.1389	8.2269	8.5067	8.8959
Ant.	10.7148	9.4058	11.6104	9.7666	8.4696	8.9940	11.3771	9.1179	7.2193
Ap.	5.7403	10.1881	7.3340	9.8677	6.9638	9.8058	7.1107	8.3595	7.6512
Archd.	11.9536	9.7224	12.5584	10.1676	9.5065	9.8409	12.3142	9.5800	7.9008
Chrm.	5.6094	9.9882	5.6856	9.3988	7.2426	9.5926	6.8454	8.1748	8.6935
Clit.	7.8823	7.6523	9.4404	8.6955	6.2772	7.9091	9.0526	7.3507	7.6672
Cra.	6.1198	8.3770	8.2888	9.4531	7.6279	9.5599	7.2020	7.4545	8.5458
Cri.		9.4752	7.0138	8.8672	6.8664	9.5029	6.6423	8.3670	8.2647
Criti.	9.4752		11.0914	8.1665	5.9709	7.9655	9.1629	5.7477	7.2671
Ep. 2	7.0138	11.0914		9.6066	8.0240	10.0398	6.9971	9.5826	9.7588
Ep. 3	8.8672	8.1665	9.6066		6.5336	7.7207	7.9575	7.5711	8.3830
Ep. 7	6.8664	5.9709	8.0240	6.5336		5.3750	6.6595	4.4952	5.5303
Ep. 8	9.5029	7.9655	10.0398	7.7207	5.3750		9.6994	6.3768	7.5935
Ep. 13	6.6423	9.1629	6.9971	7.9575	6.6595	9.6994		7.9805	8.4949
Epin.	8.3670	5.7477	9.5826	7.5711	4.4952	6.3768	7.9805		6.8278
Era.	8.2647	7.2671	9.7588	8.3830	5.5303	7.5935	8.4949	6.8278	
Euthd.	5.7860	8.6755	6.3715	8.6775	6.6101	9.5357	6.7310	7.4924	7.8024
Euthphr.	6.1707	10.7511	8.3546	9.8978	8.5500	11.1064	7.7285	9.2940	9.4217
Grg.	4.5053	9.5904	7.0115	9.1512	7.1335	9.3131	6.9020	8.3745	8.4532
H.G.	8.1819	7.1223	9.6956	9.2000	6.3716	8.1597	8.6346	8.2529	6.1122
Hipparch.	7.6309	9.4199	9.2040	9.9624	8.4817	10.7218	9.3811	8.8364	9.3278
His.	7.0758	7.3048	9.6359	8.6531	6.3301	8.5128	8.6141	7.2539	5.9474
Hp.Ma.	5.5123	9.3315	8.4372	9.9391	8.4600	10.7986	7.0844	9.0104	9.0233
Hp.Mi.	5.8182	9.6637	7.8822	8.8450	7.5233	9.6686	6.6746	8.7112	8.5896
Ion	5.4578	9.9789	7.9663	9.7018	8.3159	10.5927	7.9201	9.3154	9.2141
La.	5.0860	9.4746	7.4413	9.2368	7.3031	9.5319	6.2476	8.6689	8.2944
Lg.	7.0302	5.2687	8.8451	6.8132	3.3460	5.3642	7.7409	3.6056	6.2587
Ly.	6.0600	10.2965	6.9610	10.1496	7.9711	10.3479	7.9881	9.2190	9.2983
Mem.	4.8940	8.0078	7.3927	7.8493	5.2572	7.5705	7.0837	6.2564	6.4058
Men.	4.4045	9.7777	6.8776	10.0945	7.1897	9.9455	6.5193	8.3557	8.1455
Min.	8.7129	10.6316	10.0380	10.5723	9.4166	10.1467	9.3534	9.4714	10.4795
Mx.	6.7486	8.3711	9.1502	9.5062	6.4692	8.1735	8.5763	7.5084	5.7462
Oec.	5.1453	9.1114	6.6435	8.6869	6.8175	9.1716	6.7781	7.8332	8.2903
Ora.	8.9689	8.9970	10.0051	10.2099	7.7320	9.3301	10.1347	9.2981	6.3875
Pac.	11.9247	9.7581	12.9891	10.4533	9.6921	9.8911	12.4524	9.6854	7.6990
Pan.	11.5543	9.0764	12.9946	10.3002	9.4976	9.7347	12.3589	9.6058	7.5707
Panth.	11.3001	9.2842	12.4546	9.6714	8.8793	9.5786	11.8267	9.5481	7.4058
Phd.	4.4137	9.1721	7.2026	9.3503	6.7703	9.3291	6.7703	7.3815	7.9947
Phdr.	4.8239	7.1924	6.4725	7.3141	4.3090	7.2426	5.7575	5.6885	6.1879
Phlb.	6.3518	6.0765	8.4702	7.3135	4.8224	7.7322	6.9305	4.5170	6.3606
Plt.	8.6304	5.0647	10.2529	8.4418	5.5504	7.6512	8.1453	5.3062	6.1990
Prm.	8.4949	10.9955	9.0816	11.1440	8.6824	10.6908	8.9392	9.4909	10.4033
Prt.	4.6765	8.6627	6.2640	8.6694	6.4085	9.2565	6.5097	7.3430	7.6104
Rep.	4.6248	8.2869	6.5888	8.1772	5.3925	8.1035	6.6549	6.9892	6.9078
Smp.	4.8284	8.7897	6.4157	8.5116	6.0642	8.8170	6.1802	7.0743	7.7659
Sph.	7.9758	5.7246	9.1931	8.2536	5.0477	7.5824	7.7708	4.3157	6.9752
Thg.	5.8297	10.7683	6.2783	9.6630	7.4715	9.8923	6.5789	8.8590	8.4159
Tht.	4.5975	8.5550	6.9900	8.7086	6.0398	8.9363	6.1697	6.4726	7.2398
Ti.	9.2218	4.0927	9.8972	7.5438	4.9136	6.8088	8.3949	4.7487	6.7480
Tma.	8.4720	7.7812	9.4265	9.0372	7.0017	8.7505	9.2203	8.0500	5.8529

TABLE 8.16 continued

Document	Euthd.	Euthphr.	Grg.	H.G.	Hipparch.	His.	Hp.Ma.	Hp.Mi.	Ion
Alc. 1.	4.7942	4.5858	3.5180	8.0584	6.6146	7.6755	4.8976	5.1982	5.3689
Alc. 2	6.4873	7.2696	5.4988	7.3631	7.2070	6.6472	6.6533	6.4001	7.1343
Amat.	6.2100	8.4391	5.5835	8.8771	7.3432	7.6298	7.7152	7.1650	7.0547
Ant.	10.2494	12.1579	10.0057	9.9073	10.2973	9.1722	11.4134	11.0635	10.8141
Ap.	5.3875	6.8252	5.2181	8.6534	7.6844	7.8851	6.3050	6.6505	7.0048
Archd.	11.4525	13.6430	11.9893	10.1095	12.2873	9.5936	12.5918	12.3591	12.5744
Chrm.	4.6770	6.7354	4.5779	9.0421	7.0670	8.4599	6.0498	5.7238	6.2429
Clit.	7.6422	10.7038	8.2456	8.3310	9.3784	7.7634	9.3040	9.0834	9.3122
Cra.	6.5055	6.6907	5.2755	8.3115	6.8645	7.2831	4.6667	6.1982	6.7959
Cri.	5.7860	6.1707	4.5053	8.1819	7.6309	7.0758	5.5123	5.8182	5.4578
Criti.	8.6755	10.7511	9.5904	7.1223	9.4199	7.3048	9.3315	9.6637	9.9789
Ep. 2	6.3715	8.3546	7.0115	9.6956	9.2040	9.6359	8.4372	7.8822	7.9663
Ep. 3	8.6775	9.8978	9.1512	9.2000	9.9624	8.6531	9.9391	8.8450	9.7018
Ep. 7	6.6101	8.5500	7.1335	6.3716	8.4817	6.3301	8.4600	7.5233	8.3159
Ep. 8	9.5357	11.1064	9.3131	8.1597	10.7218	8.5128	10.7986	9.6686	10.5927
Ep. 13	6.7310	7.7285	6.9020	8.6346	9.3811	8.6141	7.0844	6.6746	7.9201
Epin.	7.4924	9.2940	8.3745	8.2529	8.8364	7.2539	9.0104	8.7112	9.3154
Era.	7.8024	9.4217	8.4532	6.1122	9.3278	5.9474	9.0233	8.5896	9.2141
Euthd.		6.1582	4.7272	8.4611	6.1947	7.7644	5.6444	6.1936	6.3683
Euthphr.	6.1582		5.4381	9.8975	7.2001	9.3712	6.1920	6.4940	6.4742
Grg.	4.7272	5.4381		8.4660	5.2047	7.6355	4.5044	4.9634	4.5510
H.G.	8.4611	9.8975	8.4660		9.4346	4.3808	8.3678	7.9975	9.2173
Hipparch.	6.1947	7.2001	5.2047	9.4346		8.3940	6.5143	6.7968	6.1666
His.	7.7644	9.3712	7.6355	4.3808	8.3940		7.7522	7.7839	7.9091
Hp.Ma.	5.6444	6.1920	4.5044	8.3678	6.5143	7.7522		5.6118	5.6864
Hp.Mi.	6.1936	6.4940	4.9634	7.9975	6.7968	7.7839	5.6118		5.4962
Ion	6.3683	6.4742	4.5510	9.2173	6.1666	7.9091	5.6864	5.4962	
La.	5.7201	6.3020	4.2590	7.9513	7.5046	7.6935	4.3493	5.8648	6.0478
Lg.	6.8525	8.5953	7.1385	7.1230	8.4365	6.3261	8.2643	8.1266	8.1462
Ly.	5.1454	6.2902	4.7975	8.8586	6.9293	9.0916	6.0660	6.3879	6.7641
Mem.	5.1294	6.5285	4.8193	6.3116	6.2381	4.8290	6.2774	5.3076	5.9779
Men.	5.4129	5.5347	3.6579	8.3176	6.4817	7.8572	5.1999	5.8186	5.9179
Min.	9.0152	8.3621	7.8788	10.0970	8.6073	9.6380	8.1580	7.4899	8.2486
Mx.	7.4929	9.4403	7.0795	5.9219	9.0421	5.2682	7.0081	6.8115	7.5082
Oec.	4.7577	6.7793	4.4136	7.9531	6.6331	6.7093	5.7837	4.4570	5.6834
Ora.	7.6837	10.1187	8.1897	8.4705	8.1209	8.0588	8.8840	8.4961	8.7511
Pac.	11.6812	13.4619	11.7130	10.4809	12.0862	10.0794	12.5710	12.4506	12.6844
Pan.	11.3320	13.3006	11.5679	9.9099	11.8551	9.5267	12.0365	11.9051	11.9936
Panth.	11.0245	12.8064	11.0511	9.8689	11.1691	9.2192	11.7169	11.5428	11.4483
Phd.	4.6182	6.0020	4.2678	7.9329	7.3023	6.7963	4.8141	5.1655	6.1378
Phdr.	4.2608	5.6782	4.0331	6.6961	6.2241	6.1670	5.4656	5.1880	5.4099
Phlb.	5.5678	7.4135	6.3277	7.8039	7.4243	6.7849	6.7070	7.4104	7.6149
Plt.	7.8677	9.4836	8.1716	7.4085	8.7916	6.9246	8.6592	8.9165	9.3513
Prm.	8.3423	9.0646	8.1365	10.1211	9.8425	9.9528	7.8553	9.0644	9.5671
Prt.	4.1507	6.0072	3.3894	7.7102	5.8580	6.5721	4.6064	5.0236	5.0226
Rep.	4.2471	6.4861	3.9875	7.1029	6.6989	6.6858	5.5183	5.5249	6.4432
Smp.	4.5375	5.9441	3.7879	7.7186	6.6851	6.6844	4.8384	4.7368	5.2452
Sph.	6.5401	8.5261	7.2663	8.1756	7.8786	7.3229	7.8467	8.8760	8.9129
Thg.	5.9118	5.0177	5.1749	9.1926	8.1235	8.4991	6.8811	5.9541	6.6158
Tht.	4.5171	6.0637	4.4728	7.7977	6.6326	6.8701	4.9862	5.8884	6.3746
Ti.	7.8634	10.1985	8.9362	7.2188	8.9335	7.4470	9.0522	8.6550	9.7814
Tma.	6.6092	8.9756	7.3841	7.4749	7.2554	6.8483	8.1205	7.8337	7.6000
	Euthd.	Euthphr.	Grg.	H.G.	Hipparch.	His.	Hp.Ma.	Hp.Mi.	Ion

TABLE 8.16 continued

Document	La.	Lg.	Ly.	Mem.	Men.	Min.	Mx.	Oec.	Ora.
Alc. 1.	4.2668	7.1233	5.0379	5.0245	3.6677	7.6343	7.4297	5.4632	8.7207
Alc. 2	5.7803	6.9278	6.6614	5.0698	4.7983	7.3758	5.8261	6.5294	7.4983
Amat.	7.1628	7.7280	6.9565	5.2282	6.4479	9.0378	8.0916	5.7950	9.5884
Ant.	10.0919	7.7355	11.3879	9.3922	10.6324	12.2543	8.7661	10.2086	7.8335
Ap.	4.4672	7.6354	5.8836	5.7746	4.5969	8.1785	6.2932	5.9895	7.7876
Archd.	11.2984	8.7016	12.8770	10.6619	12.1230	13.4735	9.0308	11.6820	9.0541
Chrm.	6.1734	7.7437	5.3813	5.8186	4.8575	9.0681	7.4770	5.3445	8.6185
Clit.	8.6718	6.2014	9.5610	7.5167	8.3549	11.1547	7.5908	7.9655	7.9584
Cra.	5.3756	7.0702	7.0355	5.8403	5.2020	5.9583	7.0933	6.1150	8.9214
Cri.	5.0860	7.0302	6.0600	4.8940	4.4045	8.7129	6.7486	5.1453	8.9689
Criti.	9.4746	5.2687	10.2965	8.0078	9.7777	10.6316	8.3711	9.1114	8.9970
Ep. 2	7.4413	8.8451	6.9610	7.3927	6.8776	10.0380	9.1502	6.6435	10.0051
Ep. 3	9.2368	6.8132	10.1496	7.8493	10.0945	10.5723	9.5062	8.6869	10.2099
Ep. 7	7.3031	3.3460	7.9711	5.2572	7.1897	9.4166	6.4692	6.8175	7.7320
Ep. 8	9.5319	5.3642	10.3479	7.5705	9.9455	10.1467	8.1735	9.1716	9.9301
Ep. 13	6.2476	7.7409	7.9881	7.0837	6.5193	9.3534	8.5763	6.7781	10.1347
Epin.	8.6689	3.6056	9.2190	6.2564	8.3557	9.4714	7.5084	7.8332	9.2981
Era.	8.2944	6.2587	9.2983	6.4058	8.1455	10.4795	5.7462	8.2903	6.3875
Euthd.	5.7201	6.8525	5.1454	5.1294	5.4129	9.0152	7.4929	4.7577	7.6837
Euthphr.	6.3020	8.5953	6.2902	6.5285	5.5347	8.3621	9.4403	6.7793	10.1187
Grg.	4.2590	7.1385	4.7975	4.8193	3.6579	7.8788	7.0795	4.4136	8.1897
H.G.	7.9513	7.1230	8.8586	6.3116	8.3176	10.0970	5.9219	7.9531	8.4705
Hipparch.	7.5046	8.4365	6.9293	6.2381	6.4817	8.6073	9.0421	6.6331	8.1209
His.	7.6935	6.3261	9.0916	4.8290	7.8572	9.6380	5.2682	6.7093	8.0588
Hp.Ma.	4.3493	8.2643	6.0660	6.2774	5.1999	8.1580	7.0081	5.7837	8.8840
Hp.Mi.	5.8648	8.1266	6.3879	5.3076	5.8186	7.4899	6.8115	4.4570	8.4961
	La.	Lg.	Ly.	Mem.	Men.	Min.	Mx.	Oec.	Ora.

	La.	Lg.	Ly.	Mem.	Men.	Min.	Mx.	Oec.	Ora.
Ion	6.0478	8.1462	6.7641	5.9779	5.9179	8.2486	7.5082	5.6834	8.7511
La.	.	7.1788	6.2533	6.0898	4.0144	7.9685	6.0551	5.4821	8.2465
Lg.	7.1788	.	8.4421	5.4145	7.5653	8.9053	6.5332	6.8799	8.2391
Ly.	6.2533	8.4421	.	6.4090	5.7965	8.4819	8.3681	6.4413	8.9268
Mem.	6.0898	5.4145	6.4090	.	5.4345	7.7060	5.9329	4.0929	8.1607
Men.	4.0144	7.5653	5.7965	5.4345	.	8.3141	6.8463	5.5896	8.6822
Min.	7.9685	8.9053	8.4819	7.7060	8.3141	.	9.1822	8.8750	10.2524
Mx.	6.0551	6.5332	8.3681	5.9329	6.8463	9.1822	.	7.4464	6.9696
Oec.	5.4821	6.8799	6.4413	4.0929	5.5896	8.8750	7.4464	.	8.6293
Ora.	8.2465	8.2391	8.9268	8.1607	8.6822	10.2524	6.9696	8.6293	.
Pac.	11.4466	8.8348	12.6143	10.7549	11.9535	13.5920	9.5238	11.8474	9.2453
Pan.	11.0995	8.5525	12.7390	10.5717	11.7742	13.2155	8.7216	11.5648	8.4573
Panth.	10.6174	8.1732	12.2876	10.1341	11.4717	13.0762	8.7471	10.9980	8.3390
Phd.	5.2928	6.7189	5.3722	4.4992	4.6783	8.0298	6.2000	4.6553	8.5819
Phdr.	5.1061	4.9164	5.3408	3.7587	4.5201	7.8010	5.7079	4.6794	7.5054
Phlb.	6.1064	3.8585	7.3855	5.7744	6.3877	8.4317	6.4124	6.7004	7.4352
Plt.	7.9251	4.3344	9.3200	7.0325	8.2767	9.4138	7.4047	8.1089	7.7259
Prm.	7.5229	8.8748	8.2352	8.2038	7.4949	10.1025	9.1259	8.9919	11.0385
Prt.	4.7294	6.5181	5.7336	4.3455	4.0187	7.9603	5.9923	3.8120	8.0644
Rep.	4.9832	5.7348	4.3720	4.3137	4.6978	8.4490	5.9432	4.8311	7.4468
Smp.	4.7037	6.2654	5.7726	4.4894	4.3960	7.8412	5.9110	4.4651	8.0437
Sph.	7.4697	4.0984	8.3579	6.5794	7.3050	9.3260	7.7880	7.7197	8.3470
Thg.	5.8422	8.0411	5.9451	5.5903	5.1524	8.0693	8.1176	5.5814	8.9375
Tht.	4.8795	6.2667	5.7287	4.6097	3.3541	8.4013	5.9809	5.2568	8.4242
Ti.	8.9107	4.8092	9.7704	7.3035	8.9992	9.7810	7.8080	8.2979	7.8306
Tma.	7.8173	6.7866	8.3481	6.5852	8.3666	9.4499	6.7064	7.3383	4.7959
	La.	Lg.	Ly.	Mem.	Men.	Min.	Mx.	Oec.	Ora.

TABLE 8.16 continued

Document	Pac.	Pan.	Panth.	Phd.	Phdr	Phlb.	Plt.	Prm.
Alc. 1.	11.8801	11.6783	11.1793	4.1939	3.8201	5.8999	7.9399	7.2680
Alc. 2	10.0269	9.9215	9.8601	5.9975	4.9283	5.8295	7.2103	8.7514
Amat.	12.1158	12.1087	11.6258	6.3831	5.9742	7.7959	8.2411	9.0741
Ant.	3.6449	4.1228	3.0319	10.9926	9.4659	7.9019	7.6576	13.0228
Ap.	11.6693	11.6572	11.1657	5.4502	5.1001	6.5234	8.7082	8.2610
Archd.	2.9303	3.3772	3.9262	11.9790	10.6666	8.8579	8.7155	13.7727
Chrm.	12.1482	11.8890	11.5700	4.1458	4.3520	6.7771	8.8252	7.5341
Clit.	9.8036	9.2106	9.4117	8.0179	7.0190	6.9104	7.5066	10.6710
Cra.	12.2088	11.9118	11.7942	4.7180	5.2252	6.0458	7.3323	7.7783
Cri.	11.9247	11.5543	11.3001	4.4137	4.8239	6.3518	8.6304	8.4949
Criti.	9.7581	9.0764	9.2842	9.1721	7.1924	6.0765	5.0647	10.9955
Ep. 2	12.9891	12.9946	12.4546	7.2026	6.4725	8.4702	10.2529	9.0816
Ep. 3	10.4533	10.3002	9.6714	9.3503	7.3141	7.3135	8.4418	11.1440
Ep. 7	9.6921	9.4976	8.8793	6.7703	4.3090	4.8224	5.5504	8.6824
Ep. 8	9.8911	9.7347	9.5786	9.3291	7.2426	7.7322	7.6512	10.6908
Ep. 13	12.4524	12.3589	11.8267	6.7703	5.7575	6.9305	8.1453	8.9392
Epin.	9.6854	9.6058	9.5481	7.3815	5.6885	4.5170	5.3062	9.4909
Era.	7.6990	7.5707	7.4058	7.9947	6.1879	6.3606	6.1990	10.4033
Euthd.	11.6812	11.3320	11.0245	4.6182	4.2608	5.5678	7.8677	8.3423
Euthphr.	13.4619	13.3006	12.8064	6.0020	5.6782	7.4135	9.4836	9.0646
Grg.	11.7130	11.5679	11.0511	4.2678	4.0331	6.3277	8.1716	8.1365
H.G.	10.4809	9.9099	9.8689	7.9329	6.6961	7.8039	7.4085	10.1211
Hipparch.	12.0862	11.8551	11.1691	7.3023	6.2241	7.4243	8.7916	9.8425
His.	10.0794	9.5267	9.2192	6.7963	6.1670	6.7849	6.9246	9.9528
Hp.Ma.	12.5710	12.0365	11.7169	4.8141	5.4656	6.7070	8.6592	7.8553
Hp.Mi.	12.4506	11.9051	11.5428	5.1655	5.1880	7.4104	8.9165	9.0644
Ion	12.6844	11.9936	11.4483	6.1378	5.4099	7.6149	9.3513	9.5671
La.	11.4466	11.0995	10.6174	5.2928	5.1061	6.1064	7.9251	7.5229
Lg.	8.8348	8.5525	8.1732	6.7189	4.9164	3.8585	4.3344	8.8748
Ly.	12.6143	12.7390	12.2876	5.3722	5.3408	7.3855	9.3200	8.2352
Mem.	10.7549	10.5717	10.1341	4.4992	3.7587	5.7744	7.0325	8.2038
Men.	11.9535	11.7742	11.4717	4.6783	4.5201	6.3877	8.2767	7.4949
Min.	13.5920	13.2155	13.0762	8.0298	7.8010	8.4317	9.4138	10.1025
Mx.	9.5238	8.7216	8.7471	6.2000	5.7079	6.4124	7.4047	9.1259
Oec.	11.8474	11.5648	10.9980	4.6553	4.6794	6.7004	8.1089	8.9919
Ora.	9.2453	8.4573	8.3390	8.5819	7.5054	7.4352	7.7259	11.0385
Pac.		3.1792	3.7514	12.0684	10.7088	8.9007	8.5275	13.9391
Pan.	3.1792		3.3694	11.7802	10.4362	8.5892	8.3145	13.8984
Panth.	3.7514	3.3694		11.5473	10.0509	8.3452	8.2449	12.9759
Phd.	12.0684	11.7802	11.5473		3.9330	5.6996	7.8550	6.7742
Phdr.	10.7088	10.4362	10.0509	3.9330		4.6470	6.3557	7.4536
Phlb.	8.9007	8.5892	8.3452	5.6996	4.6470		4.4146	8.2455
Plt.	8.5275	8.3145	8.2449	7.8550	6.3557	4.4146		10.0716
Prm.	13.9391	13.8984	12.9759	6.7742	7.4536	8.2455	10.0716	
Prt.	11.2952	11.0268	10.6425	3.6942	3.5288	5.6346	7.7062	8.0916
Rep.	10.5839	10.4671	9.9481	3.3193	3.2659	5.0597	6.7718	7.0616
Smp.	11.4966	11.1073	10.7343	3.0624	3.3565	5.4880	7.7281	7.2220
Sph.	9.6348	9.5913	9.2499	6.8433	5.5153	3.4226	3.7794	8.5633
Thg.	12.5113	12.5891	12.0389	5.3577	5.0980	7.1777	8.7227	8.1101
Tht.	11.5588	11.4215	11.1449	3.2317	3.1160	5.2100	7.3461	6.7035
Ti.	9.9353	9.2439	9.4149	8.2003	6.3494	5.6500	4.9312	9.7346
Tma.	9.0814	8.5110	8.2677	7.6100	6.0831	6.6622	6.3460	10.7020
	Pac.	Pan.	Panth.	Phd.	Phdr.	Phlb.	Plt.	Prm.

TABLE 8.16 continued

Document	Prt.	Rep.	Smp.	Sph.	Thg.	Tht.	Ti.	Tma.
Alc. 1.	4.3925	3.8503	3.8871	7.0738	4.6587	4.1345	8.6666	7.6574
Alc. 2	5.4369	5.4202	5.9320	6.8761	6.0484	5.2873	8.1213	7.7577
Amat.	6.0550	5.5645	6.1303	7.7278	6.6069	6.3922	9.3752	8.2112
Ant.	9.8490	9.3522	10.1826	8.5732	11.0927	10.6582	9.2962	7.5931
Ap.	4.8395	5.1812	4.9024	7.8390	5.4819	4.6607	9.3505	7.6698
Archd.	11.1453	10.6699	11.2253	9.6816	12.5260	11.4817	9.6962	8.9189
Chrm.	4.7488	3.9147	4.4668	7.5347	5.8035	4.3857	8.7420	8.1658
Clit.	7.5914	7.2524	7.6259	7.2966	9.8660	7.5277	6.9612	7.0504
Cra.	4.6591	5.7842	5.1679	6.3867	6.6770	4.7270	7.5941	8.1928
Cri.	4.6765	4.6248	4.8284	7.9758	5.8297	4.5975	9.2218	8.4720
Criti.	8.6627	8.2869	8.7897	5.7246	10.7683	8.5550	4.0927	7.7812
Ep. 2	6.2640	6.5888	6.4157	9.1931	6.2783	6.9900	9.8972	9.4265
Ep. 3	8.6694	8.1772	8.5116	8.2536	9.6630	8.7086	7.5438	9.0372
Ep. 7	6.4085	5.3925	6.0642	5.0477	7.4715	6.0398	4.9136	7.0017
Ep. 8	9.2565	8.1035	8.8170	7.5824	9.8923	8.9363	6.8088	8.7505
Ep. 13	6.5097	6.6549	6.1802	7.7708	6.5789	6.1697	8.3949	9.2203
Epin.	7.3430	6.9892	7.0743	4.3157	8.8590	6.4726	4.7487	8.0500
Era.	7.6104	6.9078	7.7659	6.9752	8.4159	7.2398	6.7480	5.8529
Euthd.	4.1507	4.2471	4.5375	6.5401	5.9118	4.5171	7.8634	6.6092
Euthphr.	6.0072	6.4861	5.9441	8.5261	5.0177	6.0637	10.1985	8.9756
Grg.	3.3894	3.9875	3.7879	7.2663	5.1749	4.4728	8.9362	7.3841
H.G.	7.7102	7.1029	7.7186	8.1756	9.1926	7.7977	7.2188	7.4749
Hipparch.	5.8580	6.6989	6.6851	7.8786	8.1235	6.6326	8.9335	7.2554
His.	6.5721	6.6858	6.6844	7.3229	8.4991	6.8701	7.4470	6.8483
Hp.Ma.	4.6064	5.5183	4.8384	7.8467	6.8811	4.9862	9.0522	8.1205
Hp.Mi.	5.0236	5.5249	4.7368	8.8760	5.9541	5.8884	8.6550	7.8337
	Prt.	Rep.	Smp.	Sph.	Thg.	Tht.	Ti.	Tma.
Ion	5.0226	6.4432	5.2452	8.9129	6.6158	6.3746	9.7814	7.6000
La.	4.7294	4.9832	4.7037	7.4697	5.8422	4.8795	8.9107	7.8173
Lg.	6.5181	5.7348	6.2654	4.0984	8.0411	6.2667	4.8092	6.7866
Ly.	5.7336	4.3720	5.7726	8.3579	5.9451	5.7287	9.7704	8.3481
Mem.	4.3455	4.3137	4.4894	6.5794	5.5903	4.6097	7.3035	6.5852
Men.	4.0187	4.6978	4.3960	7.3050	5.1524	3.5941	8.9992	8.3666
Min.	7.9603	8.4490	7.8412	9.3260	8.0693	8.4013	9.7810	9.4499
Mx.	5.9923	5.9432	5.9110	7.7880	8.1176	5.9809	7.8080	6.7064
Oec.	3.8120	4.8311	4.4651	7.7197	5.5814	5.2568	8.2979	7.3383
Ora.	8.0644	7.4468	8.0437	8.3470	8.9375	8.4242	7.8306	4.7959
Pac.	11.2952	10.5839	11.4966	9.6348	12.5113	11.5588	9.9353	9.0814
Pan.	11.0268	10.4671	11.1073	9.5913	12.5891	11.4215	9.2439	8.5110
Panth.	10.6425	9.9481	10.7343	9.2499	12.0389	11.1449	9.4149	8.2677
Phd.	3.6942	3.3193	3.0624	6.8433	5.3577	3.2317	8.2003	7.6100
Phdr.	3.5288	3.2659	3.3565	5.5153	5.0980	3.1160	6.3494	6.0831
Phlb.	5.6346	5.0597	5.4880	3.4226	7.1777	5.2100	5.6500	6.6622
Plt.	7.7062	6.7718	7.7281	3.7794	8.7227	7.3461	4.9312	6.3460
Prm.	8.0916	7.0616	7.2220	8.5633	8.1101	6.7035	9.7346	10.7020
Prt.	.	4.2967	2.6563	6.7364	5.5104	3.6476	7.9208	6.9017
Rep.	4.2967	.	3.9692	5.9931	5.6691	3.6121	7.5056	6.6322
Smp.	2.6563	3.9692	.	6.8780	5.0451	3.6658	7.7081	7.0093
Sph.	6.7364	5.9931	6.8780	.	8.2786	6.0381	5.2988	7.4135
Thg.	5.5104	5.6691	5.0451	8.2786	.	5.7399	9.4508	8.2839
Tht.	3.6476	3.6121	3.6658	6.0381	5.7399	.	7.5789	7.4312
Ti.	7.9208	7.5056	7.7081	5.2988	9.4508	7.5789	.	7.2114
	Prt.	Rep.	Smp.	Sph.	Thg.	Tht.	Ti.	Tma.

respect, because mutual distances between works, even when by different authors, are generally not of this magnitude, as a glance at the table will show. For most of the 52 works readings as high as 10 are found in four or five of the 51 possible distance measurements, whereas for the Isocratean speeches this value and values higher than 10 occur in 25 or 30 cases.

Such distances are at the higher end of the range for differences between works. In many cases, as we shall shortly see, authorial differences, where they are known to exist, are not matched by vast distances between works. Only in the case of Isocrates has this been found to be true, although since three of the other authors are represented by single works only (Aeschines, Lysias, Thucydides) and the Isaeus speeches have all been grouped under a single heading, as if they were part of a single composition, there may not be sufficient material to make a properly balanced comparison.

Nevertheless, it does seem that the overall impression of a distinctiveness of Isocratean style has been adequately confirmed.

The Republic

I take this work next because it is central in the Platonic corpus for both its philosophical content and its stylistic character, and probably also in its date of composition. It will therefore give a good indication of the sort of figure for distance which we might anticipate between other well-accepted Platonic works. The works closest to the *Republic* are shown in Table 8.17.[13] I have taken the list down to include the two works of Xenophon so as to show the inherent difficulty of using the distance measurement *per se* as a guide to

TABLE 8.17   *Mahanalobis distances: works closest to the* Republic

Phdr.	3.27	Euthd.	4.25
Phd.	3.32	Prt.	4.30
Tht.	3.61	Mem.*	4.31
Alc. 1	3.85	Ly.	4.37
Chrm.	3.91	Cri.	4.62
Smp.	3.96	Men.	4.70
Grg.	3.99	Oec.*	4.83

* Xen.

[13] The figure for *Smp.* + *Rep.* is found to be non-significant, i.e. the two works are judged as being probably from the same population. For the most part I have found these probability figures difficult to reconcile with anything which is known about the works. For example it is inexplicable to me that *Hipparch.* should have a non-significant difference between itself and the Ant., the figure being 10.297, yet for *Phdr.*, where the distance between the two works is only 6.224, this is regarded as being significant at the 0.0001 level.

authorship. Many of the genuine works of Plato are more distant from the *Republic* than are the *Memorabilia* and *Oeconomicus*, most notably the later group of dialogues, the lowest score for a member of that group and the *Republic* being 5.06 for *Philebus*. The highest figure to be found between the *Republic* and any other known Platonic work is that of 8.29 for the *Republic* + *Critias* distance, a figure higher than that found between Isocrates and his nearest neighbour.

As pointed out earlier, we are not in a position to assume that any set of variables will automatically reveal to us differences in authorship. These 37 variables measure many features of the samples and these will often relate to concepts such as genre, subject, treatment, linguistic usage, vocabulary, of which we may in part be aware, but it does not follow that such measurements will highlight disparity of authorship, unless we can find some means of eliminating noise and variance related to effects which stem from sources other than individuality of style determined by authorship. Candisc and discriminant analysis with an authorship-group classification will achieve this to a certain extent, and we may also use Stepdisc to select a more appropriate set of variables than the full set, if authorship discrimination is the main objective.

This approach has been used in some of the examples of discriminant analysis which I have described, but here I am trying to ascertain, by using all of the information which is available, what sort of relationship exists between these works, without at the same time appearing to beg the question of authorship. Thus while it is perhaps disturbing to find such a wide range of values for the distance figure between the main works of the Platonic corpus, it would be far more worrying if none of those works were found to resemble each other to the same degree as do the works of the Isocratean corpus, say *Panathenaicus* and *Panegyricus*, which have a separation of 3.37. We are more interested in those works which reveal proximities to each other than in the cases where vast differences are found to intervene, although the latter must testify in part to the enormous range of Plato's style.

Since the *Republic* is the second longest Platonic dialogue the average distance score between it and other dialogues is likely to be slightly lower than those for dialogues of medium length, since, on the assumption that all samples are drawn from a vast homogeneous population, the greater the number of samples taken the lower will be the variance. For shorter dialogues even greater distances are to be anticipated. I shall therefore take a figure of 4.0 or perhaps 4.5 as typical for a genuine Platonic work to exhibit as a distance measurement between itself and at least one other genuine work of the corpus. In some cases it may be valid to allow a higher figure because of special circumstances such as genre separation or peculiarity of subject.

Another point worth noting is that the comparative distance measurements are not necessarily reciprocal in the sense that where dialogue A is

found to be closest to dialogue B it does not follow that B must be closest to A. In the example above, the *Republic* is found to be closest to *Phaedrus*, but *Phaedrus* is itself closer still to the *Symposium*. The ranking of dialogues in the order of their proximity to any one given dialogue will frequently change when viewed from the position of one of those which has been ranked.

Alcibiades *1*

The closest dialogues to this work are shown in Table 8.18. The figures in the final column give the proximity ranking of this dialogue from the reciprocal viewpoint. Thus it is found to be the second closest of all the dialogues to the *Gorgias*, the third closest to *Meno* and so on. It seems astonishing that, if this work is spurious, the author should have had such success in matching the Platonic style as to be closer in many instances to genuine works than they are to each other. While one can accept that *Alcibiades* 1 might show some proximity to the *Gorgias* if it were intentionally imitative of that work (which in any case it is not), yet it is strange that the *Gorgias* itself has only *Protagoras* as a nearer neighbour, while the *Republic*, *Symposium*, and other dialogues approximately contemporary are kept far off in the distance.

There is also the evidence of the discriminant analyses to consider, where it was shown that *Alcibiades* 1 has far more of the genuine Platonic samples attributed to it than other suspect works (Tables 8.5 and 8.8). Of the group of early dialogues of Plato this must surely be one of the prime candidates for inclusion in the genuine corpus.

Amatores

The high values of distance for this dialogue are due partly to its shortness, since it contains only two samples. Its nearest neighbours are shown in Table 8.19. These results are not very convincing in terms of a bid for Platonic authorship. The proximity of the *Memorabilia* and *Oeconomicus* is disturbing, even though the ranking is not repeated in the reciprocal figures. For it is quite clear that *Amatores* is no closer to Plato than it is to Xenophon, as was more or less indicated by the first example of discriminant analysis, when the

TABLE 8.18   *Mahanalobis distances: works closest to* Alcibiades *1*

		Reciprocal position			Reciprocal position
Grg.	3.51	2	*Smp.*	3.88	6
Men.	3.67	3	*Tht.*	4.13	7
Phdr.	3.82	6	*Phd.*	4.19	7
Rep.	3.85	4	*La.*	4.27	3

TABLE 8.19   *Mahanalobis distances: works closest to* Amatores

		Reciprocal position			Reciprocal position
Mem.	5.23	13	*Oec.*	5.79	19
Rep.	5.56	20	*Phdr.*	5.97	20
Grg.	5.58	20			

division into seven authors resulted in one out of the two samples for this dialogue being classified with Xenophon (Table 8.1). Unless more evidence of stylistic resemblance comes to light it would appear to be wise not to accept this dialogue. The only reservation I have is that dialogues on the periphery of Plato's output might perform strangely in all these tests, because they occupy the extremes of his stylistic range, and are anomalous for this reason. There is no obvious internal justification for rejection and many scholars in the past have accepted its authenticity (see Guthrie, *History*, v. 390–2), but the stylometric evidence is not in its favour and at best can only be summarized as an absence of outright condemnation.

Apology

This has higher scores than are found for many of the other genuine dialogues and these are probably caused by genre estrangement, for the *Apology* is not dialogue but monologue, although it does contain one short passage of interchange between Socrates and his accuser (24 D–27 D). Its closest dialogues are shown in Table 8.20.

It is difficult to account for the proximity of the *Theaetetus*, which is a dialogue in all probability of a much later date, but perhaps the long, uninterrupted speeches in this dialogue, as also in *Protagoras* and the *Symposium*, play some part in generating the affinity. Unfortunately, the work had progressed too far to enable me to use Xenophon's short piece of the same title for comparison.

TABLE 8.20   *Mahanalobis distances: works closest to the* Apology

		Reciprocal position			Reciprocal position
La.	4.47	5	*Prt.*	4.84	17
Men.	4.60	9	*Smp.*	4.90	17
Tht.	4.66	13			

As one of Plato's works which are central to his writing, since it differs somewhat in genre it is useful to have here a record of its performance to compare with other possibly slightly inflated figures for outliers, such as the *Epistles*, which will also suffer under genre estrangement. The figures in themselves are no sure guide to authorship, since we shall find that those between Plato and Xenophon are sometimes lower, but it is in their relative values that we are looking for clues to their origins, how they compare with each other and what dialogues they are found to resemble most closely.

Clitophon

The proximities for this dialogue are given in Table 8.21. In most cases of discriminant analysis this short work is most often classed with the *Laws*, except in those cases where it is included in the main group of Plato's works, when it is found that it is not rejected as a misclassified sample. If the document classification is used (i.e. if the separation is made into individual works) as a single sample it is allocated either to the *Laws* or to *Epistle* 7.

On the above figures it ought to be rejected, but its proximity to so many of the works of the later period, if we ignore the presence of the Aeschines speech in the list, does imply that it is a later work and that the genre distinctiveness gives it an unusually high score. I am not aware that commentators have generally ascribed to it a late dating, but the fact that so many of these later dialogues are fairly consistent in their affinities to each other leads one to suspect that interlopers are rare. I am reluctant to atethize it, because of its proximity to the *Laws* and because the style of the later period is not an easy one to imitate and one would expect rather more roughness and irregularity in its relationship to other Platonic works, over and above the obvious hiccups caused by genre dissimilarity.

It is more monologue than dialogue and as such it is not altogether surprising to find it showing leanings towards the orators. However, I shall discuss its pecularities of style more fully in the following chapter, when assessing its chronological position, or rather, the position it would take if it were a genuine dialogue.

TABLE 8.21   *Mahanalobis distances: works closest to* Clitophon

Lg.	6.20	Tma.	7.05
Ep. 7	6.28	Rep.	7.25
Phlb.	6.91	Sph.	7.30
Phdr.	7.01	Epin.	7.35

TABLE 8.22   *Mahanalobis distances: works closest to* Cratylus

Prt.	4.66	Smp.	5.17
Hp.Ma.	4.67	Men.	5.20
Phd.	4.71	Phd.	5.23
Tht.	4.72	Grg.	5.28

Cratylus

This is a puzzling dialogue about the origins of language. Its aim is not fully understood. Modern opinion tends to favour grouping it with the *Theaetetus*, *Parmenides*, and the late-middle period of dialogues. Its slightly inflated scores I take to be the effect of individuality of subject, but I would not myself have detected any resemblance to the *Hippias Major*. The scores for its distance from the closest works are given in Table 8.22. All the works mentioned, with the exception of the *Hippias Major*, are dialogues whose authenticity is not in question and the proximity of *Cratylus* to the central core of Platonic writing seems to be confirmed by these figures.

Critias

The main point of interest here lies in the proximity of *Critias* to the *Timaeus*, the dialogue which it claims to follow (see Table 8.23). The fact that the two are found to be neighbours does confirm our reliance on these measurements as indicators of stylistic relationship. The inclusion of *Critias* in this group, close to all the late dialogues, also implies that it does not belong stylistically to the intermediate period of Plato's writing, but to the later group, with which it is usually placed (Chapters 7 and 9 *passim*).

Crito

One would perhaps expect this dialogue to be closest to *Phaedo*, but the distance measurement to either *Meno* or *Phaedo* differs only marginally (see

TABLE 8.23   *Mahanalobis distances: works closest to* Critias

Ti.	4.09	Epin.	5.75
Plt.	5.06	Ep. 7	5.97
Lg.	5.27	Phlb.	6.08
Sph.	5.72	Phdr.	7.19

TABLE 8.24   *Mahanalobis distances: works closest to* Crito

Men.	4.40	*Rep.*	4.62
Phd.	4.41	*Prt.*	4.68
Grg.	4.51	*Phdr.*	4.82
Alc. 1	4.60	*Smp.*	4.83
Tht.	4.60		

Table 8.24). One cannot attach great importance to small differences, for the simple reason that we are here dealing with mean values, and in the hypothetical case of Plato's perhaps choosing to alter one of the dialogues by editing, extending, or shortening it, for example, we would have been left with an adjusted mean value which would have resulted in different distance scores. What we are left with, of course, is the final work, but there must have been many influences and causes which shaped its final form, among which mere chance would have had its place, so that it would be unwise to insist too much on the niceties of placing of the proximate dialogues, when we know that if Plato had decided to expound further a particular point in any one of these the relative positions would probably have changed anyway.

The comparative seriousness of tone of the *Crito*, lacking the playfulness of some of the early- and middle-period dialogues, possibly enhances its distance measurements slightly above the norm. Its place in the main stream of Platonic writing is not in doubt, and this is confirmed by its proximity to so many of the central dialogues of the canon.

The Epinomis, Epistle 7, the Laws

I have taken these three works together for the sake of emphasizing their mutual proximity (see Table 8.25). Both *Epistle* 7 and the *Epinomis* select the *Laws* as being the one work out of the entire Platonic corpus to which they have greatest affinity. This is the most remarkable result to have emerged from the investigation so far, for these two works which have been variously dismissed as totally lacking in Platonic style and abysmally deficient in both form and content appear to be closer stylometrically to the *Laws* than to any other dialogues, either of the same period or earlier. It is not only that this result is achieved from the point of view of either the *Epinomis* or *Epistle* 7, for they cannot choose but be closest to one of the 51 other works available, and the possibility must exist that, if they were imitative, they might succeed in bringing themselves into reasonable proximity to the work which they copied or from which they were derived. But this argument does not apply in reverse, and the *Laws* ought to have far stronger affinities with other works of the same period, rather than with these two, if indeed they are not genuine.

TABLE 8.25   *Mahanalobis distances: works closest to* Epinomis, Epistle 7, Laws

		Reciprocal position			Reciprocal position
Epin.					
Lg.	3.60	2	*Ti.*	4.75	2
Sph.	4.31	4	*Plt.*	5.31	6
Ep. 7	4.50	3	*Phdr.*	5.69	20
Phlb.	4.52	4	*Criti.*	5.75	5
Ep. 7					
Lg.	3.35	1	*Ti.*	4.91	4
Phdr.	4.31	10	*Sph.*	5.05	5
Epin.	4.50	3	*Mem.*	5.26	14
Phlb.	4.82	6	*Rep.*	5.39	18
Lg.					
Ep. 7	3.35	1	*Plt.*	4.33	2
Epin.	3.61	1	*Tim.*	4.81	3
Phlb.	3.86	2	*Phdr.*	4.91	15
Sph.	4.10	3	*Criti.*	5.27	3

In addition, *Epistle* 7 does not set out to be imitative of the *Laws*, for it is an entirely independent work, differing in approach and subject from the latter, and one would anticipate that at the very least genre difference would separate it effectively from anything contemporary, even if the other works were by the same author. Yet, despite this barrier, the distance measurement between it and the *Laws* at 3.35 is distinctly at the lower end of the scale, with only a handful of the 1300 possible combinations achieving lower scores.[14] This must be acknowledged as being fairly decisive evidence in favour of authenticity, for, while one might have expected some leaning towards works of the period to which internal evidence suggests it belongs, the strength of the relationship is quite unusual in comparison with that exhibited by other genuine dialogues, showing that there is a kinship between these works which is closer than that between the *Sophist* and *Politicus* or between the *Timaeus* and *Critias*, both of which pairs are claimed to have been written as sequels to one another.

[14] Nine to be precise. These are: *Archd.* + *Pac.*: 2.9303; *Archd.* + *Panth.*: 3.0319; *Pac.* + *Pan.*: 3.1792; *Phd.* + *Rep.*: 3.3193; *Phd.* + *Smp.*: 3.0624; *Phd.* + *Tht.*: 3.2317; *Phdr.* + *Rep.*: 3.2659; *Phdr.* + *Tht.*: 3.1160; *Prt.* + *Smp.*: 2.6563. This puts the *Ep.* 7 + *Lg.* pair in the top 1% of close scores. The figure of 1300 is arrived at by squaring the total number of works, dividing by two, since the matrix is symmetric, then subtracting the original number, since a work has no distance from itself. Thus total number of scores $= \frac{1}{2}n^2 - n = n\left(\frac{1}{2}n - 1\right) = 52 \times 25 = 1300$.

The *Epinomis* must also, I think, be accepted as genuine, in view of the evidence of its proximity to the later dialogues, especially to the *Laws*, to which it is the professed sequel. These findings reinforce the impressions given by the earlier examples of discriminant analysis that *Epistle* 7, the *Epinomis*, and the *Laws* have much in common. That they were all by the same author appears to be incontrovertible, for, if the above evidence is to be turned aside as inadmissible, it is difficult to imagine what other evidence of similarity of style could be adduced which would be more convincing.

However, although it is now possible to set aside the arguments of those who pronounce these works to be disparate in style, one should perhaps look more closely at the contents, especially of the *Epinomis*, before reaching too hasty a decision. *Epistle* 7 is nowadays accepted by the majority of scholars and its thought and content has been widely discussed. It hardly seems necessary to rehearse once more the arguments *pro* and *contra*. The importance of having its authenticity confirmed is that it will now be possible to rely on the long excursus on the nature of reality (342 A–344 C) with confidence as a guide to Plato's later thought. Also we shall no longer need to distrust the biographical details which fill so much of this letter and which provide us with a much fuller account of Plato the man than it is possible to obtain from the dialogues.

The *Epinomis* is more obscure and is usually dismissed as being of little merit. For example, 'when we consider the *Epinomis* in detail, we very soon become aware of contact with an inferior mind, which feebly strays and stumbles among the last physical and metaphysical speculations of Plato'.[15] For Guthrie 'It is hard to believe that all this was written by the author of the *Timaeus*.'[16]

The premiss round which the work is written is that, although the *Laws* has provided the legal foundation for the ideal city, it has not been decided what knowledge should be imparted to the citizens to make them wise. Or, in a more general sense, what is that branch of knowledge which, when human beings partake of it, renders them possessors of wisdom? The answer is hedged round with circumlocutions, partly because it is a theme which has run through so many of the dialogues, if not all, imbuing them with a particular colour and earnestness. So many forms of knowledge seem to hold the key to wisdom, but on closer examination they fail the test of dialectic. Here at last we have the answer provided for all who wish to receive it. But it is partly obscured because Plato feared how easily it might be ridiculed. For he has already given warnings enough of the inadequacy of the written word (*Phdr.* 275 C–E, *Ep.* 7, 341 E).

However, the train of thought appears to be that it is through numeracy that the human race is distinguished from animals, and hence has the ability

---

[15] W. R. M. Lamb, Loeb translation, p. 424.   [16] Guthrie, *History*, v. 386.

to study astronomy, which teaches us about order and regularity in the universe, divine qualities which are attributable to the impelling power of the soul. By this means we are enabled to observe the heavenly bodies, which are the supreme manifest deities, and they must be regarded with due reverence. Habits of reverence inculcated by such studies should then be extended to embrace the rites and duties demanded by custom and culture towards the traditional gods. Astronomy is the key to an understanding of reality and to the achievement of wisdom, although evidently to Plato the term had a much wider meaning than that which we ascribe to it today, and embraced topics which we would more properly describe as belonging to religion. Astronomy, therefore, is the subject the citizens must make their special study, approaching it through an understanding of all the branches of mathematics and of the way in which the physical world is constructed on numerical principles. Only a few will be capable of attainment of all this knowledge and the accompanying spiritual perfection, and it is these few who by decree will be appointed to the highest office. The members of the nocturnal council are to be encouraged to pursue the attainment of this wisdom.

Now I do not claim that this is a major dialogue, or vital to an understanding or Plato, but it does answer a question left unanswered in the *Laws*, that of discovering the most appropriate subjects for the wardens to learn (968 D), and it is handled in a manner not unworthy of the writer of the *Laws*. It would be more appropriate to regard it, I suspect, as a precursor of the *Timaeus*, rather than as a substantial dialogue in its own right, and the points of detail on which it differs from the *Timaeus* are probably to be attributed to a change of perspective, or the passage of a few years, rather than to innate and damaging contradictions.[17] (See also the following chapter on chronology).

There we must leave this interesting little dialogue to its fate—*I, pedes quo te rapiunt et aurae*—and also *Epistle* 7, but they will both be considered again in the next chapter.

Epistles 2, 3, 8, and 13

These epistles cannot be used in a discriminant analysis, except as stray samples which are lodged with the group to which the discriminant function judges them to be nearest, but they themselves cannot form a group, as they each consist of only one sample. Nevertheless, the results are worth looking at, and in Table 8.26 I give the classification summary for these works taken from six different examples of discriminant analysis.

For *Epistles* 3 and 8 these results are consistent, suggesting a date contemporary with *Epistle* 7, which internal evidence in any case confirms.

---

[17] e.g. the presence of the fifth element, ether, in *Epin.* 918 C. See also Guthrie, *History*, v. 386, and references.

TABLE 8.26 *Classification summary for* Epistles *2, 3, 8, and 13 based on six examples of discriminant analysis*

Discrim[a]	Ep. 2	Ep. 3	Ep. 8	Ep. 13
All 37 variables, by document	*Chrm.*	*Ep.* 7	*Lg.*	*Alc.* 1
All 37 variables, by 14 author	*Thg.*	*Ep.* 7	*Ep.* 7	*Alc.* 1
20 variables A14, by document[b]	*Cri.*	*Lg.*	*Ep.* 7	*Phdr.*
20 variables A14, by document[b]	*Tht.*	*Criti.*	*Ep.* 7	*Phdr.*
10 PXX variables,[c] by 14 author	Pl. 2	*Ep.* 7	*Ep.* 7	*Ep.* 7
BLETS only, by document	*Thg.*	*Ti.*	*Lg.*	*Alc.* 2

[a] Variables used for each discriminant analysis and the classification scheme adopted.
[b] These examples differ in that for the latter all of Isoc. plus *Rep.* and *Lg.* were excluded.
[c] These are the first 10 variables selected by a Stepdisc analysis on a subset of works using only the established Platonic dialogues together with *Mem.* and *Oec.*

Epistle 3 appears to be closer in some cases to the *Timaeus* or *Critias*, but its proximity to later works is not in doubt. We could speculate that its ultimate position relative to the other epistles would depend on the position of the *Timaeus* (or *Critias*) relative to the *Laws*, a subject which will be treated in detail in the next chapter. These two epistles, 3 and 8, are probably genuine, since I have taken the authenticity of *Epistle* 7 as more or less proven. For the other pair, 2 and 13, the situation is more problematic.

If conjectural dating based on content turns out to be correct, then one would expect them to be close to the *Theaetetus*, or possibly to whatever is early in the later group. This would suit a classification with Plato 2 or with *Phaedrus*, as does occur in some cases, or even with *Epistle* 7, since genre similarity must play some part in the affinity measurement. For *Epistle* 2, therefore, only two of the six classifications are satisfactory, and for *Epistle* 13, three out of six.

These two epistles would present problems under any scheme of distance measurements, for they do not purport to be dialogues, which the majority of the works are, but private letters, whereas 3, 7, and 8 are much more open and appear to have been written for publication. *Epistle* 13 especially, with its domestic details of failure to store the figs and presents for Dionysius and his wife, is unique in the Platonic corpus and it is perhaps not surprising that its performance is erratic.

Looking at the distance measurements, however, the picture of affinities shown in Table 8.27 emerges. From this table it appears that *Epistles* 2 and 13 are too close to the early dialogues, although *Epistle* 13 evidently has some kinship with the *Phaedrus* and *Theaetetus*. Although it is true that we do not know the dates of *Charmides*, *Protagoras*, and *Euthydemus* (or of *Alcibiades* 1 or *Theages* if genuine), yet they are generally thought to be considerably

TABLE 8.27 *Mahanalobis distances: works closest to* Epistles *2, 3, 8, and 13*

	Ep. 2	*Ep.* 3	*Ep.* 8	*Ep.* 13
Alc. 1				5.32
Chrm.	5.69			
Ep. 7		6.53	5.37	
Epin.			6.38	
Euthd.	6.37			
Laws		6.81	5.36	
Phdr.		7.31	7.24	5.76
Phlb.		7.31		
Prt.	6.26			6.51
Smp.	6.41			6.18
Thg.	6.28			
Tht.				6.17
Ti.		7.54	6.81	

earlier than the *Theaetetus*. The argument from genre no doubt has some validity, but when there is no genre resemblance from competing contexts, one would expect the overall pattern of attraction to have some sort of chronological bias. The other disturbing feature is that *Epistles* 2 and 13 have a mutual score of 6.997, which puts them rather far apart in comparison with the dialogues listed above.

However, these high scores may spring from the fact that they all consist of only one sample each, and this tends to result in increased distance measurements. On the positive side it is clear that the affinities are towards Platonic works, or to those which in all probability are Platonic (*Epistle* 7 and *Alcibiades* 1).

On balance it seems best to accept 3 and 8 as genuine, despite their high scores, since they are lodged fairly and squarely with the later dialogues and close to *Epistle* 7 and the *Laws*, a position which internal evidence would demand. For 2 and 13 I propose to reserve judgement until the results of the chronological investigation are known.

Against Eratosthenes

This is the only work of Lysias included, the most well-known of his speeches (13) in which he attacks one of the Athenians who was instrumental in destroying his patrimony and securing the death of his brother. There are no obvious connections with any other works included here, so whatever turns out to be closest is likely to be oratorical in character. The list of proximities is given in Table 8.28.

TABLE 8.28   *Mahanalobis distances: works
closest to* Against Eratosthenes

Ep. 7	5.53	*His.*	5.95
Mx.	5.75	*Alc.* 2	5.96
Tma.	5.85	*HG*	6.11

Epistle 7 at the head of this list is rather surprising, but perhaps reflects the quasi-historical nature of both these works. They have something of the flavour of an *apologia pro vita sua* and are frequently packed with narrative of past events, so that there is more of a kinship than one might expect at first sight. One tends to forget when studying these distance measurements that each work has to be close to something and record one work out of the 51 possible as being its closest neighbour. In the case of Lysias, who is only represented by this work, the proximity is bound to be across the barrier of authorship, no matter what other work is placed close to this speech. Looking at it from the point of view of Epistle 7 we find that *Against Eratosthenes* occurs only tenth on the list, well away from the proximities among the later works of Plato (Table 8.16).

In general one could say that the main character of this speech is historical, for of the six works listed above five contain a high proportion of historical narration. *Alcibiades* 2 seems to be out of place, although it does contain two sections of pseudo-historical writing (141 c–143 a, 148 c–150 b), which could cause this anomalous result (yet they hardly amount to more than one-third of the text).

It is difficult to predict, in the case of authors represented by single works, where the affinities will lie, other than that they will inevitably entail a relationship across the authorship divide, which is perhaps stating the obvious. Thus in Table 8.28 there are at least four authors found to lie close to Lysias, possibly six if we should decide that the authors of the *Menexenus* and *Alcibiades* 2 are not Plato. The scores are such as one might expect between authors, but there is no rule which can be applied to determine what constitutes a gap sufficient to separate authors, and what is merely a local difference between works, for in many instances the separation is considerably lower than these figures for Lysias, and in others considerably higher.

Thus there is found to be a strong mutual link between Aeschines and Isaeus (*Against Timarchus* and *Orations*), but the relationship with the *Menexenus* is more remote for Aeschines than for Isaeus. Not one of these orators appears to be close to Isocrates in comparative terms, for other works by Plato and Xenophon intervene. But in reverse the Isocrates speeches are closest to other examples of oratory, firstly to other speeches by

Isocrates as has already been shown, and then to those of other authors, but the distance measurement is high even though their genre is in theory close.

Gorgias, Meno, Phaedo, Phaedrus, Protagoras, Symposium, Theaetetus

It is not necessary to deal with these works individually, since they are all from the well-established main period of Plato's writing, neither too early to present problems of classification, nor too late to be in a class on their own (see Table 8.29). *Phaedrus* is the one dialogue in a more ambiguous position than the others, as it has links with both early and late works, but it has such low scores on so many distance measurements that it seemed appropriate to include it. All have scores of less than 3.6 with at least one other dialogue. It is also noticeable how compact a group this is, with only *Alcibiades* 1 and the *Republic* appearing in the lists as the two dialogues from outside the group.

Charmides, Euthydemus, Euthyphro, Laches, Lysis

I have grouped these five dialogues together because they show some slight signs of erratic behaviour although they belong to the central core of Platonic composition (see Table 8.30). Their scores are somewhat higher than those of the previous group, starting at about 4.0 for the nearest relative, and higher for *Euthyphro* and *Lysis*. These higher scores may be linked to the fact that the dialogues here are shorter than those just dealt with and are therefore more liable to sampling fluctuation.

TABLE 8.29   *Mahanalobis distances: works closest
to* Gorgias, Meno, Phaedo, Phaedrus, Protagoras,
Symposium, *and* Theaetetus

Grg.			*Prt.*	
Prt.	3.39		*Smp.*	2.65
Alc. 1	3.52		*Grg.*	3.39
Men.	3.66		*Phdr.*	3.52
Men.			*Smp.*	
Tht.	3.59		*Prt.*	2.66
Grg.	3.66		*Phd.*	3.06
Alc. 1	3.67		*Phdr.*	3.36
Phd.			*Tht.*	
Smp.	3.06		*Phdr.*	3.11
Tht.	3.23		*Phd.*	3.23
Rep.	3.32		*Men.*	3.59
Phdr.				
Tht.	3.12			
Rep.	3.27			
Smp.	3.36			

TABLE 8.30 *Mahanalobis distances: works closest to* Charmides, Euthydemus, Euthyphro, Laches, *and* Lysis

Chrm.		La.	
Rep.	3.91	Men.	4.01
Phd.	4.15	Grg.	4.25
Phdr.	4.35	Alc. 1	4.27
Euthd.		Ly.	
Prt.	4.15	Rep.	4.37
Rep.	4.25	Grg.	4.79
Phdr.	4.26	Alc. 1	5.03
Euthphr.			
Alc. 1	4.59		
Thg.	5.01		
Grg.	5.43		

Charmides and *Euthydemus*, with their proximity to *Phaedrus* and the *Republic*, are probably the latest of the group, with *Euthyphro* perhaps the earliest, especially if we can also show that *Theages* and *Alcibiades* 1 are early works. *Lysis* is rather puzzling, since one would not expect it to have too much in common with the *Republic*, but the presence of the *Gorgias* and *Alcibiades* 1 in the list, as with *Euthyphro* and *Laches*, seems to be true to form.

Hippias Major

Proximities for this dialogue, the authenticity of which has been frequently questioned in the past, are given in Table 8.31. This certainly does not look like a spurious work, since it is so close to many of the major dialogues and does not include alien authors among its nearest neighbours. It is not as close as we might wish in order to make assurance double sure, and it does not approach the figure of 2.66 between *Protagoras* and the *Symposium*, but it still represents typical scores such as are achieved by the well-established dialogues, and there is no obvious evidence of abnormality, such as an allegiance to an outside work or to a later Platonic dialogue. On balance the

TABLE 8.31 *Mahanalobis distances: works closest to* Hippias Major

La.	4.35	Phd.	4.81
Grg.	4.50	Smp.	4.83
Prt.	4.60	Alc. 1	4.90
Cra.	4.67	Tht.	4.99

TABLE 8.32 *Mahanalobis distances: works closest to* Hipparchus

Grg.	5.20	Men.	6.48
Prt.	5.86	Alc. 1	6.61
Ion	6.17	Tht.	6.63
Euthd.	6.19	Oec.	6.63
Phdr.	6.22	Smp.	6.68
Mem.	6.23		

evidence for genuineness is fairly convincing, if taken in conjunction with the previously recorded cross-classifications of the various discriminant analyses.

Hipparchus

The figures in Table 8.32 are rather too high to inspire confidence. We may compare them with those for *Ion*, which are dealt with subsequently, where there is at least one low score for the *Ion* + *Gorgias* distance. The problem, however, with any work which perhaps lies at the extremes of an author's range is that some abnormalities are almost inevitable and its style will probably impinge also on other neighbouring authors, as we find here in the case of the *Memorabilia*. After some early apprentice works an author's style might alter considerably, so that one cannot easily predict the performance of such early pieces. The results from discriminant analysis were not particularly good either for this dialogue, so that we can at least conclude that it does not belong to any group of works showing typical Platonic features. The evidence for rejection, however, is not conclusive, for we are not in a position to say at what point the dividing line between genuine and spurious should be drawn.

Ion

As an early work, which in all probability it is, there is no reason to suppose that *Ion* should have especially strong links with many other dialogues and the figures given in Table 8.33 seem adequate to justify its inclusion in the corpus of genuine works.

TABLE 8.33 *Mahanalobis distances: works closest to* Ion

Grg.	4.55	Alc. 1	5.37
Prt.	5.02	Phdr.	5.41
Smp.	5.24	Hp.Mi.	5.49

The Laws, Philebus, Politicus, Sophist, Timaeus, Critias

I deal with all these works together as they form the group of late dialogues whose unity of style is noticeable even at the most simple level of measurement. Between members of this group and nearly all those dialogues of an earlier date of composition there is usually a larger gap than that shown for any of the distances within the group. *Phaedrus*, however, acts as a bridge between early and late and is often found as a near neighbour within this subset. The ranking lists taken from the matrix of Mahanalobis distances are shown in Table 8.34. The regularity of these lists, with the same dialogues nearly always included, gives us just cause for thinking that we are dealing with a closely related group of works which have a common stylistic bond. There is the additional consistency that the *Timaeus* and *Critias* are always found to be close, as also are the two pairs *Sophist–Politicus* and *Laws–Epinomis*. *Epistle* 7 is also found in all six cases to be a member of the group.

One might have hoped that the stability which is displayed by this group of late works would be found also in the works earlier in date, but this, unfortunately, is not so, and it is much more difficult to detect consistency of style in the earlier period. This is partly because more works are involved anyway, and partly, I suspect, because the innate variability is much greater and it is more difficult to pick out the underlying patterns. However, the existence of this late group of works will be of great assistance later in enabling us to discover the chronology of composition of the works.

Hippias Minor

Guthrie (*History*, iv. 191) takes this to be a genuine dialogue, since it is mentioned by Aristotle (*Metaph.* 1025ᵃ6) and he shows that its content may be reconciled with orthodox Platonic doctrines, if we accept it as an attempt to test by a form of *reductio ad absurdum* the paradox of Socrates' belief that virtue is knowledge. Here we are concerned with the stylometric evidence, although it is relevant to point out that a citation by Aristotle cannot be taken as a guarantee of authenticity pure and simple, but only as proof that Aristotle knew the work, whoever the author might have been. In this case he does not mention the author but only refers to 'the argument in the *Hippias* that the same man is both true and false', and this certainly does not preclude Platonic authorship but equally certainly does not guarantee it.

The stylometric evidence is not much in favour of authenticity (see Table 8.35), especially with the damaging proximity of the *Oeconomicus* indicating that the work is no closer to Plato than it is to Xenophon. The results from discriminant analysis do not show it to be especially Platonic in style either, so that we are left with rather negative evidence for those who wish to prove Platonic authorship. In its favour we could argue that, although it has been shown that it stands outside the ambit of the typical Platonic works, the earliness of date could have caused it to have the affinities attributed to it.

TABLE 8.34 *Mahanalobis distances: works closest to* Laws, Philebus, Politicus, Sophist, Timaeus, *and* Critias

Lg.			
Ep. 7	3.34	*Plt.*	4.33
Epin.	3.61	*Ti.*	4.81
Phlb.	3.86	*Phdr.*	4.91
Sph.	4.10	*Criti.*	5.27
Phlb.			
Sph.	3.42	*Phdr.*	4.64
Lg.	3.86	*Ep.* 7	4.82
Plt.	4.41	*Rep.*	5.06
Epin.	4.51	*Tht.*	5.21
Plt.			
Sph.	3.78	*Criti.*	5.06
Lg.	4.33	*Epin.*	5.31
Phlb.	4.41	*Ep.* 7	5.55
Ti.	4.93	*Era.*	6.20
Sph.			
Phlb.	3.42	*Ep.* 7	5.01
Plt.	3.78	*Ti.*	5.30
Lg.	4.10	*Phdr.*	5.51
Epin.	4.31	*Criti.*	5.72
Ti.			
Criti.	4.09	*Plt.*	4.93
Epin.	4.75	*Sph.*	5.30
Lg.	4.81	*Phlb.*	5.65
Ep. 7	4.91	*Phdr.*	6.35
Criti.			
Ti.	4.09	*Epin.*	5.75
Plt.	5.06	*Ep.* 7	5.97
Lg.	5.27	*Phlb.*	6.08
Sph.	5.72	*Phdr.*	7.19

TABLE 8.35 *Mahanalobis distances: works closest to* Hippias Minor

Oec.	4.46	*Phd.*	5.17
Smp.	4.74	*Phdr.*	5.19
Grg.	4.96	*Alc.* 1	5.20
Prt.	5.02	*Mem.*	5.31

That it is an unusual dialogue is not in dispute, but such qualities as it possesses do not in themselves prove that it could not have been written by Plato.

I confess to a certain reluctance to athetize this work, because the methods I am using do not appear to work well with dialogues which are at the extremes of a stylistic range, but only with those which can be shown to be typical. The fact that the work is fairly unusual has been demonstrated, but it is not so unusual as to stand out from all the others as being unquestionably false. I therefore include it in the chronological tests, in an attempt to establish the date of composition which should be ascribed to it if it were genuine.

The Memorabilia, Oeconomicus

The affinities of these two may cause some surprise (see Table 8.36), but we have to consider that the number of works by Xenophon which are included is limited, so that, apart from the immediate connection with either one or the other, neither the *Memorabilia* nor *Oeconomicus* has any choice but to show itself resembling in some way works that are adjacent in style, even though they are not by the same author. I exclude the *Hellenica* from this account because it seems to me that the stylometric judgement is correct in suggesting that the gap between it and these two is greater than that between either of them (the *Memorabilia* or *Oeconomicus*) and many of the works of Plato. This is caused by the genre difference, which undoubtedly in Greek results in a different use of language.

One would hope, however, that a Platonic work would not be at the top of the list in each case. For if this method fails to show an authorship difference by indicating that the *Memorabilia* and *Oeconomicus* are, from the point of view of each one separately, the two works which are closest to each other of the 52 included, then how can one rely on it to do so in other instances?

TABLE 8.36   *Mahanalobis distances: works closest to* Memorabilia *and* Oeconomicus

Mem.			Oec.		
		Reciprocal position			Reciprocal position
Phdr.	3.76	5	*Prt.*	3.81	6
Oec.	4.09	2	*Mem.*	4.09	2
Rep.	4.31	10	*Grg.*	4.41	8
Prt.	4.35	10	*Hp.Mi.*	4.46	1
Smp.	4.49	12	*Smp.*	4.47	9

The assumption that underlies this query is that any assortment of variables is necessarily going to reveal differences between authors above everything else, and it is clear that this assumption is false. They can do this if they are deliberately slanted to achieve that aim, or if the authorship differences are greater than any others which interpose between the various works. Generally speaking, a random selection of variables will give us a good average assessment of style, one which tells us how close individual works are to each other on the basis of whatever information is contained in these variables. The 37 variables used here may be regarded as having been selected at random, because the main motive in choosing them was ease of collection of the data and because they are dependent on the letter content of words which, since orthography and the alphabet as a means of recording sound seem to have been created as a congeries of historical accidents, must have an element of randomness in their application. So that it is not unreasonable to claim that these variables give us a good generalization of style.

The picture that emerges from this generalization is that authorship differences are not necessarily dominant, and there are many cases where overlap occurs and where two works by different authors are closer than two by the same author. This is true of the *Memorabilia* and *Oeconomicus*, which are closer to many of the intermediate Platonic dialogues than most of these latter are to the works of Plato's later period.

This problem may be overcome by using such techniques as discriminant analysis with an authorship classification, whereby the differences which the variables reveal between authors are enhanced to produce the maximum effect. Even at the level of simple Mahanalobis distances one can improve the authorship differentiation by using a subset of specially selected variables. For example, a typical result using the first 20 variables which Stepdisc selects from the 14-author classification described in the early part of this chapter shows that both the *Memorabilia* and *Oeconomicus* have each other as their nearest neighbours, confirming their common origin, but the overall gap between them and the next nearest Platonic dialogues is not much increased.

If we insist that style is something which may be measured, and stylometry does make this assertion, then the generalized measurements which are given by the Mahanalobis distances ought to be accepted as representative of an underlying reality, even though the results may be in some cases unpalatable. On the analogy that style covers a sort of spectrum of which it is possible to measure the wavelength of each part, then it would appear that, on the basis of stylometric measurement, some works of Xenophon are intermediate in wavelength between late Plato and the works of his middle period, while Isocrates lies at the extreme edge of the spectrum. Using enhancement techniques to discriminate between authors is equivalent to the traditional

method of selecting words for their distinctiveness ratio, the difference being that Discrim and Candisc are much more sophisticated in their approach and more far-reaching.

I believe, in any case, that these stylometric judgements do present a metric of style which is much truer to the realities of language than the traditional approach which supposes that because Plato was Plato and Xenophon Xenophon there could be no confusion between them. The later works of Plato do form a separate group which is linguistically distinct, and there is a gap between this group and the Platonic works which precede, a gap which is stylistically quite large and evident to most readers. We can make the mental adjustment to accommodate these two styles because we know that they are both the product of one man. Stylometry, however, tells us that some of Xenophon's work is intermediate between these two styles, and that it is closer to the former Plato (Plato 1 in the earlier examples of Discrim) than the two halves of Plato are to each other. In an absolute sense of style, or, more accurately, in the sense which is determined by the use of these 37 variables, it is surely right to accept this judgement. It does not imply a *rapprochement* between Plato's philosophy and the conceptual framework of Xenophon's writing, a man whose interest in philosophy probably did not extend beyond that of the average well-educated layman, but it does indicate that, at some stage in the separate orbits of these two writers, their use of language runs on a parallel course.

Hence the distance measurements show that the *Memorabilia* is close to quite a few of the Platonic dialogues, and closer to *Phaedrus* than it is to the *Oeconomicus*, while this latter work is closest of all to *Protagoras*. These results would cause us to shudder if the objective had been to highlight differences between authors and had selected variables especially for this purpose, or had used the various enhancement techniques available to us to discriminate between authors. But in fact the intention was to illustrate in a general way which works showed the greatest similarities, using the whole range of variables without any interference, in this way obtaining a much broader picture of the underlying linguistic problems.

As to the interpretation of the above figures, despite the proximity of the two works to various Platonic dialogues, there is only limited reciprocity, for the *Memorabilia* and *Oeconomicus* both occur quite low down on the respective ranking lists for *Phaedrus*, *Protagoras*, the *Republic*, *Symposium*, and *Gorgias*, as the final column shows, although from the point of view of authenticity tests it would certainly be easier if they had been even farther removed.[18] The most important lesson to be learnt from all this is that we cannot rely on a single result or just a few measurements to solve questions of authenticity, because the linguistic background is of great complexity, and an

---

[18] I mean in terms of a larger distance measurement, for that *Mem.* and *Oec.* must show *some* proximity to *some* of the dialogues is an inevitability of the construction of the distance matrix.

approach which imposes a simplistic solution disregards the many possible factors which might cause any one of a whole series of works to exhibit some degree of resemblance. What we need to know is how much and in what contexts these resemblances occur, and then make a judgement against the background of our knowledge of the behaviour of an author's work in the entire field which we have been able to study.

The Menexenus

This strange dialogue illustrates the difficulty of dealing with genre bias, for the *Menexenus* is an example of an *epitaphios* or funeral oration. Such speeches were delivered annually at Athens over the tomb of those who had fallen in battle in the preceding year. From the list of affinities in Table 8.37 it would appear that the style of this *epitaphios* is no more that of Plato than it is of Thucydides, or possibly Lysias, and, curiously enough, in the works of all these authors there is an extant example of the genre.

Discriminant analysis with a seven-author classification is marginally in favour of Plato as author, but it is certainly not possible to claim that it is typical of Plato, for everything about it seems to indicate its strangeness and difference from the norm. It is mentioned by Aristotle twice in the *Rhetoric* (1367^b8 and 1415^b30) and this is counted as decisive by most scholars in favour of authenticity, even though Plato is not mentioned as the author by name.[19] The verdict of stylometry is rather against it, but I am inclined to think that that is more a reflection of its peculiar nature than a definitive declaration against Platonic authorship. I do not know how stylometry as I have been using it could cope with a genre distortion as sharp and distinctive as this one. In the case of the epistles it did not seem to create the sort of problems which might have been expected, and the distance measurements were only slightly inflated, and for *Epistle* 7 not at all. We could allow a certain discount on its scores to cater for the genre distance, say 1.0 or 1.5, and this would bring it fairly close to *Phaedrus*, which is a dialogue boasting a similar theme, the fickleness of oratory. The historical element in the *Menexenus* is enough to bring it into contact with Thucydides, the *Hellenica*, and to a lesser extent Lysias, so that if these are removed as being too

TABLE 8.37   *Mahanalobis distances: works closest to* Menexenus

His.	5.27	Smp.	5.91
Phdr.	5.71	HG	5.92
Era.	5.75	Mem.	5.93

---

[19] Guthrie, *History*, iv. 312 f.

obviously a genre classification, we are left with *Phaedrus* and the *Symposium* as the two closest works, at distances of 4.7 and 4.9 respectively (subtracting 1.0 for genre).

I accept that this is an unsatisfactory proceeding but I cannot offer anything better. In line with the majority of scholars I am inclined to agree to its authenticity, because it is so patently a Platonic theme and one wonders who else could have dreamed up the idea of Aspasia producing a hotch-potch speech for Socrates, and for any other orator who might require it, and almost giving him a beating when he failed to remember it.

Nevertheless, I concede a weakness in the stylometric evidence. The problem is discussed further in the following chapter.

Minos

The figures given in Table 8.38 appear to be too high for this to be a genuine Platonic work, since there is nothing in the dialogue itself to indicate extreme abnormality, either by way of genre or subject matter. The position of *Cratylus* at the head of the list is inexplicable, unless they are both to be taken as very early dialogues, a proposition which would not find much support among Platonists. However, the high scores may be due in part to the small number of samples (2), and it is unwise in the circumstances to be too dogmatic about authenticity.

The Parmenides

I have left this dialogue till the last because it poses special problems. There is no doubt as to its authenticity, except perhaps for some nineteenth-century scholars,[20] but we need to ask the question 'Why is it so remote from all the other dialogues?' (see Table 8.39). None of the other dialogues achieves such high scores, even those that I have been prepared to athetize, with the single exception of *Minos*, but even that has one considerably lower score. In fact the *Parmenides* stands out on all the peripheral tests which I have conducted as the one dialogue which would have to be rejected if it were necessary to eliminate the one which was most atypical. Yet it is even more remote from

TABLE 8.38 *Mahanalobis distances: works closest to* Minos

Cra.	5.96	*Alc.* 1	7.63
Alc. 2	7.38	*Mem.*	7.71
Hp.Mi.	7.49	*Phdr.*	7.80

[20] Ibid. iv. 40.

TABLE 8.39 *Mahanalobis distances: works closest to* Parmenides

Tht.	6.70	*Phdr.*	7.45
Phd.	6.77	*Men.*	7.49
Rep.	7.06	*La.*	7.52
Smp.	7.22	*Chrm.*	7.53
Alc. 1	7.27		

other authors than it is from Plato, and there is no question of declaring it to be spurious. In addition, its proximity to the *Theaetetus* is fairly decisive, for that is the one work with which it is usually coupled on the basis of its philosophical content and of internal references.

Some attempt must be made to account for these high scores and to explain the oddity of this work. It takes no great insight to see that the source of the deviation lies in the Eleatic section of the dialogue, from 137 c onwards, in which Parmenides undertakes a dialectical enquiry into the consequences of the existence of 'the one' and 'the others' (το ἕν, τα ἄλλα). The first, shorter section is conventional dialogue. I have tried splitting the dialogue into two parts, a conventional and an Eleatic section, and including them both in analyses. The effect on the Mahanalobis distance is a considerable increase for the latter part, emphasizing still further its remoteness from all other Platonic works, and a decrease for the first section, showing its kinship to Plato's normal style.

In fact, had the Eleatic section of the *Parmenides* been detached from the Platonic corpus and been presented as a work by a different author, there would have been no difficulty in 'proving' that it could not have been written by Plato. How does one account for this abnormality of style, which makes the *Parmenides* stand out from everything else which Plato wrote? For few would accept that anyone other than Plato could have written it.

But perhaps I am making too much of an obvious difference caused by subject matter and the abstract nature of the discussion. The concepts which are used are entirely those which were the stock-in-trade of the Eleatics, Parmenides, Zeno, and Melissus. Thus we find the one and the many, the others, being and non-being, motion and rest, smallness and greatness, limited and unlimited, and all the various antinomies which arise from a consideration of these topics in relation to the Parmenidean concept of 'the one'.

The problem of assessing why it is that Plato came to be writing in a style so foreign to his own is compounded by the fact that most of the information that we have about the Eleatics derives from this dialogue—Simplicius

notwithstanding[21]—so that our enquiries tend to follow a circuitous route. Thus to conjecture that the *Parmenides* is a typical example of dialectic from the Eleatic school, and then to examine it as such, tends to confirm our opinion of how such discourses ran, without drawing on any evidence independent of Plato which might give us an insight into the relationship between Parmenides and Zeno and whether or not dialectic of this type was the staple diet of the Eleatic school.

Fortunately some fairly large portions of Parmenides' poem 'On Nature'[22] do survive, and the one thing that is certain is that its language does not resemble that of Plato's *Parmenides* in any way. One must look elsewhere for the source of the work, or at least its stylistic abnormality, for it is true that the ideas dealt with are fundamentally those of Parmenides and his poem. To Solmsen there is an obvious debt to Zeno[23] but he does not make it clear whether he considers it to be in the realm of style, or ideological, or both, or neither.

For Cherniss,[24] 'Parmenides and his poem are the butt at which the second part of the dialogue is aimed', and also, '(it) . . . is an attack on Eleaticism by the father of the school, a parody of the method used in Zeno's book, but not a parody of the form of that book'. Presumably this means that Plato is parodying Zeno's method of arriving at contradictory conclusions (by misuse of the verb 'to be'), but not his general approach of dialectical enquiry, the use of a probing sequence of questions to discover all the consequences of a given hypothesis. But whether he thinks that the dialogue is imitative of Zeno's language and style, for the sake of parody or otherwise, he does not say.

I am inclined to take the plunge at this point and state positively that this work is highly derivative, based probably on a Zenonian original, possibly even the work mentioned in the dialogue, but if not that precisely, then on some other production(s) of the Eleatic school. For the only surviving Greek which in any way resembles that of this dialogue are the fragments of Zeno and Melissus of Samos, also reputed to have been a pupil of Parmenides.[25] In addition, Zeno is claimed by Aristotle to have been the 'father of dialectic'[26] so it is possible that Plato's work is a tribute to Zeno as much as a parody of him, for it is through dialectic,

by comparing all these things with each other—names and definitions, sights and perceptions—testing them with kindly proofs and using question and answer without

---

[21] Friedrich Solmsen, 'The Tradition about Zeno of Elea Re-examined', *Phronesis*, 16 (1971), 116–41, reprinted in A. P. D. Mourelatos (ed.), *The Pre-Socratics*, (New York, 1974), p. 369. Simplicius evidently sees Zeno through the eyes of Plato.
[22] G. S. Kirk and J. E. Raven, *The Presocratic Philosophers* (Cambridge, 1962), p. 263.
[23] Solmsen, 'The Tradition about Zeno of Elea Re-examined'.
[24] H. F. Cherniss, 'Parmenides and the *Parmenides* of Plato', *AJP*, 53 (1932), 122 ff.
[25] Kirk and Raven, *The Presocratic Philosophers*, p. 298.
[26] Ibid. 287; D.L. viii. 57.

contentiousness, by such means, although with difficulty, the light of reason and comprehension will shine out concerning each of these things in the mind of him who has exerted himself to the limit of human endeavour. (*Ep.* 7, 344 b).

On this basis then I would explain the divergence of the *Parmenides*, that it is imitative of Eleatic material which, had it survived, would give us a far better chance of understanding the import of the dialogue, whether it is satirical or serious, and its points of reference to the Parmenidean concept of 'the one' and to Plato's own philosophical development. Its language is not Platonic, but closer to the surviving Zenonian fragments which deal with the contradictions inherent in the concept of 'the many', and those of Melissus dealing with reality and the infinite.[27] The verbal parallels between Zeno, Melissus, and the *Parmenides* are not exact, but they are closer than anything to be found elsewhere in Plato. Evidently he could, when he chose, write in this way, for the evidence is before our eyes, but the assumption that he alone should have been responsible for the development of this style and that it should spring naked and perfectly formed from his mind for this one occasion is astonishing.

I therefore offer the above interpretation as an interim measure, in an attempt to satisfy the stylometric evidence which seems to demand the presence of an author other than Plato.

Alcibiades 2

This dialogue has in the past been rejected emphatically for un-Platonic language,[28] so it is disappointing to find such scant evidence for this in the stylometric record (see Table 8.40). Apart from the presence of the *Memorabilia* early in the list there is nothing here to indicate that this is an unusual or spurious dialogue. It perhaps falls into the category of works such as the *Memorabilia* and *Oeconomicus*, which are intermediate in the stylometric sense between early-middle and late Plato, and therefore are difficult to

TABLE 8.40   *Mahanalobis distances: works closest to* Alcibiades 2

Men.	4.80	Tht.	5.29
Phdr.	4.93	Rep.	5.42
Ap.	5.05	Prt.	5.44
Mem.	5.07		

---

[27] Kirk and Raven, *The Presocratic Philosophers*, pp. 365–6, 381–5.
[28] Guthrie, *History*, v. 387.

separate from the genuinely Platonic. Some evidence for this is to be found in the affinities for *Phaedrus*, the *Theaetetus*, and the *Republic*, which are dialogues just preceding the late period, and the stylometric tests of misclassification in the various cases of discriminant analysis did not work well for this dialogue.

Nevertheless, it cannot be shown to be definitively non-Platonic and it is probably best in the circumstances to reserve judgement.

Conclusions

This has been an exceedingly difficult chapter. It is not simply that the problem lies in showing that one work differs from another, or that any one author at a certain period is dissimilar to many others. That for the most part can be demonstrated. But what we are required to prove is that all authors at all times differ from each other, and it may be doubted whether language possesses this capacity for such extensive and continuous differentiation. Can we show that the writings of all authors bear the impress of a unique and recognizable individuality, no matter how humble or exalted, idiosyncratic or pedestrian those writings might be, no matter how greatly he or she might have been swayed by the ordinary pressures of the use of speech which exert their influence on all who share a common language?

This problem becomes more difficult the more varied the language of the author with whom we are dealing. For as style changes to accommodate new forms and new flights of fancy it is almost inevitable that it should come into contact at some points with the styles of other authors.

Nevertheless, there have been some positive gains, for we have succeeded in showing how style may be measured and how the range of an author's output may be depicted by a series of mutual distance measurements between works. Against this background it was found that four works in particular were so similar to the main body of Platonic material, in some cases even closer to the central dialogues of the surrounding corpus than many of the other dialogues traditionally accepted as genuine, that it was impossible to deny them a place as true Platonic works. Thus *Epistle* 7 and *Epinomis* were found to be closer to the *Laws* than any other dialogues of the entire corpus. Similarly, *Hippias Major* and *Alcibiades* 1 are found to have so much in common with the *Republic* and other works of the early and middle period that it is scarcely rational to reject them. Besides these four we could also include *Epistles* 3 and 8 since their kinship with *Epistle* 7 and the late works was clearly demonstrated.

At the other end of the range the greatest difficulties lie when we are faced with dialogues that look as if they ought to be rejected. While it is reasonable to propose that works which may be shown to be absolutely typical should be

accepted, the converse does not apply, that atypical works are necessarily spurious and should be excluded from the canon. For our knowledge of the 'norm' is built on whatever we recognize as genuine and this process in itself represents a prejudice against dubious works. Merely to repeat the judgement by showing how suspect works differ from this constructed norm is in a sense tautological. For by definition a work which is atypical and recognized as such will lie outside the norm for that author. But that in itself does not constitute proof that it was written by another, unless it can also be shown that the author in question never departed from his standard established style. In the case of Plato, with works like the *Parmenides* and *Menexenus*, and to a lesser extent the *Apology* and *Cratylus*, this departure from the typical is obvious. And this should warn us to be cautious when considering the status of dubious dialogues.

On these grounds I feel reluctant to pronounce a final verdict on *Alcibiades* 2, *Amatores*, *Hipparchus*, *Hippias Minor*, *Minos*, *Theages*, and *Epistles* 2 and 13. They could be forgeries, since they seem to be outside the main ambit of Plato's style. On the other hand they could be early works, written in his formative years, before his style was properly developed, and therefore liable to abnormality. For *Epistles* 2 and 13, peculiarity of subject and genre may be sufficient to account for their anomalous behaviour.

The *Menexenus* and *Clitophon* I take to be genuine, although the stylometric evidence is not much in their favour, especially for the former. For *Clitophon*, the proximity to all the later works is not something which I believe any casual artificer could have achieved. On a purely subjective level both dialogues seem to me to be too outrageous to have been written by anyone other than Plato. For they recall such devices as the farcical claim made by Socrates in the *Protagoras* that the best philosophers in Greece originated from Sparta (342^{a-d}), or the buoyant, ebullient humour of the *Euthydemus*. Additional justification for placing *Clitophon* among the later works will be found in the relevant chronological section of the following chapter.

Also in the next chapter all of the above-mentioned dialogues (with the exception of *Alcibiades* 2) will be discussed further, for the question of date of composition cannot be entirely separated from that of authenticity. The shorter, dubious dialogues are fairly lightweight and the probability is that they are earlier works, and a failure to group with other early dialogues could tip the balance of opinion against them. On the other hand if they are placed in positions which seem to be consistent with the philosophical ideas expressed in them, so that they harmonize with the surrounding dialogues, then it would be relevant to take this as additional evidence of authenticity.

9

The Chronology of the Dialogues: Stylometric Evidence

DETERMINING the order in which a series of works has been written is a problem which differs from that of authorship identification. In the latter case it is necessary to discover some quantities which remain unchanged throughout an author's career, while in the former change is the essence of the enquiry. In a typical univariate case one might hypothesize that a certain feature in an author's style is used with varying frequency and that the increase (or decrease) of usage is linearly related to the passage of time. Thus Shakespeare's use of 'and' could be deemed to be a chronologically related variable and by calculating average values of frequency of usage for this word for each play the sequence of composition could be graphically represented, provided of course the dates of at least two plays were known (preferably one early, one late) so that the slope of the line could be correctly plotted.

Such a scheme could also be applied to Plato, but since I have no figures for the use of individual words I will illustrate the method using readings for some of the BLET variables averaged out over the sample values for each dialogue. Table 9.1 gives the values for BLET4 and BLET5 (words ending in ι and ν respectively) for 25 of Plato's dialogues.

These two variables are chosen deliberately because, as will become apparent later in the chapter, they are closely linked to Plato's changing style (see especially Table 9.3 and also Chapter 10). If we were to be restricted to the use of only one or two variables in the analysis of chronological variation, then these two, BLET4 and 5, would be at the top of the list of preferred candidates, namely all those capable of demonstrating to a greater or lesser extent the differences between early and late works of Plato.

In Table 9.1 the 25 Platonic dialogues are listed alphabetically. For the purpose of argument we are to suppose that the readings of these variables are linearly related to the date of composition of each work. Thus for BLET4 *Crito* is the earliest work (21.97) and *Critias* the last (12.47). For BLET5 *Hippias Major* (21.99) is first, the *Sophist* (31.40) being the final work of the list. If we assume that Plato commenced writing in 399 and continued until his death in 347, then the position of each of the works of Table 9.1, as indicated by either variable, could be shown graphically, as in Figure 9.1.

In this graph the lower line represents the BLET4 figures, the upper those for

TABLE 9.1   *Mean values of* BLET4 *and* BLET5 *for 25 dialogues*

Work	BLET4	BLET5	Work	BLET4	BLET5
Ap.	21.75	23.67	*Men.*	21.90	24.37
Chrm.	19.06	24.57	*Phd.*	20.58	23.25
Cra.	21.27	23.16	*Phdr.*	17.94	25.72
Cri.	21.97	22.97	*Phlb.*	16.86	28.72
Criti.	12.47	30.32	*Plt.*	15.54	31.04
Epin.	15.25	30.90	*Prm.*	18.52	26.03
Euthd.	18.97	24.24	*Prt.*	21.43	24.79
Euthphr.	20.90	22.74	*Rep.*	18.95	25.10
Grg.	20.83	24.26	*Smp*	20.61	24.48
Hp.Ma.	21.69	21.99	*Sph.*	15.68	31.40
La.	21.51	23.59	*Tht.*	19.80	25.06
Lg.	15.31	30.04	*Ti.*	13.08	29.77
Ly.	19.02	22.75			

BLET5. In each case the extreme values, quoted above, have been taken as termini for the period of writing, spanning from 399 to 347, and the lines have been drawn between the appropriate coordinates for these two points. Other dialogues have then been placed on the line at the point where the BLET value indicated that they should lie. Thus for BLET4 *Crito* and *Critias* are at the extremes and determine the position and slope of the line, the other dialogues being fitted wherever the BLET4 value shows that they should lie. Their dates of composition may then be read off on the x-axis. For BLET5 *Hippias Major* and the *Sophist* are the terminal dialogues. Not all of the dialogues given in the list have been shown on the graph, since the space scarcely allows it, especially as the earlier works are mostly clustered together and cannot be separated with clarity.

Now it is obvious that such a system of dating leaves much to be desired. However, one should bear in mind that here its weakness becomes apparent through reasons of comparison, because different sources of information have been shown to give conflicting results, whereas if only one set of figures had been used and these had been claimed to have validity above all others for the purposes of dating the dialogues, the discrepancies might well have remained hidden, with no evidence of disagreement being brought to light. But the method is intrinsically flawed in other ways, not so much due to the potential contradictions between the results achieved by different variables, serious though these contradictions are, but because an unwarranted assumption is made about the relationship between time or date of composition (the dependent variable), and the BLET measurement. Although we are free to hypothesize that this relationship is linear, giving a straight-line
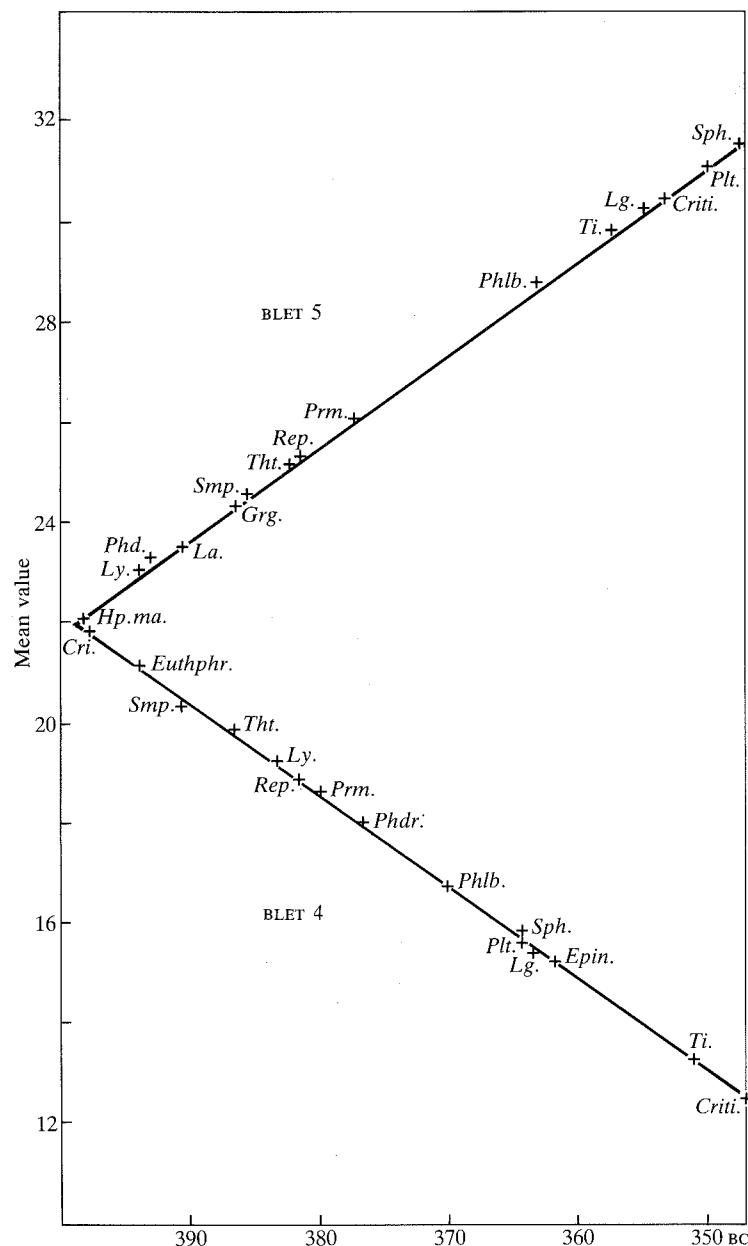
FIG. 9.1   Graph derived from Table 9.1

graph, there is no means of proving this. Indeed the evidence seems to suggest that the phenomena are far more complex and that simple mathematical assumptions will not be adequate to describe them.

After all, in terms of what we might reasonably expect from an author's style and its development, any one feature might at some time become a favourite trick of style, consciously or unconsciously pursued and then gradually abandoned or neglected in favour of other forms. Such a feature, if measurable, would be best described by a quadratic function in its relationship to time, a function which, as time progresses, reaches a maximum or minimum point, falling or rising on either side of that point depending on whether the feature was a positive or negative aspect of style, one assiduously cultivated or one 'more honoured in the breach than the observance'. A simple example is the use of a particular word or construction, one for which, for whatever reasons, the author develops a liking, and which he then gradually abandons in favour of other stylistic idiosyncrasies. Even the commonly occurring words could possibly be shown to behave in this way if the author's style were looked at over a wide range of works spanning a long period.

But to use the example which we have to hand, the BLET4 variable, the graph of Figure 9.1 clearly shows the *Timaeus* and *Critias* as having the lowest scores on this variable. We could take these scores as being representative of the minimum point of a quadratic function. With such an assumption these two works would no longer be the final compositions of Plato, as Figure 9.1 seems to indicate, but they would occur possibly somewhere in the centre of the later group of works, with dialogues of earlier and later date on either side. Figure 9.2 illustrates this possibility, using just a few dialogues, since in any case we are not in a position either to confirm or deny the accuracy of such an interpretation. The main purpose of this discussion is to demonstrate that the problems associated with chronological analysis are in fact considerably more complex than is generally admitted.

As regards the high scores of the BLET4 variable which occur with *Crito* and *Meno*, if the assumption of a quadratic function is to be maintained, then these cannot be regarded as a maximum or additional turning-point on the curve (in which case the function would be cubic or polynomial), but as a continuation of the curve away from the single turning-point already noted as the *Critias* coordinate.[1] In theory this curve continues to infinity on either side, although in practice the line ceases at the limiting dates of Plato's writing career, 399 and 347 BC. A similar curve could be constructed for BLET5.

[1] Precise coordinates of the minimum value can only be calculated mathematically if enough data regarding dates of composition is already available (specifically, exact dates for at least three dialogues). The minimum point may be on either side of *Crit.*, not necessarily *Crit.* itself. For the curve represents a theoretical approach to a practical problem and has an infinite number of points corresponding to an infinite number of possible dialogues.
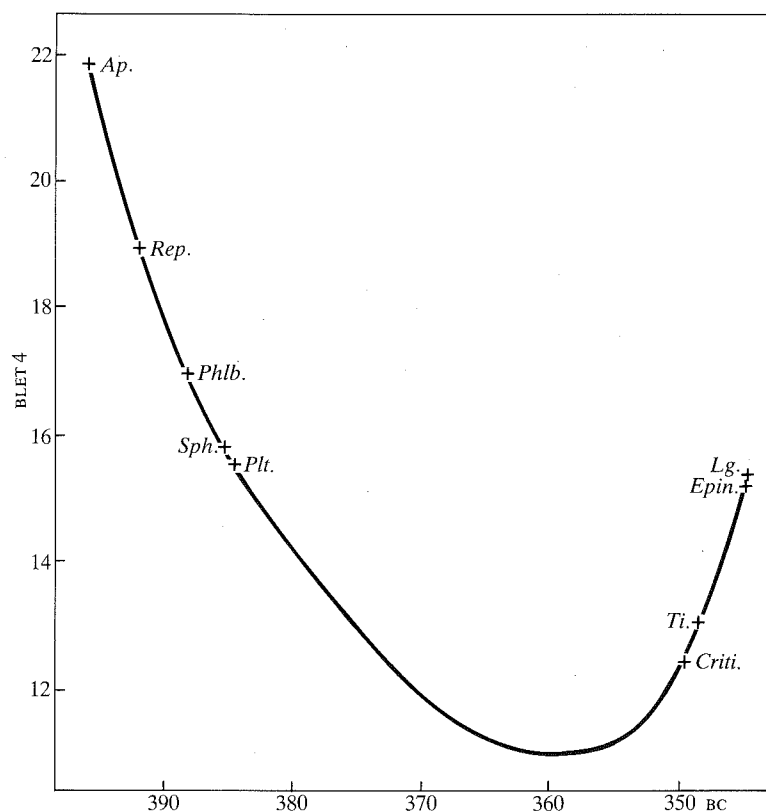
FIG. 9.2   BLET4 as a quadratic function, derived from Table 9.1

Yet even allowing for the possibility that a quadratic function might be the appropriate explanation of these results, a proposition which cannot be verified, we are still a very long way from exhausting all the possible interpretations which might be applied to the data. There are no reasons a priori justifying a denial that the features observed might be best modelled as cyclical or exponential functions, or based on cubic or higher powers than the simple linear or quadratic functions discussed above. Nor should one forget that what appears to be a pattern in a set of figures may, on closer investigation, turn out to be no more than a random scatter of values which when emphasized in certain ways give the impression of order or regularity.

The fact is that we do not know, for any given author, or for authors in general, how linguistic usage changes with the passage of time. Plato provides a good example of an author whose work has been studied for signs

of chronological variations. Lutoslawski[2] gathered and collated the work done by many scholars who had measured and counted certain distinctive features of style in the Platonic canon, and he showed how a summation of these characteristics for each individual dialogue could be used as a measurement of the distance of each from the *Laws*, a work which he, along with most scholars, believed to be Plato's last work. A later survey was made by Brandwood,[3] who also used computer counts of clausulae rhythms to date the dialogues. Underlying all this work is the assumption of linearity, the belief, usually unstated, that a shift in value of the linguistic feature measured is a direct representation of an exactly proportionate shift in the date of composition.

However, as shown above, this hypothesis is difficult to sustain and there is little evidence to support it, more especially in the case of ancient authors where the biographical information is scanty, usually anecdotal, and often contradictory.

Whether or not multivariate analysis provides a better approach to these problems remains to be seen, but it should be stated at the outset that any attempt to reduce a multiple observation (i.e. a series of readings which are descriptive of one sample) to a single number must be fraught with uncertainties. Each of the 493 samples of Plato is characterized by readings on 37 variables and the task is to reduce the 37 to a single reading by combining them in some way, at the same time ensuring that the technique of combination reflects and enhances the chronological influences which altered Plato's style over the span of his writing career. This is not a simple problem and there is no one correct solution to it. The approach which I offer is tentative, although I believe it to be sound. But the reader must form his or her own conclusions as to whether or not the methods used are adequate to support the conclusions which are presented.

As stated above, we are dealing with a set of 37 readings for each one of the 493 samples of the Platonic corpus, each sample being 1000 words long. It is perhaps most helpful to consider the information that is contained within these variables as forming a sort of profile. We are interested in observing how stable this profile is and how it may be observed to alter over a period of time.

We have seen how by using the techniques of discriminant analysis with an authorship classification[4] the stability of the profile may be demonstrated. But now the interest is in using the same variables with a different objective, to see how the profile alters gradually with time, and to use this alteration as the basis for our chronology.

Before proceeding to an outline of the methods used, I think it is important

[2] *The Origin and Growth of Plato's Logic.*
[3] L. Brandwood, *A Word Index to Plato* (Manchester, 1976).
[4] Ch. 8, *passim.*

first to look briefly at the variables in order to suggest how the use of them might possibly overcome some of the difficulties associated with univariate analysis. In the first place each individual variable is itself a composite variable. Thus ALET1, which measures the percentage of words containing alpha, will be affected by many different features of vocabulary, grammar, syntax, and subject. Therefore, the sort of fluctuation which is associated with single features, be it the use of a particular word or grammatical construct, will be swamped by the great mass of shifting detail which may not necessarily be in phase with the movement observed in other features, features which are nevertheless included in the orthographic measurement. Consequently, one should obtain from this ALET variable, and from all the others, a sort of wave motion or underlying movement which tells us something significant about the style of that author.

It is clear that these variables will not have been influenced directly by deliberate, conscious choice on the part of the author, so that we may justly make the claim that only unconscious features of style have been measured.[5] In addition the use of several of these variables together, in some cases the full set of 37, apart from the fact that each one separately taps many sources of individuality, will make it most unlikely that two closely similar samples, or sets of samples, will be so by accident.

Equally, because of the complexity of the profile, especially where the variables have been chosen out of the main group so as to emphasize chronological change, we may be fairly sure that proximity of works indicates a proximity in date. Evolution, it is found, is not regressive, so that even where superficially similar species have been seen to exist owing to the return or recreation of identical environments, the underlying structure and basic design is fundamentally different. In the same way an author would not return to a style after a lapse of many years, because the bones and sinews of thought would have altered in the intervening period, and any superficial similarities would only serve to mask a basic structure which had changed considerably.

The problem, therefore, is one of species identification, but in this case the species are to be defined by a chronological factor. Two works may be judged to be of the same species if their profiles as measured by certain chronologically significant variables are so close as to make separation difficult. Identity of species will then imply that the date of composition of the works concerned is very close. What we have to do is to attempt to separate the works from each other as effectively as possible, using these chronological variables, and see whether or not in the process a sequence of proximity is discovered.

---

[5] No doubt conscious elements of style are also measured by the orthographic variables. It is doubtful if anyone could successfully alter their style consciously in this respect except in rather trivial ways.

Discriminant analysis is usually found to be the most efficient way of classifying species, but for our purposes it tends to be too successful, as in this instance we are more interested in the failure of the discriminant process, when two or more dialogues are found to be very difficult to separate from each other. It is not very helpful to obtain a discriminant function which successfully classifies all samples from *Cratylus*, the *Parmenides*, and the *Theaetetus*, when what we really wish to know is how close the dialogues are to one another. It is possible to obtain a printout of the Mahanalobis distance between the dialogues, this being a measure of proximity which takes into account the fact that many of the variables correlate, so that they are in a sense only duplicating information which has already been supplied.[6] However, with large numbers of classes the figures become difficult to interpret, since 40 dialogues will yield 780 separate Mahanalobis distances between these dialogues, and in each case the distance is a measurement on a multi-dimensional scale based on the variables which have been chosen for the analysis. This can lead to apparent contradictions, especially for dialogues placed near the middle of the series.

Thus it may be found that many dialogues seem to have the *Republic* indicated as their nearest neighbour, whereas the *Republic* itself can only be nearest to one of these, or possibly to some other dialogue which does not appear in the original list of those which have the *Republic* as their closest neighbour. In a multi-dimensional world such apparent contradictions can arise and it becomes difficult to construct a sequence of composition from such information because of the conflicting claims for position made by competing dialogues.

I have found it to be more helpful to use canonical discriminant analysis (Candisc in the SAS system), an approach which still uses the Mahanalobis distances, but has the advantage that it produces an actual value of the first canonical variable for each sample,[7] as well as the mean value of this variable for each group (or dialogue).

The theory on which Candisc is based is that there is some unique combination of the original variables which will cause this combined variable to have maximum correlation with group membership. In other words the samples are to be classified according to their score on the canonical variate, and this variate is calculated in such a way as to obtain the highest proportion of correct classifications. In this way it is related to discriminant analysis, because it attempts to separate all the dialogues by making each one achieve a unique score on the canonical variate.

---

[6] Ch. 8, *passim*.

[7] *SAS User's Guide*, pp. 369 ff. This is true also for the subsequent canonical variables, of which there are as many as the original variables, or groups, whichever is the lesser. However, our interest is mainly in CAN1, since the subsequent canonical variates usually highlight unpredictable and unexpected differences between the dialogues, owing to the requirement that they should have zero correlation with all preceding canonical variables.

This variate takes the form

$$\text{CAN}1 = ax_1 + bx_1 + ..., nx_n$$

where x_1, x_2, etc. are the scores on the original variables, the ALETS, BLETS, or CLETS, and a, b, c,..., n, are the unique coefficients calculated for the first canonical variate, CAN1. CAN1 by definition summarizes most effectively all the available information in the data, at least in so far as classification in the correct group is concerned, so that the most extreme dialogues will yield the highest or lowest scores on this variable, while those in the middle section will have appropriate scores and will be found to resemble each other fairly closely. In this way we can construct a sequence from the lowest to the highest, or vice versa, and if the original variables are carefully selected so as to give information which is chiefly orientated towards temporal variation of style, we may then claim that the sequence is approximately that of the date of composition of each dialogue.[8]

The procedure to be adopted is, therefore, twofold and consists of, first, selecting a subset of variables which gives information relevant to changes of style occurring over a period of time and, secondly, using these variables in a Candisc analysis of all those dialogues which we wish to date.

Selecting a subset of supposedly temporal variables is not too difficult, as we may use a stepwise discriminant analysis on two groups of dialogues which are known to belong to different periods. This may appear to be begging the question, since we are presupposing that certain dialogues either post-date or pre-date others and this may constrain the subsequent analysis to mimic the paradigms. There is a certain amount of truth in the criticism, but in selecting dialogues for the groups we are only using such information as is universally acknowledged, such as that the *Laws* post-dates the *Republic*, or that *Protagoras* and the *Gorgias* are earlier than the *Sophist* and *Politicus*. If we did not have such knowledge the safest option would be to use all 37 variables in the Candisc analysis, since the chances are that, with such a large number of variables, most of the fortuitous and irrelevant variations would be ironed out.

In any case we shall be using four different group pairs to find the optimum subset of variables to use in the Candisc analysis. The stepwise discriminant analysis will list in descending order of importance those variables which contribute most to the power of the discriminant function, or, to use layman's language, those variables which are most useful for separating the two groups.[9] Final results from the analyses based on these four different

[8] I have simplified in this summary of Candisc. Strictly speaking it is the multiple correlation, R, between the variables and the group-membership variable which is maximized. The theory allows for the calculation of a unique set of coefficients for the group-membership variable also. Technically minded readers may consult the *SAS User's Guide* and the publications referred to therein.

[9] Discriminant analysis works by maximizing the F-ratio for the groups. In stepwise

TABLE 9.2   *Composition of sets A–D*

Set	Early dialogues	Late dialogues
A	*Rep.*	*Lg.*
B	*Grg., Phd., Prt., Smp.*	*Phlb., Plt., Sph., Ti.*
C	*Ap., Chrm., Cri., Euthd., Euthphr., Grg., La., Ly., Men., Phd., Prt., Rep., Smp.*	*Lg., Phlb., Plt., Sph.*
D	*Grg., Prt.*	*Plt., Sph.*

groups may thus be used as a cross-check against each other to ensure that no wildly aberrant chronologies are given credence.

Table 9.3 shows the results obtained from each of the four Stepdisc analyses, based on the four group pairs A, B, C, and D. The composition of these is shown in Table 9.2. As the table shows, set C is the most wide-ranging and set A the most restricted, the latter containing only two dialogues, albeit the two that are by far the longest. The danger with using a more limited choice of dialogues for deriving the most suitable set of variables to be used in chronological studies is that the differences in style may arise from circumstances unconnected with temporal variation. Thus, if two dialogues only are used, the fact that one is reported speech and the other direct speech may create differences of style which are reflected in the Stepdisc selection of variables. Or the differences may be due to nothing other than subject matter, and this also would affect the final selection of variables. The computer cannot know that we are only interested in variation produced by temporal change in style, but must make the best use of the material which is presented to it.

However, it is most unlikely that the differences between the *Republic* and *Laws* is due only to trivial and fortuitous causes, for anyone who reads the two dialogues cannot but be aware of an essential difference in the use of language, so much so that one could be forgiven for thinking that they were by two different authors. Nevertheless, I shall attach more importance to the results obtained from sets B and C, because these are more likely to dilute the effects of any coincidental or fortuitous discrepancy between individual dialogues. It is more important to find variables which are closely related to the changing features which Plato's style reveals as he grows older, than to discover that others are efficient discriminators of a small group of works.

We should not expect that the variable sets will necessarily resemble each other. This is due to the nature of discriminant analysis and the fact that

discriminant analysis each variable is selected in turn if it contributes most to the value of the F-ratio, or is discarded if it fails to contribute above a certain threshold level to the increase of this value.

TABLE 9.3 *Stepwise discriminant analysis on early and late groups using 37 variables*

TABLE 9.3 *continued*

Step	Variable Entered	Variable Removed	No. in	Partial R²	ASCC*	Probability >ASCC
Set A: 170 samples						
1	ALET11		1	0.6201	0.62007235	0.0001
2	ALET10		2	0.2216	0.70425985	0.0001
3	CLET7		3	0.1943	0.76171690	0.0001
4	ALET15		4	0.1543	0.79849265	0.0001
5	BLET4		5	0.1116	0.82098957	0.0001
6	CLET9		6	0.0724	0.83394743	0.0001
7	ALET19		7	0.0549	0.84305709	0.0001
8	BLET2		8	0.0519	0.85120138	0.0001
9	ALET14		9	0.0513	0.85883731	0.0001
10	ALET4		10	0.0764	0.86962468	0.0001
11	ALET9		11	0.0418	0.87507251	0.0001
12	CLET5		12	0.0419	0.88030829	0.0001
13	ALET13		13	0.0344	0.88442627	0.0001
14	ALET6		14	0.0382	0.88884587	0.0001
15	ALET12		15	0.0337	0.89259240	0.0001
16	ALET1		16	0.0269	0.89547766	0.0001
17	BLET1		17	0.0264	0.89823311	0.0001
18	CLET4		18	0.0165	0.89990782	0.0001
19		ALET12	17	0.0124	0.89864819	0.0001
20	ALET5		18	0.0324	0.90193182	0.0001
21	ALET17		19	0.0302	0.90489778	0.0001
22	ALET7		20	0.0251	0.90728641	0.0001
23		BLET4	19	0.0040	0.90691336	0.0001
Set B: 140 samples						
1	BLET5		1	0.6657	0.66571297	0.0001
2	BLET4		2	0.2326	0.74345906	0.0001
3	BLET9		3	0.1304	0.77692465	0.0001
4	ALET10		4	0.0741	0.79344663	0.0001
5	CLET7		5	0.0836	0.81071419	0.0001
6	BLET7		6	0.1305	0.83541591	0.0001
7	ALET4		7	0.0613	0.84550430	0.0001
8		BLET9	6	0.0065	0.84448748	0.0001
9	ALET19		7	0.0440	0.85132855	0.0001
10	ALET7		8	0.0556	0.85959990	0.0001
11		BLET4	7	0.0000	0.85959546	0.0001
12	CLET6		8	0.0504	0.86667394	0.0001
13	BLET3		9	0.0835	0.87781296	0.0001
14	ALET9		10	0.0257	0.88095858	0.0001
15	BLET6		11	0.0211	0.88346456	0.0001
16	ALET15		12	0.0211	0.88592630	0.0001
17		BLET7	11	0.0161	0.88406005	0.0001
18	ALET1		12	0.0509	0.88996264	0.0001
19	CLET9		13	0.0330	0.89359526	0.0001
20	ALET3		14	0.0290	0.89668254	0.0001
21	ALET14		15	0.0240	0.89916479	0.0001
22		ALET9	14	0.0138	0.89775705	0.0001
23	CLET2		15	0.0200	0.89980053	0.0001
24	CLET4		16	0.0193	0.90173383	0.0001
25	BLET9		17	0.0205	0.90374764	0.0001
26		CLET2	16	0.0118	0.90260016	0.0001
27	ALET6		17	0.0210	0.90464130	0.0001
Set C: 346 samples						
1	BLET5		1	0.6377	0.63765349	0.0001
2	BLET4		2	0.1522	0.69282029	0.0
3	BLET8		3	0.1114	0.72703333	0.0
4	BLET3		4	0.1454	0.76672392	0.0
5	ALET9		5	0.0783	0.78499575	0.0
6	BLET9		6	0.0723	0.80054796	0.0
7	ALET10		7	0.0672	0.81395251	0.0
8	CLET9		8	0.0380	0.82102519	0.0
9	CLET7		9	0.0398	0.82815171	0.0
10	ALET4		10	0.0297	0.83326236	0.0
11	ALET14		11	0.0363	0.83932189	0.0
12	ALET11		12	0.0197	0.84248123	0.0
13		BLET5	11	0.0007	0.84237224	0.0
14	BLET2		12	0.0229	0.84598492	0.0
15	ALET12		13	0.0231	0.84954744	0.0
16	ALET1		14	0.0293	0.85395482	0.0
17	ALET19		15	0.0270	0.85790477	0.0
18	ALET15		16	0.0328	0.86256216	0.0
19		BLET9	15	0.0046	0.86192126	0.0
20	ALET6		16	0.0208	0.86479082	0.0
21	ALET5		17	0.0138	0.86666075	0.0
22	ALET13		18	0.0185	0.86912536	0.0
23	ALET17		19	0.0133	0.87086014	0.0
24		BLET8	18	0.0050	0.87021589	0.0
25	ALET7		19	0.0160	0.87229886	0.0
26	CLET5		20	0.0178	0.87457376	0.0
27	CLET4		21	0.0146	0.87640368	0.0
28	ALET3		22	0.0068	0.87723975	0.0
Set D: 70 samples						
1	BLET5		1	0.7414	0.74139568	0.0001
2	ALET10		2	0.2441	0.80451045	0.0001
3	BLET9		3	0.1454	0.83293171	0.0001
4	ALET7		4	0.1514	0.85823387	0.0001

TABLE 9.3 *continued*

Step	Variable		No. in	Partial R^2	ASCC*	Probability $>$ ASCC
	Entered	Removed				
5	BLET7		5	0.0650	0.86744895	0.0001
6	CLET7		6	0.1346	0.88529423	0.0001
7	ALET19		7	0.0552	0.89163053	0.0001
8		BLET9	6	0.0330	0.88793702	0.0001
9	CLET6		7	0.0741	0.89623882	0.0001
10	ALET8		8	0.0802	0.90455972	0.0001
11	ALET5		9	0.0481	0.90915237	0.0001
12		CLET7	8	0.0280	0.90654001	0.0001
13	CLET5		9	0.0588	0.91203100	0.0001

Note: ASCC = average squared canonical correlation; two class levels (early and late); significance levels: to enter = 0.15, to stay = 0.15.

many of the variables are correlated. Thus BLET5 is chosen first in sets B, C, and D but does not appear at all in set A, which has ALET11 as first choice. These two variables are probably quite closely correlated, since one records words ending in v and the other words containing v. Consequently, the absence of either one or other of these from the list is explained by the presence of the other, as a high correlation implies that the two correlated variables contain approximately the same information and using them both in a discriminant analysis would be superfluous.

It is interesting to see also that set C, which is based on the most catholic data set consisting of nearly all the undisputed dialogues, selected five BLETS among the first six variables of greatest importance for discrimination. Yet it is the BLETS which we suppose to be principally related to grammatical measurements, as they measure the percentage of words ending in specified letters, so that they would be directly influenced by the inflexional use of language. In the broadest sense we may perhaps interpret this to mean that the difference between early and late Plato is not so much one of vocabulary, in which ALETs would chiefly figure, but in peculiarities in the use of language, which expresses itself as a greater or lesser reliance on inflexional values.

It does not follow that all differences between authors and works could be similarly categorized, for even in the case of the *Republic* and *Laws*, only two BLETS appear among the first 10 variables chosen. In any case it is best not to rely too much on these rule-of-thumb interpretations, because the complexity of multivariate analysis often renders such interpretations misleading or false. One would need to look at a full chart of the correlations of all the variables to be able to make sensible judgements about the significance of a

subset chosen by stepwise discriminant analysis, and even then one would have to express oneself with caution.

I will therefore pass on to the more immediate problem of deciding how many of these chosen variables to use in the subsequent Candisc analysis. Here there seems to be no hard-and-fast rule which can be applied, for if we decide to use only the first two or three, it seems to be approaching perilously close to the univariate situation, with all its attendant possibility of error. Whereas increasing the number used must also bring in a lot of extraneous sources of variance which have nothing to do with changes caused by temporal features of style. An upper limit of 10 is in practice reasonable, for it is found that Candisc with this number achieves a high value of canonical R^2, the correlation of CAN1 with the group membership variable being in the region of 0.8.[10]

To present a broad picture of the possibilities that this type of analysis offers, and perhaps its limitations, I have used the Candisc analysis on each of the four subsets of variables, but employing initially the first 10, then the first nine, and so on down to the first three. This gives eight separate analyses for each of the sets A, B, C, and D. In addition I have done a Candisc on the Platonic dialogues using all 37 variables, the ALETs separately, the BLETs, and the CLETs.

I will now attempt to summarize the procedures outlined above and, I hope, make them clearer, as I am sure that the explanations have been rather difficult to follow. Firstly, a subset of variables is chosen based on paradigms which represent early and late Plato. This is then used in the entire corpus of Platonic dialogues in a canonical discriminant analysis (Candisc), a procedure which enhances the effectiveness of the variables as discriminators between the dialogues. Thus we are making use of these time-related variables to see how successfully the dialogues may be distinguished from each other, although we are more interested in the failure of the process than in its implementation. The figure which is of greatest importance to us is the mean score of each dialogue on the first canonical variate (CAN1). It is calculated by averaging the CAN1 scores over all the samples for the named dialogues.

The importance of the CAN1 variable is that it has the highest possible correlation with group membership, so that it would effectively be the one variable which could classify most successfully all the samples with the correct parent dialogue. For that is the basic objective in its calculation, and the mean score represents a unique score attributable to each dialogue, a

[10] Canonical R^2 gives the proportion of variance in the group-membership variable which is accounted for by the specified canonical variable. Broadly speaking this means that it is an indicator of how successfully the CAN variable sorts the samples into the correct group. The greater the number of variables used, the higher the value of R will be, since each variable adds some new information, however small. Values higher than 0.8 could perhaps imply that the information being used is related to factors other than purely temporal ones.

score which distinguishes each dialogue from those adjacent to it in the sequence. In the ideal situation where all the dialogues belonged genuinely to distinct categories, that is where each one formed its own group and did not overlap with its neighbours, these mean values would be well defined and separated by large intervals from neighbouring values.

But in practice individual samples from various dialogues have scores falling on both sides of the mean values and there is a considerable amount of overlap, especially as, in the way which by implication we are defining style for this analysis, there is not a great difference between the style of any two dialogues which are close in date. Thus the mean CAN1 value for the *Republic* might be 1.5, that for the *Symposium* being 1.0, but it is fairly certain that many individual scores of samples from the *Republic* will be below 1.0 and many scores for the *Symposium* above 1.5. In general, the greater the similarity between the samples of any two dialogues, the greater will be the amount of overlap of the CAN1 scores, and the mean values of CAN1 for the two dialogues will gradually merge, while conversely, if the differences between the two dialogues are great, the mean CAN1 scores will be far apart.

In the sequence of scores on this variable we thus have a sequence of affinities, but since these affinities are based on original variables which were selected to show changes dependent on temporal stylistic features, the sequence is also one of date of composition. Those dialogues with scores near the middle of the range will be the ones written somewhere near the middle period, while those at the extremes must be either the first or last in the series. Where two scores are found to be fairly close, we shall not be in a position to assert that one or other of them must have priority, because the sample values have a considerable spread on either side of the mean value, as has already been mentioned. The standard deviation of CAN1 is often a substantial fraction of the mean value.

Taking an example to illustrate this point, we find that a Candisc using BLETs only gives mean CAN1 values ranging from 2.71 for the *Hippias Major* to − 4.46 for *Critias*. *Crito* and *Lysis* have values of 2.36 and 2.34 respectively and are placed as numbers 7 and 8 in the series. Yet looking at individual scores on these dialogues one finds that sample values range between 3.62 and 0.60 for the four samples of *Crito*, with a slightly smaller spread for the six samples of *Lysis*. This information is displayed in Table 9.4. From the figures in this table it may be seen that *Lysis* is much more of a homogeneous dialogue than *Crito*, although that of course is only in the sense implied by the use of the BLET variables to define style.

These scores indicate that individual samples could potentially be placed far from the parent dialogue, since both the 3.62 for *Crito* and the 3.10 for *Lysis* are higher than the highest mean score attained by any dialogue (*Hippias Major*: 2.71); while the 0.60 for the low-scoring sample of *Crito* would place it later than the *Republic*, at position 27 approximately.

TABLE 9.4 *Canonical discriminant analysis on* Crito *and* Lysis *using* BLETs *only*

Work	No. of samples	Mean	Standard deviation	High	Low
Cri.	4	2.36	1.29	3.62	0.60
Ly.	6	2.34	0.60	3.10	1.44

Evidently these results and sequences need to be interpreted cautiously, since it is apparent that the CAN1 mean score values conceal a large amount of fluctuation in the individual sample scores. This is only to be expected, as we are dealing with works by one author, many of which we anticipate to have been written with no great gap of time to separate them, so that differences between individual samples of any one work are likely to be as great as those between whole dialogues, for they are effectively from the same population.

It may be argued that this apparent heterogeneity of the dialogues invalidates completely the approach we are taking, since, if interpreted literally, it could be used to prove that sample 1 of the *Crito* was written before any other of Plato's dialogues, while sample 4 was delayed for some 20 or 30 years, until after the *Republic* was completed. It could be further argued that previous studies of Plato have not shown such uncertainty in the placing of the dialogues, or uncovered these apparent contradictions, and that they must therefore be more reliable.

The simplest answer to this criticism is that Plato has never been studied (stylometrically) with this level of detail before, but it is fairly certain that, if previous studies had used samples of this size (1000 words), they would have revealed just as much, if not more, variation than this research has shown. It does not invalidate a set of results to show that, if sample sizes are reduced, a different set of average values is obtained, for on that basis we could continually reduce sample sizes until, arriving at the level of single-word samples, we could conclude that those that were not identical were from a different population. Arguments of this type involve a *reductio ad absurdum* and miss the essential point that the implications of sample size and its effect on sample means is well covered by statistical theory. What is affected is the reliability of the results and the degree of confidence which may be attached to them. We are entitled to use average values because they are the best guide to the population mean values which it is possible for us to obtain, and no one supposes (or should suppose) that individual sample values will equal, or even be very close to, the mean value. That would imply a world with almost no variation, and the whole science of statistics would be invalidated.

However, as stated above, where it is found that the CAN1 mean values fall close together for two or more dialogues, we will interpret that proximity to imply that the order of composition may be as shown, but that the priority of one or other of the dialogues cannot be guaranteed. Besides, we shall not be concerned with only one set of results, but with 36 or more, which we hope will show some sort of consistency by placing frequently the same dialogue in the same position across a broad range of background variables.

Each run of Candisc uses in turn 10 variables, then nine, and so on down to three, for each of the four sets A, B, C, and D. This gives a total of 32 analyses and to these are added a separate run for each group of variables (ALETS, BLETS, and CLETS) and one other for all the 37 variables. I have not attempted to reproduce the entire printout for each analysis since the sheer bulk of the material precludes it. We have therefore the output for 36 runs of Candisc (32 plus four) and from these it is principally the CAN1 mean value for each work that we are interested in.

All the Platonic dialogues have been included, with the exception only of *Alc.* 2, the Spuria and Definitiones, and the eight shortest epistles. *Alcibiades* 2 is omitted in deference to the arguments of scholars who claim its language is incompatible with a date contemporary with Plato.[11] When included it is placed most often close to *Phaedrus*, and such a late dating is scarcely compatible with its content, providing perhaps additional evidence against authenticity. Results for the 36 Candisc analyses are given as a series of lists of the mean CAN1 values for each of the 40 dialogues. These results are given in Table 9.5.

Each one of these 36 lists claims potentially to present the correct order of composition of the dialogues, but it is probable that, allowing for the amount of extraneous noise which each variable brings along with it, as well as giving information relating to the temporal aspects of style, not one of the lists gets the sequence absolutely correct. I shall be concerned with interpreting the balance of the evidence, to see in which direction it leans, and in this I shall attach more weight to the results from sets B and C, as they are based on a wider range of dialogues and are less likely to be subject to the chance influence of dissimilarities than when only two or at most four dialogues are used as paradigms.

Interpretation

It is clear that the evidence for any one interpretation is not unequivocal, and we shall have to look for trends, rather than undisputed claims for position, although even a modest perusal of these lists shows that the level of agreement is far more secure with the later dialogues. Earlier than the

[11] Guthrie, *History*, v. 387; J. Souilhé, *Platon: Œuvres complètes*, xiii, pt. 2, p. 7 (Budé edn.; Paris, 1930).

Republic, which is the main dialogue of the middle period, a considerable amount of fluctuation is observed. Obviously, adjacent lists will show less difference than those far apart, because only one variable will have been added (or taken away), whereas in the latter case there may be a difference of as many as seven variables. As the number of variables is reduced, the fluctuations in the positions of the dialogues become more extreme, for it is an approach to the univariate situation with all its attendant uncertainties.

It is not possible to decide if there is an optimum number of variables to use from each set, for we cannot tell how much information from each variable is relevant to chronological aspects of style, and how much, for our purposes, is merely irrelevant noise. But I think it is probably best to exercise caution at the lower end of the scale, and to attach greater importance to the middle group of five, six, and seven variables. It is sufficiently far removed from the univariate case to avoid extreme errors, yet not using so many variables as to clutter the enquiry with masses of extraneous detail.

Starting with the later dialogues, I shall attempt to deal with the results in their approximate order of reliability.

Late dialogues

This group remains consistent throughout for all sets of variables. One might expect such results from sets B and C, because they already contain a large proportion of these dialogues as paradigms, but this does not explain the presence of *Critias*, *Epistles* 3, 7, and 8, and the *Epinomis* in both sets, or of the *Laws* with set B and the *Timaeus* with set C.

The full list of late dialogues is as follows:

Philebus
Clitophon
Epistles 3, 7, and 8
Sophist
Politicus
Laws
Epinomis
Timaeus
Critias

Only the set using all CLETS shows any deviation from this list by including *Cratylus* and *Crito*, but it is in other ways an impossibly unreliable list, with *Ion*, *Meno*, *Protagoras*, and *Phaedo* all being shown as post-dating the *Republic*.

35 out of the 36 lists concur in selecting these dialogues as the late group, a group which we feel justified in isolating as self-contained because of the large gap (in the CAN1 value) which always separates the first member of it from the dialogue immediately preceding (see also the previous chapter).

TABLE 9.5 *List of CAN1 mean values for 40 Platonic works based on 36 separate Candisc analyses with selected sets of variables*

Set A

10 variables		9 variables		8 variables		7 variables	
Ly.	−3.2968	Ly.	−3.3580	Ion	−3.2208	Ion	−3.2189
Ep. 2	−3.1913	Ep. 2	−3.1404	Ep. 2	−3.0602	Ep. 2	−3.1498
Hp.Mi.	−3.1864	Hp.Mi.	−3.0978	Hp.Mi.	−3.0334	Ly.	−3.0947
Euthphr.	−2.9719	Euthphr.	−2.9936	Euthphr.	−3.0307	Euthphr.	−3.0191
Ion	−2.9230	Ion	−2.9652	Ly.	−3.0140	Hp.Mi.	−2.9506
Alc. 1	−2.6580	Grg.	−2.7241	Hp.Ma.	−2.8045	Hp.Ma.	−2.6846
Grg.	−2.6560	Alc. 1	−2.6556	Grg.	−2.6370	Grg.	−2.6235
Hp.Ma.	−2.6210	Hp.Ma.	−2.6403	Alc. 1	−2.6162	Alc. 1	−2.6062
Chrm.	−2.6136	Chrm.	−2.5836	Prt.	−2.4028	Men.	−2.4224
Men.	−2.4302	Thg.	−2.3901	Chrm.	−2.3706	Prt.	−2.3769
Thg.	−2.3909	Men.	−2.3591	Men.	−2.3043	Cri.	−2.3695
Phd.	−2.3710	Phd.	−2.3591	Cri.	−2.2440	Chrm.	−2.2787
Prt.	−2.3680	Cri.	−2.3544	Phd.	−2.2205	Thg.	−2.2648
Cri.	−2.3501	Prt.	−2.3513	Thg.	−2.2064	Euthd.	−2.1456
Euthd.	−2.1987	Euthd.	−2.1989	Euthd.	−2.1934	Phd.	−2.1212
Ap.	−2.0282	Ap.	−1.9563	Hipparch.	−2.0080	Hipparch.	−1.9644
Smp.	−1.9546	Smp.	−1.9329	Amat.	−2.0022	Ap.	−1.9337
La.	−1.8270	Hipparch.	−1.7553	Smp.	−1.9913	Amat.	−1.9195
Ep. 13	−1.6749	Amat.	−1.7191	La.	−1.8018	Smp.	−1.9194
Amat.	−1.6468	La.	−1.7154	Ap.	−1.7794	La.	−1.8141
Hipparch.	−1.6440	Ep. 13	−1.6131	Ep. 13	−1.5059	Ep. 13	−1.6221
Prm. 1	−1.3979	Prm. 1	−1.2751	Cra.	−1.2360	Tht.	−1.1914
Rep.	−1.2970	Rep.	−1.2683	Min.	−1.2295	Min.	−1.1515
Cra.	−1.1426	Min.	−1.2321	Prm. 1	−1.1091	Cra.	−1.1309
Tht.	−1.1216	Cra.	−1.1925	Rep.	−1.0965	Rep.	−1.0370
Min.	−0.9434	Tht.	−1.0815	Tht.	−1.0932	Prm. 1	−1.0082
Prm. 2	−0.3935	Phdr.	−0.2821	Phdr.	−0.3803	Phdr.	−0.4634
Phdr.	−0.3197	Prm. 2	−0.1206	Prm. 2	−0.0492	Prm. 2	−0.2201
Mx.	0.7483	Mx.	0.9079	Mx.	0.8132	Mx.	0.7191
Phlb.	1.4887	Phlb.	1.3587	Phlb.	1.3336	Phlb.	1.3179
Clit.	2.0493	Clit.	1.9800	Clit.	1.9855	Clit.	1.8487
Ep. 3	2.2469	Ep. 3	2.2464	Ep. 3	2.2281	Ep. 3	2.1822
Ep. 7	2.4053	Sph.	2.4756	Sph.	2.4285	Ep. 7	2.3667
Sph.	2.5612	Ep. 7	2.5075	Ep. 7	2.5559	Sph.	2.4352
Lg.	3.1355	Lg.	3.0975	Lg.	3.0298	Lg.	2.9843
Epin.	3.3777	Epin.	3.3327	Plt.	3.1279	Epin.	3.1097
Plt.	3.5398	Plt.	3.4385	Epin.	3.2272	Plt.	3.2045
Ti.	4.0466	Ti.	4.1152	Criti.	3.8873	Criti.	8.8584
Ep. 8	4.2762	Ep. 8	4.2481	Ti.	3.9011	Ep. 8	3.9003
Criti.	4.3303	Criti.	4.3732	Ep. 8	4.0597	Ti.	3.9550

TABLE 9.5 *continued*

Set A
continued

6 variables		5 variables		4 variables		3 variables	
Ly.	−3.1522	Ion	−3.1983	Ly.	−3.5459	Hp.Mi.	−3.3246
Euthphr.	−2.9036	Euthphr.	−3.1001	Euthphr.	−3.1311	Euthphr.	−3.0943
Ion	−2.9004	Ly.	−3.0476	Hp.Mi.	−2.9184	Ion	−3.0277
Hp.Mi.	−2.8977	Hp.Mi.	−2.9214	Alc. 1	−2.4142	Ly.	−2.8524
Hp.Ma.	−2.5896	Hp.Ma.	−2.7003	Hp.Ma.	−2.3249	Cri.	−2.4707
Alc. 1	−2.5840	Cri.	−2.6057	Cri.	−2.1837	Alc. 1	−2.2983
Men.	−2.4476	Alc. 1	−2.5858	Grg.	−2.1727	Hp.Ma.	−2.1628
Cri.	−2.4180	Men.	−2.4264	Hipparch.	−1.9356	Grg.	−2.1092
Grg.	−2.4093	Grg.	−2.3902	Chrm.	−1.9301	Hipparch.	−1.9644
Ep. 2	−2.3490	Ep. 2	−2.2562	Men.	−1.8507	Men.	−1.7848
Phd.	−2.1031	Thg.	−2.1026	Phd.	−1.8445	Chrm.	−1.7445
Thg.	−2.0622	Phd.	−2.0935	Euthd.	−1.8317	Euthd.	−1.6751
La.	−1.8947	La.	−1.9413	Thg.	−1.5187	Phd.	−1.5342
Chrm.	−1.8576	Prt.	−1.7589	La.	−1.3224	Thg.	−1.3918
Hipparch.	−1.8203	Chrm.	−1.7276	Ep. 2	−1.2742	Smp.	−1.2009
Prt.	−1.8112	Hipparch.	−1.7228	Prt.	−1.1599	Prt.	−1.1872
Amat.	−1.7244	Ap.	−1.6209	Smp.	−1.1326	Ep. 2	−1.1602
Euthd.	−1.6770	Euthd.	−1.6016	Rep.	−1.1217	La.	−1.1177
Ap.	−1.5846	Smp.	−1.5050	Prm. 2	−1.0204	Rep.	−0.9808
Smp.	−1.5032	Amat.	−1.4492	Ep. 13	−0.8953	Min.	−0.8771
Min.	−1.3748	Ep. 13	−1.3088	Prm. 1	−0.8528	Ep. 13	−0.8630
Ep. 13	−1.2424	Min.	−1.2722	Min.	−0.7768	Prm. 1	−0.5892
Prm. 1	−1.1493	Rep.	−1.0766	Ap.	−0.7224	Amat.	−0.5683
Rep.	−1.1485	Prm. 1	−1.0711	Tht.	−0.7047	Phdr.	−0.4763
Cra.	−1.0977	Cra.	−1.0464	Amat.	−0.6432	Tht.	−0.4149
Prm. 2	−1.0339	Tht.	−0.9977	Phdr.	−0.3965	Ap.	−0.2574
Tht.	−1.0162	Prm. 2	−0.9846	Cra.	−0.3203	Cra.	−0.1528
Phdr.	−0.1776	Phdr.	−0.1989	Phlb.	1.0118	Clit.	0.0382
Mx.	0.5031	Mx.	0.4259	Mx.	1.2267	Prm. 2	0.5142
Phlb.	1.1455	Phlb.	1.1627	Clit.	1.3571	Mx.	0.7114
Clit.	2.1509	Clit.	2.1433	Ep. 3	2.0309	Ep. 3	0.8295
Sph.	2.2926	Ep. 7	2.3678	Ep. 7	2.0643	Phlb.	1.0191
Ep. 7	2.4018	Sph.	2.4188	Sph.	2.3206	Ep. 7	1.6693
Ep. 3	2.6385	Lg.	2.7577	Lg.	2.4569	Criti.	1.8367
Plt.	2.7267	Ep. 3	2.7668	Criti.	2.5553	Lg.	2.0916
Lg.	2.8785	Plt.	2.7779	Plt.	2.6392	Ti.	2.4631
Epin.	3.1137	Epin.	3.1020	Epin.	2.9260	Plt.	2.4886
Ep. 8	3.6833	Ep. 8	3.6858	Ti.	3.0056	Ep. 8	2.5688
Criti.	3.8164	Criti.	3.7295	Ep. 8	3.2669	Sph.	2.6864
Ti.	3.8338	Ti.	3.8760			Epin.	2.8047

TABLE 9.5 *continued*

Set B*

8 variables		7 variables		7 variables		6 variables	
Euthphr.	−3.4267	Euthphr.	−3.1411	Euthphr.	−3.1062	Euthphr.	−3.1081
Ly.	−3.1767	Hp.Mi.	−3.0513	Hp.Mi.	−3.0577	Hp.Mi.	−2.9533
Hp.Mi.	−3.0229	Ion	−2.8796	Ion	−2.8733	Ly.	−2.8996
Ion	−2.9372	Ly.	−2.7903	Hp.Ma.	−2.8024	Hp.Ma.	−2.8671
Hp.Ma.	−2.6390	Hp.Ma.	−2.7894	Ly.	−2.7797	Ion	−2.7153
Thg.	−2.4136	Ep. 2	−2.5603	Ep. 2	−2.5459	Ep. 2	−2.5114
Ep. 2	−2.4011	Alc. 1	−2.4060	Alc. 1	−2.4009	Thg.	−2.3515
Alc. 1	−2.3893	Thg.	−2.3863	Thg.	−2.3658	Alc. 1	−2.3287
Grg.	−2.3883	Grg.	−2.3192	Grg.	−2.2999	Grg.	−2.3172
Min.	−2.2148	Men.	−2.2587	Men.	−2.2552	Men.	−2.2728
Phd.	−2.0968	Cri.	−2.2374	Phd.	−2.2334	Phd.	−2.2004
Men.	−2.0578	Phd.	−2.2201	Cri.	−2.2307	Cri.	−2.1869
Cri.	−2.0522	Min.	−2.0161	Min.	−2.0033	La.	−2.1517
Euthd.	−2.0188	La.	−2.0055	La.	−1.9912	Ap.	−2.0982
Chrm.	−1.9998	Euthd.	−1.9330	Euthd.	−1.9408	Euthd.	−1.9975
La.	−1.8624	Chrm.	−1.9279	Chrm.	−1.9283	Min.	−1.9698
Prt.	−1.7453	Ap.	−1.8963	Ap.	−1.8968	Prt.	−1.8967
Ap.	−1.7238	Prt.	−1.8723	Prt.	−1.8690	Chrm.	−1.8187
Hipparch.	−1.6289	Amat.	−1.5963	Amat.	−1.5911	Amat.	−1.6000
Smp.	−1.4376	Smp.	−1.5281	Smp.	−1.5345	Smp.	−1.5289
Amat.	−1.4043	Hipparch.	−1.4665	Ep. 13	−1.4702	Hipparch.	−1.4962
Prm. 1	−1.4010	Ep. 13	−1.4594	Hipparch.	−1.4698	Cra.	−1.4150
Ep. 13	−1.2497	Cra.	−1.3940	Cra.	−1.4072	Ep. 13	−1.3549
Cra.	−1.1808	Prm. 1	−1.3692	Prm. 1	−1.3749	Prm. 1	−1.3353
Tht.	−1.0667	Tht.	−1.2257	Tht.	−1.2392	Tht.	−1.2127
Rep.	−0.9743	Prm. 2	−0.9835	Prm. 2	−1.0390	Prm. 2	−1.0638
Phdr.	−0.7081	Rep.	−0.9493	Rep.	−0.9439	Rep.	−0.9617
Prm. 2	−0.7037	Phdr.	−0.5355	Phdr.	−0.5292	Phdr.	−0.4650
Mx.	0.3621	Mx.	−0.0046	Mx.	−0.0106	Mx.	−0.0025
Phlb.	1.8139	Phlb.	1.6298	Phlb.	1.6076	Phlb.	1.5663
Ep. 7	2.1083	Ep. 7	2.2800	Ep. 7	2.2903	Ep. 7	2.3377
Clit.	2.3077	Clit.	2.3984	Clit.	2.3924	Clit.	2.3536
Ep. 3	2.3936	Ep. 3	2.6132	Ep. 3	2.6315	Sph.	2.6840
Lg.	2.7345	Sph.	2.7250	Sph.	2.7086	Ep. 3	2.7601
Sph.	2.8161	Lg.	2.8344	Lg.	2.8461	Lg.	2.8564
Ep. 8	2.8437	Ep. 8	3.0973	Epin.	3.0901	Plt.	3.0827
Plt.	3.1218	Plt.	3.1029	Plt.	3.1040	Epin.	3.1836
Ti.	3.1661	Epin.	3.1082	Ep. 8	3.1215	Ep. 8	3.2117
Epin.	3.1808	Ti.	3.4032	Ti.	3.3862	Ti.	3.4668
Criti.	3.7323	Criti.	3.8614	Criti.	3.8310	Criti.	3.8481

* The sequence of no. of variables in sets B and D is non-standard owing to the rejection process implicit in Stepdisc.

TABLE 9.5 *continued*

Set B
continued

6 variables		5 variables		4 variables		3 variables	
Euthphr.	−3.1115	Hp.Mi.	−2.9475	Hp.Ma.	−2.8686	Ap.	−2.8395
Hp.Mi.	−2.9484	Euthphr.	−2.8624	Euthphr.	−2.8556	Hp.Ma.	−2.7581
Ly.	−2.9070	Hp.Ma.	−2.6710	Hp.Mi.	−2.5337	La.	−2.4976
Hp.Ma.	−2.8720	Ion	−2.6366	Ap.	−2.2626	Euthphr.	−2.3147
Ion	−2.7163	Cri.	−2.3076	Ion	−2.2250	Ep. 2	−2.2747
Ep. 2	−2.5046	Ap.	−2.2907	Cri.	−2.1494	Hp.Mi.	−2.2306
Thg.	−2.3513	Men.	−2.1834	Men.	−2.1438	Cri.	−2.1951
Alc. 1	−2.3285	La.	−2.1821	La.	−2.1289	Ion	−2.0735
Grg.	−2.3226	Grg.	−2.1610	Ly.	−2.0167	Mi.	−2.0657
Men.	−2.2673	Ly.	−2.1534	Grg.	−1.9961	Thg.	−1.9715
Phd.	−2.2019	Ep. 2	−2.0996	Ep. 2	−1.9951	Men.	−1.9454
Cri.	−2.1871	Thg.	−2.0393	Prt.	−1.9557	Prt.	−1.8684
La.	−2.1476	Alc. 1	−2.0337	Alc. 1	−1.9248	Cra.	−1.7774
Ap.	−2.0948	Prt.	−1.9985	Thg.	−1.8413	Ly.	−1.7571
Euthd.	−1.9987	Min.	−1.9636	Phd.	−1.8077	Alc. 1	−1.6923
Min.	−1.9887	Phd.	−1.7751	Min.	−1.7969	Grg.	−1.6873
Prt.	−1.8949	Smp.	−1.7258	Cra.	−1.6563	Phd.	−1.6596
Chrm.	−1.8166	Euthd.	−1.4659	Smp.	−1.6207	Smp.	−1.5082
Amat.	−1.6027	Hipparch.	−1.3686	Euthd.	−1.4672	Prm. 1	−1.3994
Smp.	−1.5276	Amat.	−1.3420	Prm. 1	−1.3554	Euthd.	−1.3731
Hipparch.	−1.5032	Cra.	−1.2442	Hipparch.	−1.2421	Ep. 13	−1.0884
Cra.	−1.4194	Chrm.	−1.1399	Amat.	−1.0754	Amat.	−1.0430
Ep. 13	−1.3478	Prm. 1	−0.9530	Chrm.	−1.0021	Tht.	−1.0192
Prm. 1	−1.3294	Ep. 13	−0.8295	Tht.	−0.9132	Mx.	−0.8862
Tht.	−1.2079	Rep.	−0.8011	Ep. 13	−0.8499	Chrm.	−0.7374
Prm. 2	−1.0458	Mx.	−0.7970	Rep.	−0.6790	Hipparch.	−0.6688
Rep.	−0.9622	Tht.	−0.7835	Mx.	−0.4195	Rep.	−0.5915
Phdr.	−0.4592	Phdr.	−0.3736	Prm. 2	−0.3156	Phdr.	−0.1555
Mx.	0.0101	Prm. 2	0.2286	Phdr.	−0.2744	Prm. 2	0.2310
Phlb.	1.5575	Clit.	1.2434	Clit.	1.2710	Clit.	0.9351
Ep. 7	2.3487	Phlb.	1.3921	Phlb.	1.3487	Phlb.	1.2474
Clit.	2.3513	Ep. 7	1.8531	Ep. 7	1.8693	Ep. 7	1.8142
Sph.	2.6792	Ep. 3	2.2714	Lg.	2.3404	Ep. 3	2.1384
Ep. 3	2.7651	Lg.	2.3180	Ep. 3	2.3882	Lg.	2.1412
Lg.	2.8551	Ep. 8	2.7519	Sph.	2.5506	Sph.	2.4363
Plt.	3.0761	Plt.	2.8404	Plt.	2.8789	Plt.	2.5817
Epin.	3.1856	Sph.	2.8771	Epin.	2.8838	Epin.	2.6540
Ep. 8	3.2160	Epin.	3.0895	Criti.	2.9179	Ep. 8	2.8452
Ti.	3.4737	Criti.	3.1193	Ep. 8	3.0496	Ti.	3.1148
Criti.	3.8541	Ti.	3.3347	Ti.	3.0631	Criti.	3.2246

TABLE 9.5 *continued*

Set C

10 variables		9 variables		8 variables		7 variables	
Ly.	−2.9000	Ly.	−2.9362	Ly.	−2.8826	Hp.Ma.	2.8854
Hp.Ma.	−2.7115	Hp.Ma.	−2.7340	Hp.Ma.	−2.8526	Euthphr.	2.8445
Euthphr.	−2.6383	Euthphr.	−2.6569	Euthphr.	−2.6636	Ion	2.8341
Ion	−2.5839	Ion	−2.5846	Ep. 2	−2.4747	Ly.	2.7143
Ep. 2	−2.5710	Ep. 2	−2.5469	Grg.	−2.4717	Cri.	2.5126
Grg.	−2.5162	Grg.	−2.5382	Ion	−2.4595	Alc. 1	2.4614
Alc. 1	−2.4385	Alc. 1	−2.4397	Alc. 1	−2.4180	Grg.	2.4574
Thg.	−2.4120	Thg.	−2.4116	Phd.	−2.3785	Thg.	2.4268
Men.	−2.3682	Phd.	−2.3518	Thg.	−2.3354	Phd.	2.3933
Phd.	−2.3519	Men.	−2.3457	Men.	−2.3115	Ep. 2	2.3671
Cri.	−2.3498	Cri.	−2.3424	Cri.	−2.2996	Men.	2.2832
Hp.Mi.	−2.3068	Hp.Mi.	−2.2923	Prm. 1	−2.2347	Prm. 1	2.1838
Amat.	−2.1411	Amat.	−2.1483	La.	−2.0743	Hp.Mi.	2.1426
La.	−2.1235	La.	−2.1090	Hp.Mi.	−2.0544	La.	2.0587
Chrm.	−2.0664	Chrm.	−2.0518	Ap.	−2.0104	Ap.	1.9688
Ap.	−2.0587	Ap.	−2.0493	Amat.	−1.9930	Chrm.	1.9045
Prm. 1	−2.0357	Prm. 1	−2.0017	Chrm.	−1.9823	Amat.	1.7607
Prt.	−1.8587	Prt.	−1.8563	Prt.	−1.8026	Prt.	1.7129
Smp.	−1.7344	Smp.	−1.7290	Euthd.	−1.6793	Smp.	1.6945
Euthd.	−1.7036	Euthd.	−1.7137	Smp.	−1.6746	Euthd.	1.5372
Hipparch.	−1.3200	Min.	−1.3993	Cra.	−1.5231	Cra.	1.4477
Cra.	−1.3135	Hipparch.	−1.3589	Prm. 2	−1.4920	Prm. 2	1.3915
Min.	−1.3100	Cra.	−1.3360	Min.	−1.2786	Min.	1.3092
Prm. 2	−1.2729	Rep.	−1.2639	Hipparch.	−1.2712	Hipparch.	1.1680
Rep.	−1.2657	Prm. 2	−1.1924	Rep.	−1.2135	Tht.	1.1645
Tht.	−1.1683	Tht.	−1.1468	Tht.	−1.2100	Rep.	1.1554
Ep. 13	−0.9397	Ep. 13	−0.9159	Ep. 13	−0.9500	Ep. 13	0.9718
Phdr.	−0.5326	Phdr.	−0.5081	Phdr.	−0.4861	Phdr.	0.5334
Mx.	0.1563	Mx.	0.2093	Mx.	0.4051	Mx.	−0.2572
Phlb.	1.7404	Phlb.	1.6986	Phlb.	1.6990	Phlb.	−1.7291
Clit.	2.3093	Clit.	2.2985	Clit.	2.3716	Ep. 7	−2.3376
Ep. 7	2.3297	Ep. 7	2.3749	Ep. 7	2.3802	Clit.	−2.3663
Sph.	2.6422	Sph.	2.6193	Sph.	2.4752	Sph.	−2.6542
Ep. 3	2.9077	Ep. 3	2.9149	Ep. 3	2.9834	Lg.	−2.8873
Lg.	3.0118	Lg.	3.0085	Lg.	3.0189	Ep. 3	−3.0860
Plt.	3.2468	Plt.	3.2145	Plt.	3.2462	Plt.	−3.2794
Ep. 8	3.3008	Ep. 8	3.3213	Ep. 8	3.4524	Ep. 8	−3.3622
Epin.	3.7323	Epin.	3.7367	Epin.	3.6682	Epin.	−3.6784
Ti.	3.8097	Ti.	3.8311	Ti.	3.7114	Ti.	−3.7395
Criti.	4.0940	Criti.	4.1125	Criti.	4.0109	Criti.	−3.9772

TABLE 9.5 *continued*

Set C
continued

6 variables		5 variables		4 variables		3 variables	
Hp.Ma.	2.7431	Ion	2.6741	Ion	−2.5417	Ion	−2.6739
Ion	2.6659	Cri.	2.5526	Ep. 2	−2.5369	Thg.	−2.5740
Ep. 2	2.6637	Prm. 1	2.4968	Thg.	−2.5350	Min.	−2.5601
Thg.	2.5772	Thg.	2.4663	Prm. 1	−2.5333	Ep. 2	−2.5590
Cri.	2.5420	Ep. 2	2.4406	Cri.	−2.4994	Euthphr.	−2.5290
Ap.	2.5372	Hp.Ma.	2.4158	Phd.	−2.3205	Alc. 1	−2.4779
Ly.	2.4220	Alc. 1	2.3782	Alc. 1	−2.3113	Hp.Ma.	−2.4764
La.	2.4111	Phd.	2.2695	Hp.Ma.	−2.2871	Cri.	−2.2748
Euthphr.	2.3248	Euthphr.	2.0449	Euthphr.	−2.2025	Ap.	−2.1157
Phd.	2.2481	Ly.	2.0431	Min.	−2.1124	Prm. 1	−2.1104
Prm. 1	2.2429	Min.	2.0057	Men.	−2.0721	Hp.Mi.	−2.0804
Alc. 1	2.2250	Men.	2.0043	Hp.Mi.	−1.9037	Men.	−2.0542
Grg.	2.1251	Grg.	1.9765	Grg.	−1.8717	Cra.	−2.0405
Men.	2.0868	La.	1.9174	La.	−1.8561	La.	−2.0211
Hp.Mi.	1.8268	Hp.Mi.	1.8929	Chrm.	−8.8494	Phd.	−1.8376
Amat.	1.7000	Ap.	1.8184	Ly.	−1.8377	Grg.	−1.7983
Chrm.	1.6573	Amat.	1.7602	Cra.	−1.8106	Ly.	−1.7513
Prt.	1.6149	Chrm.	1.7487	Ap.	−1.8042	Smp.	−1.5867
Min.	1.6011	Cra.	1.6982	Amat.	−1.6617	Prt.	−1.5653
Cra.	1.5866	Smp.	1.4588	Smp.	−1.4676	Ep. 13	−1.2820
Smp.	1.5797	Prt.	1.3411	Prt.	−1.3509	Chrm.	−1.1690
Euthd.	1.4492	Tht.	1.1922	Tht.	−1.2279	Amat.	−1.0482
Tht.	1.2741	Ep. 13	1.1589	Euthd.	−1.1064	Hipparch.	−1.0232
Ep. 13	1.1936	Rep.	1.0456	Prm. 2	−1.0824	Tht.	−1.0063
Rep.	1.0464	Euthd.	1.0281	Ep. 13	−1.0369	Euthd.	−0.9685
Prm. 2	0.8901	Prm. 2	0.8526	Rep.	−0.9344	Rep.	−0.5967
Hipparch.	0.6035	Hipparch.	0.5494	Hipparch.	−0.6052	Prm. 2	−0.4914
Phdr.	0.4018	Phdr.	0.3516	Phdr.	−0.2488	Phdr.	−0.3083
Mx.	0.1655	Mx.	0.1518	Mx.	0.0369	Mx.	−0.2042
Phlb.	−1.6014	Phlb.	−1.4756	Phlb.	1.3598	Phlb.	1.0773
Clit.	−2.0077	Clit.	−1.9444	Clit.	1.8054	Clit.	1.4854
Ep. 7	−2.2913	Ep. 7	−2.2100	Sph.	2.2181	Ep. 3	1.7090
Sph.	−2.5138	Sph.	−2.3183	Ep. 7	2.2801	Ep. 7	2.0075
Lg.	−2.6820	Ep. 3	−2.4536	Lg.	2.5786	Sph.	2.2086
Ep. 3	−2.8442	Lg.	−2.6055	Ep. 3	2.6282	Plt.	2.3702
Plt.	−2.9669	Plt.	−2.6548	Plt.	2.6347	Lg.	2.4666
Ep. 8	−3.1905	Ep. 8	−2.9036	Epin.	2.9925	Ti.	2.7906
Epin.	−3.4225	Epin.	−3.1406	Ep. 8	3.1183	Epin.	2.8556
Ti.	−3.7530	Ti.	−3.4246	Ti.	3.2947	Ep. 8	3.1464
Criti.	−4.3002	Criti.	−4.2446	Criti.	4.3618	Criti.	3.6322

Table 9.5 *continued*

Set D*

9 variables		8 variables		7 variables		7 variables	
Euthphr.	-3.5877	Euthphr.	-3.5857	Euthphr.	-3.4349	Euthphr.	-3.3891
Ly.	-3.2233	Ly.	-3.2230	Ly.	-3.1897	Ly.	-3.1460
Hp.Mi.	-2.9891	Hp.Mi.	-2.9885	Hp.Mi.	-3.0159	Hp.Mi.	-2.9717
Ion	-2.9452	Ion	-2.9468	Ion	-2.9467	Ion	-2.7856
Hp.Ma.	-2.5782	Hp.Ma.	-2.5769	Hp.Ma.	-2.6475	Mp.Ma.	-2.6492
Thg.	-2.5615	Thg.	-2.5602	Thg.	-2.4152	Grg.	-2.3857
Ep. 2	-2.5379	Ep. 2	-2.5372	Grg.	-2.4002	Alc. 1	-2.3593
Alc. 1	-2.4474	Alc. 1	-2.4458	Alc. 1	-2.3919	Thg.	-2.3516
Grg.	-2.3321	Grg.	-2.3320	Ep. 2	-2.3892	Ep. 2	-2.3006
Min.	-2.3163	Min.	-2.3148	Min.	-2.2602	Min.	-2.1206
Men.	-2.1290	Men.	-2.1286	Phd.	-2.1009	Phd.	-2.0981
Phd.	-2.0912	Phd.	-2.0923	Cri.	-2.0542	Men.	-2.0630
Euthd.	-2.0742	Euthd.	-2.0750	Men.	-2.0454	Cri.	-2.0608
Chrm.	-2.0594	Chrm.	-2.0611	Euthd.	-2.0190	Euthd.	-2.0281
Cri.	-2.0033	Cr.	-2.0041	Chrm.	-1.9983	Chrm.	-1.9859
La.	-1.8490	La.	-1.8482	La.	-1.8469	La.	-1.9701
Ap.	-1.8089	Ap.	-1.8083	Prt.	-1.7416	Ap.	-1.7353
Prt.	-1.7622	Prt.	-1.7623	Ap.	-1.7099	Prt.	-1.6961
Hipparch.	-1.5974	Hipparch.	-1.5966	Hipparch.	-1.6453	Hipparch.	-1.5107
Prm. 1	-1.4604	Prm. 1	-1.4607	Smp.	-1.4362	Amat.	-1.3799
Amat.	-1.4198	Amat.	-1.4214	Amat.	-1.4114	Smp.	-1.3551
Smp.	-1.4183	Smp.	-1.4180	Prm. 1	-1.3902	Prm. 1	-1.3206
Ep. 13	-1.2941	Ep. 13	-1.2928	Ep. 13	-1.2383	Ep. 13	-1.1806
Cra.	-1.1789	Cra.	-1.1780	Cra.	-1.1913	Cra.	-1.1498
Tht.	-1.1024	Tht.	-1.1026	Tht.	-1.0565	Tht.	-1.0639
Rep.	-0.9400	Rep.	-0.9407	Rep.	-0.9743	Rep.	-1.0198
Prm. 2	-0.8295	Prm. 2	-0.8262	Phdr.	-0.6978	Prm. 2	-0.7248
Phdr.	-0.7263	Phdr.	-0.7266	Prm. 2	-0.6616	Phdr.	-0.6341
Mx.	0.4259	Mx.	0.4243	Mx.	0.3902	Mx.	0.3123
Phlb.	1.7444	Phlb.	1.7449	Phlb.	1.7967	Phlb.	1.7406
Ep. 7	2.1215	Ep. 7	2.1214	Ep. 7	2.1303	Ep. 7	2.1911
Clit.	2.3991	Clit.	2.3975	Clit.	2.3047	Clit.	2.2721
Ep. 3	2.4130	Ep. 3	2.4150	Ep. 3	2.4008	Ep. 3	2.3977
Lg.	2.7454	Lg.	2.7455	Lg.	2.7318	Lg.	2.7116
Sph.	2.7653	Sph.	2.7654	Sph.	2.8063	Sph.	2.8052
Ep. 8	2.8605	Ep. 8	2.8620	Ep. 8	2.8505	Ep. 8	2.8405
Epin.	3.1074	Epin.	3.1082	Plt.	3.1078	Plt.	3.0839
Plt.	3.1748	Plt.	3.1744	Ti.	3.1782	Epin.	3.2393
Ti.	3.2373	Ti.	3.2376	Epin.	3.1816	Ti.	3.2667
Criti.	3.8676	Criti.	3.8672	Criti.	3.7458	Criti.	3.7689

* The sequence of no. of variables in sets B and D is non-standard owing to the rejection process implicit in Stepdisc.

Table 9.5 *continued*

Set D
continued

6 variables		5 variables		4 variables		3 variables	
Euthphr.	-3.3518	Euthphr.	-3.2903	Euthphr.	-3.0745	Euthphr.	-2.7872
Ly.	-3.2003	Hp.Ma.	-2.7931	Hp.Mi.	-2.6134	Hp.Mi.	-2.6537
Hp.Mi.	-2.9461	Ly.	-2.6822	Hp.Ma.	-2.3926	Hp.Ma.	-2.4830
Ion	-2.6808	Hp.Mi.	-2.1122	Ly.	-2.3173	Ly.	-2.3975
Hp.Ma.	-2.6298	Alc. 1	-1.9688	Min.	-2.0619	Min.	-1.9834
Grg.	-2.3760	Prm. 1	-1.9614	Ion	-2.0530	Ion	-1.8921
Alc. 1	-2.3447	Grg.	-1.9569	Grg.	-1.8901	Alc. 1	-1.7502
Thg.	-2.3244	Phd.	-1.8819	Alc. 1	-1.8063	Phd.	-1.6502
Ep. 2	-2.2500	Min.	-1.8351	Cri.	-1.7589	Grg.	-1.6380
Min.	-2.1296	Men.	-1.8217	La.	-1.6524	Euthd.	-1.5492
Phd.	-2.0883	La.	-1.7865	Men.	-1.6059	Prm. 1	-1.5249
Men.	-2.0721	Ion	-1.7739	Phd.	-1.4965	Cri.	-1.5144
Cri.	-2.0219	Euthd.	-1.7687	Ap.	-1.4940	Men.	-1.3981
Euthd.	-2.0219	Cra.	-1.7397	Prt.	-1.4335	Hipparch.	-1.3552
La.	-1.9995	Thg.	-1.7071	Thg.	-1.4229	La.	-1.2942
Chrm.	-1.9620	Cri.	-1.6503	Euthd.	-1.3816	Smp.	-1.2174
Ap.	-1.7378	Ep. 2	-1.6472	Prm. 1	-1.3332	Prt.	-1.2083
Prt.	-1.6664	Ap.	-1.5577	Smp.	-1.2897	Cra.	-1.2005
Hipparch.	-1.5226	Prt.	-1.4884	Hipparch.	-1.2829	Ap.	-1.1992
Amat.	-1.4064	Chrm.	-1.4178	Cra.	-1.1060	Chrm.	-1.1769
Prm. 1	-1.3379	Hipparch.	-1.2158	Chrm.	-1.0508	Thg.	-1.0653
Smp.	-1.3207	Prm. 2	-1.1854	Ep. 2	-1.0159	Rep.	-0.7194
Cra.	-1.1457	Smp.	-1.0957	Rep.	-0.7562	Tht.	-0.5847
Ep. 13	-1.1191	Tht.	-1.0129	Phdr.	-0.5313	Ep. 2	-0.5809
Tht.	-1.0485	Ep. 13	-0.7694	Tht.	-0.4673	Phdr.	-0.5374
Rep.	-1.0463	Rep.	-0.7547	Amat.	-0.3677	Prm. 2	-0.4457
Prm. 2	-0.8034	Amat.	-0.6067	Ep. 13	-0.1355	Ep. 13	-0.1367
Phdr.	-0.6025	Phdr.	-0.4321	Mx.	-0.0430	Amat.	-0.0780
Mx.	0.3181	Mx.	0.8403	Prm. 2	0.3236	Mx.	0.2358
Phlb.	1.7228	Phlb.	1.6062	Clit.	0.4945	Clit.	0.3790
Ep. 7	2.2008	Clit.	1.7113	Ep. 7	1.4052	Phlb.	1.2784
Clit.	2.2817	Ep. 7	1.9002	Phlb.	1.4858	Ep. 7	1.4863
Ep. 3	2.4229	Ep. 3	2.2238	Ep. 3	1.5249	Criti.	1.5326
Lg.	2.7263	Sph.	2.2351	Lg.	1.9224	Ep. 3	1.5875
Sph.	2.7720	Lg.	2.4279	Criti.	2.0734	Ti.	1.8552
Ep. 8	2.8396	Ti.	2.4768	Ep. 8	2.2237	Lg.	2.0332
Plt.	3.0408	Epin.	2.8038	Ti.	2.3136	Ep. 8	2.4005
Ti.	3.2603	Criti.	2.8802	Sph.	2.7493	Sph.	2.5264
Epin.	3.2622	Plt.	2.9685	Plt.	2.7687	Epin.	2.6635
Criti.	3.7753	Ep. 8	3.0072	Epin.	2.8946	Plt.	2.8054

TABLE 9.5 *continued*

Miscellaneous sets

All ALETS		All BLETS		All CLETS		All variables	
Ly.	−3.6908	Hp.Ma.	2.7127	Ly.	−2.0914	Ly.	−3.7769
Ion	−3.3458	Ep. 2	2.6839	Thg.	−2.0149	Hp.Ma.	−3.4130
Euthphr.	−3.0792	Euthphr.	2.5851	Prm. 2	−1.9831	Ion	−3.3554
Hp.Ma.	−2.8851	Thg.	2.5377	Euthd.	−1.8421	Euthphr.	−3.1939
Hp.Mi.	−2.8507	Ap.	2.5051	Hp.Mi.	−1.7645	Grg.	−3.1478
Grg.	−2.7134	Phd.	2.3764	Hp.Ma.	−1.6777	Hp.Mi.	−3.0957
Thg.	−2.5285	Cri.	2.3600	Euthphr.	−1.6494	Men.	−3.0516
Men.	−2.5093	Ly.	2.3462	Hipparch.	−1.6401	Thg.	−3.0011
Cri.	−2.4857	Ion	2.3445	Chrm.	−1.3899	Chrm.	−2.9855
Chrm.	−2.4542	La.	2.2248	Ep. 2	−1.3631	Alc. 1	−2.8310
Alc. 1	−2.4503	Prm. 1	2.1914	Alc. 1	−1.3423	Ep. 2	−2.6315
Ep. 2	−2.2448	Alc. 1	2.1810	Amat.	−1.1886	Phd.	−2.4824
Ap.	−2.0167	Men.	2.1513	La.	−1.0202	Ap.	−2.4811
La.	−1.9941	Grg.	1.9263	Ap.	−0.9863	La.	−2.4384
Hipparch.	−1.9518	Chrm.	1.8639	Ep. 13	−0.9727	Cri.	−2.4074
Prm. 1	−1.9153	Cra.	1.7451	Grg.	−0.9423	Hipparch.	−2.3361
Phd.	−1.8647	Euthd.	1.6879	Prm.	−0.9194	Amat.	−2.2583
Prm. 2	−1.8306	Hp.Mi.	1.6706	Smp.	−0.9087	Prm. 2	−2.2116
Amat.	−1.7183	Prt.	1.6091	Rep.	−0.8640	Prm. 1	−2.2002
Prt.	−1.6132	Min.	1.5687	Phd.	−0.8261	Prt.	−2.1683
Smp.	−1.4748	Smp.	1.5405	Prt.	−0.8008	Smp.	−1.9918
Euthd.	−1.3631	Prm. 2	1.5115	Men.	−0.7940	Cra.	−1.6427
Rep.	−1.3191	Amat.	1.4470	Tht.	−0.7871	Euthd.	−1.5812
Min.	−1.2403	Tht.	1.4186	Min.	−0.6702	Tht.	−1.5559
Cra.	−1.1411	Ep. 13	1.0729	Ion	−0.5060	Rep.	−1.5374
Tht.	−1.0830	Rep.	0.9309	Phdr.	−0.2524	Min.	−1.4841
Ep. 13	−0.6255	Hipparch.	0.7031	Phlb.	0.0532	Ep. 13	−0.7774
Phdr.	−0.4938	Phdr.	0.2905	Cra.	0.0815	Phdr.	−0.5000
Mx.	0.1615	Mx.	−0.1375	Cri.	0.1198	Mx.	0.1352
Ep. 3	1.6847	Phlb.	−1.4010	Ep. 3	0.3208	Phlb.	2.1612
Ep. 7	1.8410	Clit.	−1.8052	Mx.	0.4826	Clit.	2.5865
Clit.	1.8874	Sph.	−2.1996	Sph.	0.4876	Ep. 7	2.8711
Phlb.	2.0090	Ep. 7	−2.3773	Plt.	0.9575	Sph.	3.0327
Sph.	2.9552	Lg.	−2.7554	Ep. 7	1.3110	Ep. 3	3.1282
Lg.	2.9797	Plt.	−2.9972	Epin.	1.7274	Lg.	3.7149
Ep. 8	3.0568	Epin.	−3.1162	Ti.	1.8175	Plt.	3.8939
Epin.	3.4567	Ep. 3	−3.1309	Lg.	2.1493	Epin.	4.2727
Ti.	3.5792	Ep. 8	−3.4795	Criti.	2.4708	Ti.	4.4753
Plt.	3.8743	Ti.	−3.6100	Clit.	2.6175	Ep. 8	4.5011
Criti.	4.1864	Criti.	−4.4605	Ep. 8	2.8317	Criti.	5.1568

As a group it corresponds with the traditionally accepted list of late dialogues, with the exception of *Clitophon*, a minor dialogue, the authenticity of which has in the past been doubted.

The important points to observe at this stage are that the group contains the six major dialogues, *Philebus*, the *Sophist*, *Politicus*, *Laws*, *Timaeus*, and *Critias*. There is no evidence at all, *pace* Owen, to suggest that the *Timaeus* can be anything other than a late dialogue. If it belonged at all to the middle group we would expect that some traces of this would emerge in at least some of the sequences. But the fact is that the evidence nearly all points in the opposite direction, suggesting that not only are the *Timaeus* and *Critias* late dialogues, but that they were the last productions of Plato's pen, *Critias* being cut short by Plato's death. This displaces the *Laws* from its traditionally accepted position as the final dialogue, and I shall examine this revision of the sequence shortly. For the moment I wish to keep to general points.

The group also contains the three epistles, 3, 7, and 8, and the supposed sequel to the *Laws*, *Epinomis*. The authenticity of these works is a topic which requires separate treatment (see the previous chapter), but if any further confirmation is needed it surely lies here with the correct positioning of them all in the later group. Not only does the stumbling block of genre difference appear to have been surmounted in the case of the epistles, but in most of the lists they are placed approximately where the evidence of dating would conjecturally place them.

Epinomis also seems to be confirmed as authentic by its proximity to the *Laws* and the fact that mostly it follows it in date, differing from it by about the same order of magnitude as the *Politicus* does from the *Sophist*. It appears to be justifiable to accept that this late group consists of the 11 dialogues listed above and I shall now proceed to an examination of the evidence for a more exact chronology of the individual dialogues.

One final point, however, should be made and that is to discount the possibility that these sequences are the result of chance factors unrelated to temporal changes in style. A sufficient proof that this is not so must surely be the tenacity with which the known facts of the dating of the dialogues relative to each other is confirmed in almost every sequence. Thus the content of the late group is consistently maintained and it always post-dates the *Republic*. The *Politicus* follows the *Sophist*, *Critias* is placed after the *Timaeus*, and *Epinomis* post-dates the *Laws*, and these sequences are maintained in about 90% of the lists. This cannot be coincidence, and the interpretation of these lists as chronological sequences is fully vindicated by the discovery that all these relative internal dates are confirmed and the probable dating of the remaining major dialogues is quite compatible with the order in which the majority of the lists appear to arrange them.

Philebus

Here the verdict is almost unanimous in placing *Philebus* as the first dialogue of the late group. Set C, which I take to be the most reliable set of variables, always places it at position 30, at the start of this group. With set B it is preceded in three instances by *Clitophon* (B5, 4, 3). Sets A and D show slightly more fluctuation, with two instances in the former when it moves to 29 or 32 (A3, 4), and two in the latter when *Clitophon* precedes it (D3, 4). However, the movement is not vast and we are dealing only with a minor uncertainty about the relative positioning of *Philebus*, *Clitophon*, and possibly *Epistle* 3. However, the latter only occurs once before *Philebus* and its more likely placing is slightly later, close to *Epistle* 7. The only uncertainty is that of the relative position of *Philebus* and *Clitophon*, and in this I propose to accept the majority verdict in placing *Clitophon* after *Philebus*. It is in a sense a disclaimer of the Socratic approach and hereafter his role in the dialogues diminishes, while in *Philebus* he is still the chief protagonist. Hence the position of *Philebus* as the first dialogue in the late group fits rather well with the Socratic character of the work, and the *Clitophon* may be taken to be the herald of the final dialogues in which Socrates is either a silent listener or completely absent.

The important point, however, is that *Philebus* precedes the other five major dialogues, the *Sophist*, *Politicus*, *Laws*, *Timaeus*, and *Critias*. None of these dialogues is ever shown preceding it and the evidence for its priority is too weighty to ignore. I am aware that this contradicts the conclusions of Brandwood, who placed it as the penultimate work, followed only by the *Laws* (if we ignore *Epinomis* and the epistles).[12]

It also contradicts the opinion of Ross and the sequence favoured by Guthrie (see Chapter 7). But, as far as the stylometric evidence is concerned, I think it is fair to claim that it is a simple case of the unreliability of univariate statistics when applied in such a complex field. Any one variable, when used to determine chronology, is liable to be swamped by variance associated with other factors, genre, subject matter, and so on, and to be an inadequate guide to chronological position. Whether the variable chosen be hiatus avoidance, clausulae rhythm, or an orthographic feature, taken alone it will produce wildly fluctuating results for dates of composition.[13] By using a greater number of variables the results are more reliable because a better definition is achieved of the differences between the samples.

A fuller assessment of the stylometric evidence, therefore, does not lend support to Brandwood's conclusions. Rather, it indicates that we must shift

[12] *A Word Index to Plato*, introduction. See Skemp, *Plato*, for a summary of Brandwood's work.

[13] A single variable can produce only one set of results, which can therefore hardly be described as fluctuating. What I wish to imply is that the results might well disagree violently with the results obtained using any other single variable.

Philebus back to an earlier date at the head of the group. This confirms a conjectural dating of 355 suggested by Wilamowitz and based on Plato's reaction to the death of Eudoxus.[14]

I shall adopt the order *Philebus*, *Clitophon* for the commencement of the late group and will assume an approximate date of 355 for *Philebus*, since, apart from being a convenient starting-point for this group, it fits in well with other indications of absolute dating which we shall have occasion to use for other works.

Epistles *3, 7, and 8*

The position of all three of these among the late dialogues seems to confirm their authenticity, since it is hardly likely that a forger, or pupil, or whoever it might have been, could so successfully hit off the Platonic style that, in respect of 37 rather abstruse measurements which they would never have imagined would be taken, they are fairly and squarely placed among the late dialogues at a position which their conjectured date would require them to be. These conjectured dates are 355 for *Epistle* 3 and 353 for the other two, a difference of dating which is not likely to appear as a significant difference in stylometric studies. However, in most cases *Epistle* 7 precedes *Epistle* 3 (B all, C all except three variables), and *Epistle* 8 is frequently shown as following the *Laws*. I have, nevertheless, grouped them all together, in the sequence 7, 3, 8, and placed them after *Clitophon*, since the date of 355 for *Epistle* 3 would require the placing of it to be close to *Philebus*.

I do not anticipate such accuracy from stylometric dating that it would be able to pinpoint the exact month in which a work was written, and in the case of these epistles, since the authenticity may be taken as well established, the evidence of context and content should be a more satisfactory guide to their relative positions. I have adopted the position and sequence for them which I have indicated, since the stylometric evidence is not sufficiently clear to justify a more certain inference, and we cannot be sure that, for example, part or all of the *Laws* was not written within the period that spans their composition.

Epinomis

I suspect that future generations will find it astonishing that this dialogue should ever have been athetized by commentators. No doubt the original source of this distrust was the comment made by Diogenes Laertius that some people considered it to be the work of Philippus of Opus.[15] But I am convinced that, as we continue to find better ways of unravelling the mysteries of style, the proximity of this work to the *Laws* will become more

[14] U. von Wilamowitz-Moellondorf, *Platon*, 2 vols. (Berlin, 1920); Field, *Plato and His Contemporaries*, p. 75 n.

[15] D.L. iii. 37.

and more apparent. Apart from the tests for authenticity dealt with in the previous chapter, which confirm it as genuine, the fact that in these chronological tests it is shown so consistently to be close to the *Laws* must be a weighty argument in its favour.

All sequences of sets A, B, C, and D show it as following fairly closely after the *Laws*, and their scores are always close. The full set of 37 variables and the ALET and BLET exclusive sets also confirm the proximity of the two dialogues. Its position immediately after the *Laws* seems to be adequately justified by the evidence, and it is that position which we allocate to it in the overall sequence.

The Timaeus *and* Critias

This is perhaps the most controversial part of the revision of the sequence for the late dialogues, for not only is the evidence decisive for including them in this late group, but it also favours a position for them at the extreme end of the list, as the two final dialogues of Plato's life. Set C, which we claim as the most reliable set of variables, gives in seven out of the eight lists the *Timaeus* and *Critias* as the two final dialogues of the series, in that order. For the exception, C3, the closing sequence is the *Timaeus*, *Epinomis*, *Epistle* 8, and *Critias*. Similar results are found with set B, with both dialogues always shown as later than the *Laws*, with *Critias* being placed last in six out of eight instances. In only two out of 36 sequences does *Critias* precede the *Laws*, (A3, D3) and in the last of these both the *Timaeus* and *Critias* precede it.

The weight of evidence is strongly in favour of taking these two dialogues as the closing works of Plato's life. The only ancient evidence which supports this view is found in Plutarch's *Life of Solon*[16] and I give the quotation in full.

Plato was particularly ambitious to create an elaborate masterpiece out of the subject of Atlantis, as if it were a site on some fine estate, which was still unbuilt on, but to which he had a special claim by virtue of his connection with Solon, and he began the task by laying out great porches and enclosures and courtyards on a magnificent scale, such as no story or myth or poetic creation had ever received before. But he was late in beginning and the task proved too long for his lifetime, so that the more we enjoy what he actually wrote, the more we must regret what he left undone. Like the great shrine of Olympian Zeus among the temples of Athens, so among the many beautiful works which Plato's vision conceived, the tale of the lost Atlantis is the only one to be left unfinished (sect. 32).

Perhaps it is surprising that this story has not been given much credence, for it is clear that Atlantis refers to the unfinished work which we know as *Critias*. But, for Plutarch, fidelity to historical truth or chronological detail was never the guiding principle in the composition of the biographies. He was

[16] In *The Rise and Fall of Athens*, trans. and with an introduction by I. Scott-Kilbert (Harmondsworth, 1967).

more interested in the moral portrait, and the typical statement of his attitude in these matters occurs in this *Life of Solon* from which the story of Plato and *Critias* have just been quoted.

However, when a story is so celebrated and is vouched for by so many authorities and, more important still, when it is so much in keeping with Solon's character and bears the stamp of his wisdom and greatness of mind, I cannot agree that it should be rejected because of the so-called rules of chronology, which innumerable authors have continued to revise, without ever being able to this day to reconcile their inconsistencies. (sect. 27)

The source of Plutarch's information for this anecdote of Plato is not known, but, in any case, our concern here is not to test the veracity of Plutarch, but to assess the stylometric evidence for placing the *Timaeus* and *Critias* at the end of Plato's life. And this the most reliable source of evidence confirms overwhelmingly (sets B and C) or, if one prefers the wider approach of utilizing all 36 variable sets, the verdict is 34 against two, or 94% in favour of putting these two dialogues after the *Laws*. The fact that Plutarch's anecdote confirms this dating is an added bonus, but in view of the uncertain nature of so much of the ancient evidence, it is not something which we need to rely on.

The question of course does arise as to whether or not the final position of the *Timaeus* creates difficulties and anomalies in our interpretation of Platonic philosophy, or, indeed, whether or not there lurks a fundamental flaw at the root of his system which he himself came to realize in the closing years of his life. Interpretations of the *Timaeus* are varied, but I am more inclined to read it as an extension of the former philosophy rather than a curtailment. The entire universe, its structure and development must be described and explained, so that we may understand how the whole of material existence fits into the scheme of a divine and transcendent reality which it is the duty of us all to consider and contemplate.

In a sense this concern with what is basic and fundamental, the raw material of everyone's experience, should be the first and abiding interest for the philosopher, and not left till the end of his life. But perhaps it is only possible to write coherently about such matters after a lifetime of thought and probing of the contradictions of existence. However, there may be a more mundane explanation for Plato's late interest in natural philosophy (if that is how one should interpret the *Timaeus*), namely the stimulus provided by his contact with Western science in the person of Archytas of Tarentum and the Pythagoreans, with their emphasis on a mathematical explanation for the phenomena of the visible world.

The philosophical consequences of accepting a late dating for the *Timaeus* are, however, considerably more complex than my brief outline suggests. Owen perceived a need for relocating this dialogue at an earlier period in

Plato's career because of its apparent separation from Plato's later stance regarding the Theory of Forms.[17] Yet it is clear from the evidence presented above that the stylometric characteristics of the *Timaeus* do not support his claims, however we may choose to interpret its philosophical content. In any case, the questions which arise in connection with the development of Plato's philosophical thought cannot be dealt with here.

On a purely biographical level, however, one cannot fail to be struck by the irony of the situation in which Plato, at the very end of his life, appears to cling still to the fiction that Critias, that noble reformer, was the ideal person to reveal the workings of a perfect society as it existed in the golden age of the age of the legendary history of Athens. For this is the Critias who was the leader of the Thirty Tyrants and the most bloody exponent of the methods of that interim government, with a string of murders, wholesale massacres, and sequestrations of property to his credit.[18] Yet here, in Plato's hands, he becomes the urbane and genteel historian who is to reveal the ideal State in action by giving a description of Athens and Athenian society as it existed several thousand years previously when it was forced to contend with the mighty empire of Atlantis. 'For indeed,' says Critias, 'the Athenian State of that time was the finest with regard to warfare, and superlatively governed. For it is said that the most beautiful works of art and architecture were to be found there, and its constitution was the most perfect of all those that we have heard tell under the sun' (*Ti.* 23 C). By comparison with this golden age it would be easy to see the extent to which present-day society had become rank and corrupted.

But how are we to react to this presentation of Critias by Plato in this his final dialogue? For there can be little doubt that the Athens to be described will be yet another utopian tyranny. There are several possibilities, some of them highly disturbing. Are we to read it as the final slap in the face by an embittered old man of the democracy which had murdered Socrates and

---

[17] See G. E. L. Owen, 'The Place of the *Timaeus* in Plato's Dialogues', in R. E. Allen (ed.), *Studies in Plato's Metaphysics* (London, 1965), and the reply by H. F. Cherniss, 'Parmenides and the *Parmenides* of Plato', *AJP*, 53 (1932), 122 ff. The suggestion made by Owen that *Ti.* should properly belong to the middle period rather than to that of the later dialogues by no means commanded universal support. Details of the controversy may be found in Guthrie, *History*, v. 243.

[18] Details may be found in *HG* ii. 3–4 and *Mem.* i. 2, and Lysias, *Oration* 12. If, as has been suggested (according to Guthrie it is now the general opinion, *History*, v. 244 n. 1), Plato's Critias was in fact the grandfather of the tyrant, then of course Plato is absolved from the charge of bias and sympathy towards the Tyrants. But he certainly does not go out of his way to make it clear that it is not the character of recent times who is the chief speaker in the dialogue. Critias, in any case, is introduced in other dialogues (*Chrm.*, *Prt.*) where he shows no promise of future evil, and Charmides, another of the Thirty, also gives his name to a dialogue, in which he appears as a bashful and charming youth—not exactly the best way to placate those who had suffered under the Thirty, even though Plato's account of their characters, as they were at that stage of their lives, may not have been a departure from historical truth. (For a recent appraisal of the political sympathies of Plato and his circle see I. F. Stone, *The Trial of Socrates* (London, 1988).)

---

dishonoured philosophy? For a work which lauded Critias cannot have been other than offensive to the many who had suffered under his regime and its terrors. Perhaps Plato could count on a sympathetic audience within the limited circle of the Academy, many of its members probably owing allegiance to the oligarchic faction in Athens by virtue of wealth or family connections.[19] In such a circle perhaps the tradition concerning Critias would have been less severe, and he may have been seen as the thwarted philosopher rather than the evil tyrant.[20] Plato himself owed him some measure of family loyalty (he and Critias were cousins), yet it is odd that he should choose him as the central character of a dialogue which was essentially a work of fiction, when so many other choices would have been open to him.

There is also the possibility of some sort of psychological blockage, an inability to accept that the man he had known as a friend, relative, philosopher, dramatist, conversationalist, and man of letters could have been guilty of so much wrong. Secretly, perhaps he blamed the people of Athens for their intractability, for their refusal to submit to reforms, and the innate perversity of those who persisted in maintaining the democratic traditions.

We must conclude either that he did not consider that his outbursts against democracy, whether open and explicit (as for example in the *Republic* and *Gorgias*) or implicit (in the *Charmides* and *Critias*), were offensive,[21] or that he expected them to be read by sympathetic readers. Alternatively, he might have intended them to be deliberately provocative and believed passionately that right was on his side. For the only note of reserve to be found in his works concerning this short interlude of Athenian history (the period of the Thirty Tyrants) occurs in an autobiographical section of *Epistle* 7.

The constitution at that time was reviled by many, and a revolution occurred . . . in which thirty rulers were appointed having absolute power. Some of these were relatives and acquaintances of mine, and they invited me immediately to participate in what (they implied) would be congenial employment. . . . I imagined that they would rule the city in such a way as to lead it from its evil path into a just way of life, and consequently watched them very carefully to see what they would do. And indeed I observed those men in a short time making the former constitution appear to be one of gold by comparison . . . So when I saw all these things happening, and many others of no small moment, I was appalled and withdrew myself from the evils of the time. Shortly afterwards the rule of the Thirty crumbled, together with the whole of the government then existing. (324 C–325 A)

---

[19] Among these people there were probably some who had belonged to the Council of 500 or the list of 3000 enfranchised citizens set up by the Thirty, many of whom, by their participation in the events of the time, would have compromised their standing with the democracy. (See N. G. L. Hammond, *A History of Greece to 322BC* (Oxford, 1959), 443–7).

[20] Aristotle (*Pol.* 1305ᵇ26) gives Charicles as the leader of the Thirty and in the *Constitution of Athens* (35–38) no names of the Tyrants are given. Perhaps this is due to his long period of study and tutelage in the Academy.

[21] Several passages in these dialogues are critical of democracy, e.g. *Grg.* 518 E–519 B, *Rep.* 557 A–558 B. *Critias* and *Charmides* offend by their choice of central character.

He also concedes, generously, that the re-established democracy exercised considerable restraint—'the returning exiles behaved with great moderation' (325 B).

In the light of this passage it is perhaps rather unfair to dwell on Plato's unsympathetic attitude to the democracy at Athens and to democracy in general. Yet the whole episode of the tyranny and Plato's response to it revealed in this letter, as well as his subsequent depiction of two of its architects, Critias and Charmides, in the nonage and untainted period of their lives (one might also add his portrayal of Alcibiades in *Alcibiades* 1 and the *Symposium*) must cause us to view with misgivings his proposals for the participation by philosophers in government. If he could depict Critias, Charmides, and Alcibiades so rosily, should we really be so uncritical of his judgement of Dion as to accept that, in the débâcle at Syracuse, all the fault was on the side of Dionysius?[22]

I return now to the chronological ordering of the remaining works. Thus far then the sequence for the late dialogues is that they commence with *Philebus*, *Clitophon* is second, and the group ends with the *Timaeus* and *Critias* following on from the *Laws*.

The Laws

This work is by far the longest of Plato's dialogues[23] and as such may have taken several years to write. Its lateness seems to be confirmed by the statement of Diogenes Laertius that it was left 'in the wax' at Plato's death,[24] by the reference to a battle which is thought to date to 356, and by overwhelming tradition. Ryle is perhaps the only dissenter,[25] for he sees part of it as having been written for an earlier Sicilian visit (the second), but stylometry reveals no evidence of a dichotomy of style for this work, a split between early and late dates of composition. Ryle's work on the chronology of Plato's dialogues reads, in any case, like an exercise in iconoclasm rather than a serious contribution to Platonic scholarship. However, it does help to emphasize the point that perhaps it is rather meaningless to insist on a fixed date or position for the *Laws* in the sequence of composition. It may be that works such as the *Sophist* or *Politicus* were written contemporaneously, or that continuous revision of the *Laws* was interrupted only by Plato's death.

There is of course a slight contradiction between this report in D.L. that

---

[22] Dionysius was the young tyrant of Syracuse. Plato had hoped that, with the help of Dion, uncle of the tyrant, he would be able to establish in Syracuse a model form of government which would show to the world a State governed by a philosopher-king. The plan did not work and Dion was banished, to be assassinated some years later on his return from exile. Plato's embroilment in Sicilian affairs was not only painful, through the loss of a close friend, but one suspects that it was also damaging to his reputation.

[23] Approximately 100 000 words. This total does not include the character names which provide the speech headings for the dialogue, all of which are ignored in this study. TLG gives it as 106 297 words. It provides 90 1000-word samples in my analyses.

[24] D.L. iii. 37.   [25] *Plato's Progress.*

the *Laws* was 'in the wax' (ἐν κηρῷ) at the time of Plato's death and the story told in Plutarch that *Critias* was the final work, although the former dialogue cannot claim to be unfinished in any sense other than that of perhaps lacking a final polish, whereas *Critias* breaks off in mid-flow like Schubert's unfinished symphony. I would interpret D.L.'s comment, if it is true, to mean that the *Laws* had not, like the other works of Plato, been published when Plato died, but existed only on the waxen tablets from which the final fair copies would have been made. The post-dating of the *Laws* by the *Timaeus* and *Critias* suggests that it was put aside to allow work to be done on these two other dialogues, although it is clear that the unfinished state of *Critias* implies that it too must have been 'in the wax' when Plato died. Its publication would have depended on the beneficent offices of his friends.

However, to turn to the stylometric evidence, which is obviously inter-linked for all these dialogues which are close in date, we find that the *Laws* fluctuates between positions 34 and 36, the most common sequence being the *Laws*, *Politicus*, *Epinomis*, *Epistle* 8, *Timaeus*, *Critias*. For convenience I take *Politicus* alongside the *Sophist*, to which work it is the professed sequel, and I place *Epistle* 8 with the other two late epistles, thereby effectively shifting the *Laws* to position 37. This does not represent a gross violation of the evidence and is more satisfactory than having the *Sophist* and *Politicus* split asunder, two dialogues which have a strong natural affinity, although there must be something in the character of the latter which causes it to be placed on three occasions as the final or penultimate dialogue (D5, 4, 3).[26] The closing sequence, therefore, becomes the *Laws*, *Epinomis*, *Timaeus*, *Critias*, a se-quence which has a certain tidiness which perhaps belies the reality of the situation. The fact that *Epistle* 8 so frequently occurs later than the *Laws* may imply that it was written before that work was completed and, should anyone insist that such must have been the case, the stylometric evidence would bear that interpretation.

However, these are surely minor matters, for the main import of the sequence which I have adopted is that the *Laws* precedes the *Timaeus* and *Critias*, and for that I believe stylometry provides adequate proof.

The Sophist *and* Politicus

The sequence of these two dialogues is occasionally reversed (e.g. A3, B5). Of course it is not impossible that Plato could have written the *Sophist* after *Politicus* and merely inserted the necessary references to the latter work so as to maintain the pretence that it was written as the second of the two works. There is no obvious reason, however, why Plato should choose to play such

---

[26] The set of D variables is based on *Sph.* and *Plt.* only as the late group, and I suspect that this results in a selection which emphasizes certain characteristics of these two dialogues which are not found elsewhere in the late group. This results in a distortion of the chronological pattern when applied to all the dialogues.

tricks, and it seems perfectly reasonable to accept the sequence the *Sophist, Politicus* as it is indicated in both of the dialogues.

They form in fact two parts of a proposed trilogy,[27] the third unwritten or at least not extant part having the philosopher as its chosen subject.

The bulk of the evidence (90%) shows the expected sequence the *Sophist, Politicus*, but they are also usually shown as being separated by the *Laws*, a point which was mentioned in the preceding section. Nevertheless, I have kept the two together, as they do form an undoubted pair, and nothing much is to be gained from splitting the two asunder. 21 out of the 36 lists, or 58%, show the *Laws* occurring between these two works, and in some cases the distances between the CAN1 values are not high, so that, with this sort of slightly hazy evidence, it is pointless to insist on an absolutely rigid relative position for the three works. The *Sophist* does in fact precede the *Laws* in the majority of cases (69%), and it is often very close to it in value. It is often also shown to be close to *Epistle* 7 and immediately following it, and I therefore adopt the sequence the epistles, *Sophist, Politicus, Laws*. Should anyone insist that the *Laws* must have preceded *Politicus*, I shall be the first to admit that the stylometric evidence does allow some latitude of movement in this respect.

Clitophon

In this dialogue Clitophon praises Socrates for his exertions in turning men to the path of justice and the pursuit of virtue, but follows this with a criticism that Socrates' teaching goes no further than being merely exhortational—his art is προτρεπτική but not substantial, for he cannot follow up his preaching with a declaration of the true products of justice. The art of medicine produces health, that of carpentry produces wooden utensils, therefore what does this art of justice actually produce? Socrates has shown himself unwilling to answer these questions in the past, either because he does not know the answers, or because he is unwilling to share his knowledge with others. His closest disciples also prove themselves to be unable to resolve the dilemma. Clitophon will therefore betake himself to Thrasymachus, or to anyone else who might help him, and will in future mix his praise of Socrates with some element of blame, when he speaks with Lysias or others with whom he is accustomed to converse.

To all this Socrates makes no reply and we are left in doubt as to the sequel. The criticism may perhaps be self-criticism on the part of Plato and constitutes, in some sense, a recognition of a deficiency in some of the dialogues which pre-date this one. The aporetic tactics of Socrates were not in themselves sufficient to provide the basis for building the good life, the life of virtue, as that requires something in the shape of more positive instruction.

---

[27] Guthrie, *History*, v. 123 and n. 1.

However, this still leaves the great difficulty of dating the dialogue, for the purely aporetic works (also known as Socratic) are usually taken to be the earliest. By the time of *Phaedo* Socrates was already putting forward his own ideas about virtue, rather than merely refuting those of others. *Phaedo, Crito, Meno*, the *Symposium*, and the *Republic* are all 'positive' dialogues in that they show Socrates expounding doctrines about love, immortality, obedience, justice, which are in no sense negative or refutative for the sake of refutation. The *Republic* undertakes to show us what justice is in itself, αὖτο το δίκαιον shorn of all accretions which might make it attractive superficially to the individual, as power, prestige, position, respect of one's fellows, and so on. In this context *Clitophon* appears to be rather meaningless.

The references to Lysias and Thrasymachus seem to bring us to a later date, subsequent to the *Republic* and *Phaedrus*, if mere names are anything to go by. And there are some undoubted stylistic quirks which point to the later group rather than to an earlier date of composition. The last sentence is a good example of this, commencing as it does with an indirect object, the direct object, σε, buried almost out of sight within it, and the main verb, φήσω, occurring half-way through and acting as a sort of pivot for the two halves of the sentence.

μὴ μὲν γὰρ προτετραμμένῳ σε ἀνθρώπῳ, ὦ Σώκρατες, ἄξιον εἶναι τοῦ παντὸς φήσω, προτετραμμένῳ δὲ σχεδὸν καὶ ἐμπόδιον τοῦ πρὸς τέλος ἀρετῆς ἐλθόντα εὐδαίμονα γενέσθαι.

(To a man who has not yet been converted, Socrates, I would declare you to be of immense value, but to one who has already seen the light you are more of a hindrance than a help in the achievement of acquiring perfect virtue.)

A similar case of inversion occurs at 408 B (τούτοις δὴ τοῖς λόγοις, etc.), and generally, throughout the dialogue, there is a more conscious use of antithesis and other rhetorical devices than one is accustomed to find in dialogues of the early and middle groups (e.g. 407 E, καὶ δεῖν ἐπιμέλειαν, etc.).

However, such linguistic habits cannot be interpreted decisively, but the stylometric evidence is quite decisive in placing *Clitophon* with the late dialogues, and very early in the sequence, more or less contemporary with *Philebus*. Sets B and C are in favour of putting *Clitophon* in second place, by a margin of 13 to three, and on that recommendation I will adopt the sequence *Philebus, Clitophon, Epistle* 7.

My interpretation of the dialogue is that it represents for Plato 'a long farewell to Socrates and all his greatness'.[28] The ensuing dialogues owe less to the inspiration of Socrates than they do to Plato's own philosophical ideas. If Socrates figures at all in them it will be only as a silent listener, but he will not himself contribute anything. The rest of Plato's life was to be spent building the ethical and ontological foundations which were to sustain the life of the

---

[28] Shakespeare, *Henry VIII*, III. ii, Wolsey on his downfall.

true philosopher. The *Sophist* deals with the problems of definition and reality, *Politicus* with good and evil in States and with statesmanship; the *Laws* examines the principles and practice of law-making and its influence on society, the *Timaeus* seeks to explain the visible universe in terms of abstract elements, so that, by linking the visible world and the world of phenomena to their models in abstract ideas, the reality of the Ideas or Forms is assured. On this basis the task of the philosopher is to attain to knowledge of the forms, for only then will true knowledge of reality be possible, and knowledge also of that mimic of reality, the physical world which is accessible to the senses.

Summary of dating for late group

The order which I have deduced for the late dialogues is the following:

> *Philebus*
> *Clitophon*
> *Epistle* 7
> *Epistle* 3
> *Epistle* 8
> *Sophist*
> *Politicus*
> *Laws*
> *Epinomis*
> *Timaeus*
> *Critias*

All these dialogues would have been written after Plato's return from his final visit to Sicily in 361–360. If we take 355 as a plausible date for *Philebus* and 369 as an approximate terminus for the *Theaetetus*, perhaps the last of the middle dialogues, we have a gap of approximately 14 years, during which Plato visited Sicily twice, in 367–6 and in 361–0, and within which period very little was written by him. The exception is *Phaedrus*, which, as we shall see later, acts as the bridge between the middle and late dialogues.

This gap should be sufficient to explain the great shift of style which is found to occur between the middle- and late-period dialogues, a shift which is emphasized in the 36 sequential listings by a sharp change of values for those dialogues which are close to the *Republic* and those which are more akin to *Philebus*. A typical value of CAN1 for the latter is 1.5, while the *Republic* might score − 1.0, so that the gross difference between these two works is of the order of 2.5 units. This compares with a typical difference between adjacent dialogues of 0.1 or 0.2, only a fraction of the difference between the middle and late dialogues. In fact for all dialogues earlier than the *Republic* the change in value between adjacent dialogues is often only marginal, and clearly well within the limits of statistical variation; whereas the change which is heralded by *Philebus* is sufficient to suggest that here we

are dealing with a different author, and, were our knowledge of Plato more restricted, it is probable that we would be tempted to draw that conclusion.

However, the change is perhaps best explained by a lapse from writing which lasted for some 12 or 14 years, during which time Plato changed from a youthful 58 (if one can be youthful at that age) to an old man of 70 or more. His mind cannot but have been affected by the Sicilian venture. His great friend Dion had been banished and finally murdered and he had devoted much energy, mostly wasted, to the task of reforming Dionysius' character, an effort which had threatened his own life and left him stranded in Sicily at the mercy of the tyrant's whims. All this would have been sufficient to produce such a radical alteration of style, if old age itself were not considered to be an adequate cause.

Tentatively, therefore, I would append the following dates to the dialogues of this period:

361–0	Final visit to Sicily
355	*Philebus*
	Clitophon
	Epistle 7
353–2	*Epistle* 3
	Epistle 8
352?	*Sophist*
	Politicus
− 351?	*Laws*
	Epinomis
349–8	*Timaeus*
	Critias
348–7	Death of Plato

This list should be compared with those given in Table 7.1.

The early and middle dialogues

Phaedrus

This dialogue gives every indication of being transitional. It occupies the middle ground of the two major stylistic periods of Plato's writing, its score usually bringing it closer to the group surrounding the *Republic* than to the *Philebus* group. This must place it somewhere in the period 369–355, to set the limits as widely as possible. If one succumbs to the youthful charm of the work which caused one nineteenth-century critic to choose it as the first of Plato's dialogues, then it should belong somewhere close to the beginning of this period, say in 368. It seems to be preferable to date it to about that year, rather than later, because it is in any case much closer to the earlier of the two groups, and because the second visit to Sicily, which belongs to the years

367–6, must have cast a considerable shadow over his mind, involving, as it did, the expulsion of Dion and the rather awkward detention of himself in Sicily, a detention which on occasion brought his life into some danger.

But *Phaedrus* does not appear to be tinged with any disillusionment or darkened with the shadow of disaster, so that one tends to favour a date which places it before these adventures.

It could be argued that the stylometric evidence is in this instance misleading, as so much of the dialogue is imitative. But there is so much consistency in the stylometric dating of this dialogue that the argument is not very convincing. With sets B and C it is always placed at positions 27 or 28, while with A and D some more uncertainty is evident, so that it ranges between 24 and 29. It is never found with the later group, but always with the group of dialogues immediately preceding *Philebus*. This contrasts strongly with other dialogues which are difficult to place, such as the *Apology*, which is found as early as 1 or as late as 25 in the sequence (B3, A3), or *Minos*, a dialogue of doubtful authenticity, which ranges from 5 to 26 in the sequence of composition (D4 and 5, all 37).

One has to be consistent in either rejecting or accepting the stylometric evidence, for, generally, speaking, the arguments in favour of a different interpretation of a set of results are often just as applicable to other dialogues. There are just as many alien speeches in the *Symposium*, and *Protagoras* also, with its many characters each seeking to shine in his own way, must fall under suspicion as being not entirely Platonic. The eschatological myths of the *Gorgias* and *Republic* could also be claimed as unrepresentative of the typical Platonic dialogue, resulting in some distortion of their true positions.

However, where there is reasonable consistency, it seems to be right to accept the judgement of stylometry, and to see whether or not it is possible to fit some explanation of the development of Plato's mind to the sequence as it is revealed.

On these grounds, and using the available evidence, I shall therefore place *Phaedrus* after the middle period, but prior to the later group, as a transitional dialogue, but one which bears stronger affinities with the earlier than the later group. I will not affix a date to it, since we do not have any indications from other sources, either internal references, or by its proximity to other datable dialogues, as to what it should be, except to state that it does fall within the period 369–355, and probably towards the earlier end of these limits.

Menexenus

This dialogue presents considerable difficulties for dating by stylometric means alone, because it does not have any parallels in the rest of the corpus and because its genre is distinctive enough to isolate it from the predominant

stylistic features of the other dialogues. It belongs to that special class of oratory reserved for the funeral oration, or *epitaphios*, but in this case that very distinctive *epitaphios* which was delivered annually at Athens on the occasion of a ceremony performed to commemorate and laud all those citizens who had fallen in battle over the preceding year while defending their country.

Other extant examples of the type are Thucydides' version of Pericles' speech, delivered in 431,[29] the *epitaphios* of Lysias,[30] and one by Hyperides, perhaps as it was originally composed.[31]

The common theme of all, it would appear, is 'Greater love hath no man than this' and 'My country, how glorious, how noble, past, present, and future'. It would be facile to claim that having heard one of them you've heard them all, but that is almost what Plato's introductory comments imply. It is an easy matter to praise Athenians before Athenians (235 D) and Aspasia, Socrates claims, patched up the speech which he subsequently recites using bits which had previously been prepared for Pericles' famous examples (236 B). Any worthwhile orator could throw one of these things together, even improvise it, and Socrates, when he listens to their productions, is usually spellbound and imagines himself to have become taller, nobler, more handsome, and more majestic in the eyes of any visiting foreigners (235 A–B). Only on the fourth or fifth day after does he come down to earth and realize that he is no longer living in the Islands of the Blest (235 C).

This is all pretty entertaining stuff, but it is the implications for stylometry which we must consider, especially as four-fifths of the *Menexenus* consists of Aspasia's speech, a far greater proportion than the alien speeches in any of the other dialogues. Not only this, but, apart from being imitative (and in that sense un-Platonic), its genre is so distinctive that it must stand isolated from the other dialogues as an individual item having few or no links with any of them.

In fact, in the CAN1 lists, it is usually found occupying the middle area between late and middle dialogues, i.e. between *Phaedrus* and *Philebus*, although sometimes it creeps higher to be placed somewhat above the *Republic* (e.g. B5). While it is not impossible that the *Menexenus* could have been written this late, this is not the opinion of most scholars, who prefer to date it to *c*.387 on the basis of Socrates' reference to the Peace of Antalcidas of that year (245 E), when Socrates, who recites the speech, had been dead for 12 years, and Aspasia probably even longer, a piece of buffoonery which would perhaps be meaningless if composed 20 years later, say in 366.

My own misgiving about this comparatively late dating springs from a different source, namely that the CAN1 values for the first sample of the four

[29] Th. ii. 34.   [30] Lysias, *Oration* 2.
[31] Hyperides, *Funeral Oration*, delivered in 322. It does not survive complete.

into which the *Menexenus* is divided, the sample which contains the introductory section and only a small portion of Aspasia's speech, scores consistently with a higher (or lower) value than the other three samples which consist entirely of Aspasia's speech. The score is usually not sufficient to place it as early as the *Gorgias*, but nevertheless implies that, freed of the bias of genre, it would not occupy the comparatively late position which the CAN1 lists assign to it.

This is not to be interpreted as implying different dates of composition, but that genre is dominant for the *epitaphios* and sufficiently strong to overcome the temporal characteristics which we are attempting to isolate. As a general rule I am not in favour of looking at scores for individual samples, because there is so much fluctuation that it would be virtually impossible to arrive at any conclusions on that basis, and because it is the macro-effect that is the chief interest in this investigation.

Nevertheless, in the case of the *Menexenus*, I am wary of its position and suspect that genre bias is responsible for flinging it into the comparatively empty gap between early and late dialogues. I have therefore placed it, somewhat arbitrarily, after the *Gorgias*, not on the basis of computer evidence, but because that is where many scholars insist it belongs.[32] The stylometric results, when examined in detail, do suggest that its position could well be earlier than that shown in the series of 36 CAN1 listings, and therefore there is some justification for making the shift, but it is in any case a dialogue (or speech) which will always remain an anomaly in Plato's writing (see previous chapter for discussion of stylometric tests on authenticity for this dialogue).

The middle group

Proceeding further up the list we find that the task of discovering the most probable sequence becomes increasingly difficult. Greater fluctuations are observed in the positions of individual dialogues and the distances which separate them are less pronounced. However, it is still possible to observe certain fairly constant features, although they do not have the definition and consistency of recurrence that was found for the dialogues of the later period.

There seems to be a reasonably well-defined group of middle dialogues, although the point at which the upper boundary is to be drawn is somewhat arbitrary. I include in the group the following dialogues: *Cratylus, Euthydemus*, the *Parmenides*, the *Republic*, the *Symposium*, the *Theaetetus, Epistle* 13.

I have split the *Parmenides* into two sections, the first four samples being the introductory part, the remainder consisting of Parmenides' discourse on the prospect of 'the one' or 'the others' existing. This I felt to be necessary because of the idiosyncrasies of Parmenides' discourse. All the stylometric

[32] Dodds, *Plato, Gorgias*, pp. 23–5; Guthrie, *History*, iv. 313.

tests show it to be very different from Plato's typical style, a fact which is self-evident to any reader and does not require the mechanisms of stylometry to prove. The extent to which it differs from the other dialogues is dealt with in the previous chapter, where I point out that most tests of authorship would lead us to conclude that it was not written by Plato. I interpret it, however, as a sustained piece of imitative writing based on a possible Zenonian original, or some other philosopher of the Eleatic school, and the most derivative of all Plato's compositions. The peculiar nature of this work, with its limited vocabulary and frequent repetition of key words ἕν, ἄλλα, its staccato series of question and response, and its sustained intensity of purpose makes it so unlike anything else of Plato's that it seemed probable that the chronological tests, when applied to this dialogue, would not work in the same way and would be liable to give a misleading result. By splitting it into two parts, a Platonic and a Parmenidean section, one at least has the chance of seeing how close the two are to each other, and this possibly provides some check on the idiosyncratic bias of the second section (*Prm.* 2).

In eight out of the 36 listings of CAN1, or 25%, the two sections are found to be adjacent or separated by only one other dialogue, while 50% are found to separate the two parts by five or fewer dialogues. I suspect that any other dialogue of comparable length and belonging to this or an earlier period, if split into two sections, would show only a marginally improved variation on that found with the *Parmenides*, although there might well be fewer of the extreme values recorded, as for example in list C4 where *Prm.* 1 is at position 4 and *Prm.* 2 at position 24.

All the dialogues of this period, in fact all prior to *Phaedrus*, differ from the adjacent dialogues by small amounts relative to their absolute scores. A typical range of values is from 3.0 for the top CAN1 score of the earliest dialogue, falling to 1.0 for the dialogue closest to *Phaedrus*. For the 27 dialogues prior to the later group this only gives an average separation of 0.074 between adjacent values of CAN1, so that random fluctuations are liable to cause considerable variation in the position of individual dialogues. This is a reflection of the fact that, stylistically, there is not a great deal of difference between these dialogues, as indeed one would expect if they were all by the same author and fairly close in date.

As far as the position of the *Parmenides* is concerned, when taken as a single dialogue the Mahanalobis distance usually reveals that it is closest of all to the *Theaetetus*, a position which *Prm.* 2 is often found to occupy in the CAN1 lists (e.g. B6, C9). This fits in well with supposed cross-references between the two dialogues (193 E) and a conjectured date of *c.*367 for the *Parmenides*, or at any rate close to the time of the *Theaetetus*.[33]

For the dialogues of this middle group I have adopted what approaches to the traditional order, thus:

[33] Guthrie, *History*, v. 32 f., 61.

Euthydemus
Symposium
Cratylus
Republic
Parmenides
Theaetetus
Epistle 13.

This is the same as Brandwood's sequence, except that he places *Cratylus*, the *Symposium*, and *Euthydemus* in group 2, a group which precedes the middle one of *Republic*, *Parmenides*, and *Theaetetus*. However, since he gives no order for the composition of this earlier group, there is no reason why *Cratylus*, *Symposium*, and *Euthydemus* should not be placed at the end of it, thus creating the same sequence as the one I have opted for. At the same time I should stress that differences between the dialogues of this group are relatively small and the various lists do not give a clear indication of sequence, other than that the *Theaetetus*, *Parmenides*, and *Republic* are at the end of the group and the other three dialogues are earlier. Strictly speaking, the *Republic* is more frequently found placed last in the group in sets B and C, pushing both the *Theaetetus* and *Parmenides* to an earlier date. Set A usually makes it earlier, often preceding *Cratylus*, while with set D it is succeeded in most cases only by the *Parmenides* (2) and *Epistle* 13. The set of 37 variables, which perhaps represents a large random selection of variables, and is therefore unbiased, gives the following sequence: *Prm.* 2, *Prm.* 1, *Protagoras*, *Symposium*, *Cratylus*, *Euthydemus*, *Theaetetus*, *Republic*, *Minos*, *Epistle* 13.

The presence of *Protagoras* and *Minos* in the group shows how difficult it is to draw precise conclusions from all the conflicting information, since various other dialogues appear from time to time in positions close to the *Republic* (*Hipparchus*: C5, 6, 8; *Amatores*: D5; *Minos*: A10). Without being precisely mathematical, however, I have tried to include in this group only those dialogues which appear to be placed there with reasonable consistency.

As to the sequence within the group, the *Theaetetus* is found to post-date the *Republic* in only 15 of the 36 lists, so that one should, strictly speaking, place the *Republic* at the latter end, followed only by *Epistle* 13, although even that is not confirmed by the evidence. The difficulty is essentially a historical one, for if we accept 369 as being a probable date for the *Theaetetus*, then either the *Republic* must pre-date it or its composition would probably have coincided with Plato's sojourn in Sicily on the second of his three visits. This is not impossible, especially as the theme of Plato's dealings with Sicily and Dion was the creation of the perfect society, the establishment of the philosopher-king, so one could conjecture that a large part of the stimulus for this work arose from Plato's Sicilian involvement and his attempts to reform society in Syracuse. Possibly the *Republic* was

commenced on Plato's return from Sicily and bears the fruit of his experience there. We could then construct the sequence as follows:

> *Euthydemus*
> *Symposium*
> *Cratylus*
> *Parmenides*
> 369–8 *Theaetetus*
> 367–6 Second visit to Sicily. Commencement of *Republic*
> 366   Return. *Republic* continued. *Epistle* 13.

Assuming that the *Republic* was completed by 365 it would still give a gap of about 10 years during which Plato wrote very little, apart from *Phaedrus*, which could perhaps be dated to 365. The evidence would support this interpretation, although it is not overwhelming. I confess to an inclination to place the *Parmenides* after the *Republic*, because it seems to introduce a new mood of questioning of the Theory of Forms which the *Republic*, if it were the later dialogue, does not appear to answer, whereas the rigorous dialogues of definition which are of the later period seem to be more in accord with the challenge posed by the *Parmenides* and attempt to set the Theory of Forms on more secure foundations.

Possibly the positioning of *Epistle* 13 could help to resolve this dilemma since, if genuine, it can be dated to *c*.365, very soon after Plato's return from Sicily. It is in itself a remarkable document, revealing to us a side of Plato which we encounter nowhere else in the dialogues. He sends, for Dionysius' children, 12 jars of sweet wine and two of honey, observing at the same time that they had returned to Athens too late to store the figs and that the myrtle berries which they had put away into storage had rotted. He also sends a statue which he had bought as a present for Dionysius' wife, who had looked after him 'in a manner worthy of both you and me' not only when he had been well but also in sickness (360 A–B). He sets out his financial commitments: the responsibilities he has towards the daughters of his nieces, for whom he will have to provide marriage portions, and the expenses he will be put to (no more than 10 minas) should his mother die (361 D–E). Then follows detailed advice as to how Dionysius should conduct his financial affairs at Athens (362 A ff.).

This wealth of personal detail, which is unlike anything found in any of the dialogues, or even in the other epistles, has led scholars to query the authenticity of the work. As far as stylometry is concerned, both the epistolary style and the rarity of the subject matter would tend to push the scores for this work to extreme values, so that it would not be entirely surprising to find that it did not submit easily to the discipline of the chronological investigation.

But it is quite remarkable how, in the event, the stylometric investigation

shows the proximity of *Epistle* 13 to both the *Republic* and the *Theaetetus*, for it is doubtful if anyone reading these three works would pick out much stylistic resemblance between them. If we look at the results for set C, we find the following sequences for the relevant dialogues (omitting extraneous dialogues which we have not included in this group):

  10. *Republic, Theaetetus, Epistle* 13
   9. *Republic, Parmenides* 2, *Theaetetus, Epistle* 13
   8. *Republic, Theaetetus, Epistle* 13
   7. *Theaetetus, Republic, Epistle* 13
   6. *Theaetetus, Epistle* 13, *Republic, Parmenides* 2
   5. *Theaetetus, Epistle* 13, *Republic, Euthydemus, Parmenides* 2
   4. *Theaetetus, Euthydemus, Parmenides* 2, *Epistle* 13, *Republic*
   3. *Epistle* 13,..., *Theaetetus, Euthydemus, Republic, Parmenides* 2

I have generally attached greater importance to the results from this set than to any other of the lists because it is based on a wider selection of dialogues, but in this case the wide separation of the two parts of the *Parmenides* must give us pause. Otherwise one would be inclined to take the precedence of the *Theaetetus* as decisively proved. Set B overcomes this defect, and frequently has the two parts of the *Parmenides* adjoining each other, but it places *Epistle* 13 before both the *Theaetetus* and the *Republic*, sometimes even before *Cratylus*. Set D also places *Epistle* 13 somewhat early.

Allowing for the great variability of all these works, their lack of resemblance to each other in all but the most superficial respects, and the strangeness or quirkiness of some of them, notably *Cratylus*, the *Parmenides*, and *Epistle* 13, which would tend to throw into disarray any attempt at classification, it is perhaps not surprising that these efforts to find the correct sequence of composition seem to be stumbling. However, looking at the more positive aspects of the results, we do see that the group itself is fairly well defined, although some occasional stray dialogues do seem to find their way into it (the *Apology, Protagoras, Hipparchus, Amatores, Minos*). Above all, the presence of *Epistle* 13 is confirmed and its proximity to the *Theaetetus* and the *Republic*.

Cratylus does present considerable problems, and its relative dating has always been the subject of dispute. Placing it in this group, for which the stylometric evidence is quite adequate, confirms the arguments of M. M. McKenzie[34] for a later dating of this dialogue. Its contents link it more with problems studied in the *Parmenides* and *Theaetetus* than those dealt with in the earlier aporetic dialogues.[35] Five lists in sets B and C (B3, 4, C3, 4, 5) do, however, imply an earlier dating for *Cratylus*, although, with fewer variables being utilized, the results tend to become unreliable, as is evinced by the greater separation of the two parts of the *Parmenides*. We are accepting,

---

[34] 'Putting the *Cratylus* in Its Place', p. 150.   [35] Ibid. 156.

therefore, a majority verdict in taking *Cratylus* into the middle group, a verdict which is confirmed by all the results of set A, set D being slightly ambiguous. Most of the lists place *Cratylus* securely with the middle group, which starts at about position 20, although in some instances it rises to position 14. The proportion in favour of a late dating is, however, at least as high as 5 to 1.

The absolute dating of this group cannot be divorced from the sequence and dating of the remaining 18 or so dialogues, since their proximity in terms of scores and the fact that some appear to be earlier, some later, than the traditionally accepted date, makes interpretation difficult.

In particular, the fact that there is no evident discontinuity anywhere in the sequence prior to *Phaedrus*, calls for some explanation, for the 30-year period 399–369 included not only the founding of the Academy, conjecturally in 387,[36] but also Plato's first visit to Sicily in 389–8. Did he continue writing throughout this period, or are we to assume that there was a lapse from creative labours while he devoted himself to teaching and elucidating the principles on which the Academy was to be run? The stylometric evidence does not give any indication of such a break, but on the contrary shows that many of the dialogues are very closely related, so much so that many of them are almost impossible to place in a securely indicated position, and this suggests that most of the dialogues prior to the *Republic* were written in an unbroken sequence.

Another acute difficulty arises from the attempt to reconcile the proposed dates of the *Gorgias* and *Symposium* with the position assigned to them by the stylometric evidence. For the *Gorgias* is shown as being comparatively early, separated from the *Symposium* by *Protagoras* and the entire *Phaedo* group, i.e. *Phaedo*, the *Apology, Crito*, as well as *Meno, Laches, Charmides*, and some shorter dialogues. Yet the proposed dates for the *Gorgias* and *Symposium* are 387 and 385 respectively, so that a large number of dialogues have to be crammed into a space of about two years, and the period 385 to 369 would be almost empty, with perhaps only the *Cratylus* and *Euthydemus* belonging to that period, even though no great hiccup is evident in the development of Plato's style over that time. Something must give way: either the stylometric evidence is at fault or one or other of these dates (for the *Gorgias* and *Symposium*) must be abandoned.

Since I am committed to an interpretation of the former, an interpretation which implicitly accepts its viability, I must follow where it leads and see if any adjustment in the conjectural dates is possible. In fact Dodds himself, who gives a detailed analysis of the evidence for dating the *Gorgias*, has already pointed to the difficulty of placing the *Symposium* so close to it.[37] He suggests that the date for the latter is at fault. I am inclined to accept his

---

[36] Dodds, *Plato, Gorgias*, pp. 24–5.   [37] Ibid. 25.

arguments, but more especially for the dating of the *Gorgias*, for which he assembles many detailed and cogent indicators, some of which, it is true, may not be entirely convincing, but taken cumulatively they present a very strong case for dating this dialogue to *c*.387, immediately after Plato's first visit to Sicily.

However, in accepting Dodds' choice of 387 for the *Gorgias* I am compelled to adopt a considerably more radical version of Plato's career than he did, as I must also accommodate the stylometric evidence, which places after the *Gorgias* so many dialogues which Dodds considered to pre-date it. My proposals, based on an interpretation of the CAN1 lists, are to limit the productions of the 390s to a very few, possibly only half a dozen, short dialogues. (Kahn makes a similar proposal,[38] but our choice of dialogues for the first period does not tally.)

It was the next 20 years from 388 onwards which witnessed the real bursting-forth of Plato's genius, a period which included the publication of the five great masterpieces, the *Gorgias*, *Protagoras*, *Phaedo*, the *Symposium*, and the *Republic*, as well as many shorter dialogues, perhaps of lesser merit but of compelling interest. The foundation of the Academy belongs also to this period, so we must view it not as an obstacle or hindrance to Plato's creativity but as the chief stimulus and motive of his thought and writing.

To fill in some of the details, the following are the dialogues which I suggest should belong to the 390s:

Lysis
Euthyphro
Minos
Hippias Minor
Ion
Hippias Major
Alcibiades 1
Theages
Crito

The degree of certainty of the evidence which allowed for the reasonably accurate chronology of the later dialogues does not apply here, for whatever dialogues are chosen for this group they cannot be placed there on the unanimous verdict of the CAN1 lists, for each one of them is found to occur at a later position (sometimes very much later) in one or more cases. Thus *Lysis* is shown to be early preponderantly by sets A, C, and D, by all 37 and ALET listings. In many cases it occupies first position (10 out of 36), thereby giving some colour to the anecdote told by Diogenes Laertius of Socrates listening to Plato's reading of this dialogue. Perhaps it was written before Socrates'

---

[38] 'Did Plato Write Socratic Dialogues?'.

death, but it is doubtful if the actual dates of composition of these early dialogues can ever be established with anything approaching certainty.

The fact that so few works were composed before the date of the *Gorgias* may possibly be taken as a partial confirmation of Plato's self-imposed exile in Megara immediately after the trial and execution of Socrates.[39] *Minos* and *Theages* belong to the list of dialogues of doubtful authenticity, while the others, excepting the *Hippias Major* and *Crito*, are comparatively slight, especially when set alongside the works which follow from the *Gorgias* onwards. Assuming, therefore, that the years from 399 to *c*.387 were fairly empty years, it also allows for the possibility that Plato during this period visited Egypt. In view of the mercenary details of *Epistle* 13, which we have found to be genuine, we could perhaps also accept Plutarch's statement that 'Plato paid for the expenses of his stay in Egypt by selling oil' (*Life of Solon*, sect. 2) especially as it is coupled with the claim that 'both Thales and Hippocrates the mathematician engaged in trade' and that trade carried with it no mark of inferiority (at least in the days of Solon). The house in which Plato was lodged in Egypt at Thebes was still standing in Strabo's day, five centuries later, and shown to enquiring tourists.[40] However, these matters are irrelevant to our main line of enquiry.

I should stress that the dialogues of this group were chosen using a visual assessment of the data rather than a strict mathematical calculation of their mean position. In some cases I have not followed my own rule of attaching more weight to sets B and C, especially in the case of *Minos*. If it really were a sort of introductory sketch for the *Laws*, as Shorey seemed to think,[41] it should belong to the later group, which it clearly does not, and I have therefore accepted the evidence of set D, C5, 4, 3, and B8, although it is probably not a genuine dialogue.

The authenticity of these works I have dealt with more fully in the previous chapter, but at this point I should mention that the apparent wild fluctuation of *Minos* and, to a lesser extent, *Theages*, is not incompatible with genuine-ness, since some of the assured dialogues are subject to this vice, for example *Charmides* and the *Apology*. On the other hand one cannot fail to notice that *Minos* is extreme in this respect and betrays very little stability compared with *Theages*, which remains around positions two and eight for set C and only occasionally strays lower than position 10 in some of the other sets, or *Alcibiades* 1 which hovers between positions five and 14 in all 36 of the CAN1 lists.

This must put a question mark against the authenticity of *Minos*, which does, however, suffer from the possible effects of genre, as one part of it is devoted to a long speech by Socrates extolling King Minos. This accounts for

---

[39] Guthrie, *History*, iv. 14.
[40] Str. xvii. 29. See also Cic. *De Rep.* i. x. 16, *De Fin.* v. xxix. 87. Guthrie, *History*, iv. 15.
[41] Ibid. v. 390.

approximately one third of the dialogue, a substantial proportion of the second part, perhaps 60%, and this no doubt causes a discrepancy between the two halves, as indeed is revealed by the individual scores for CAN1 on *Min.* 1 and *Min.* 2.

However, we cannot discount the possibility that an excessive fluctuation in the position of a dialogue as revealed by the CAN1 lists is in itself a sign of lack of authenticity, because it could indicate that the work in question does not obey the rules of chronological variation which are the cause of the shift in style of the other dialogues. In such circumstances it would be only random features which were being highlighted by the successive choice of variables, although a superficial resemblance to the early or late dialogues would tend to confine its allocation to one half or the other of the chronological sequence.

One would expect that an alien work would be located at one of the extremes of the range of scores, but this may be an unrealistic expectation because, apart from the fact that the variables have not been selected to emphasize authenticity but to determine chronology, there may be as many as seven or eight spurious dialogues in the corpus and they cannot all be placed at the extremes of the range. This is simply a physical impossibility and, in any case, if two spurious works were to be placed as neighbours in a Candisc analysis with mean values at the extreme end of the range, this would imply that they were closer in style to each other than they were to the works of Plato as individuals. But there is no presupposition that if two or more works are spurious they are necessarily by the same author, and in the probable case that different authors are responsible for the forgeries we would expect some of these works to be pushed into a closer acquaintance with the Platonic corpus as a result of their more compelling divergence from the other spurious works.

However, we are not in a position, with Plato, to state in advance which are the unauthentic dialogues, so that we can only guess what would be the effects of including a dozen or more spurious works in the corpus, and we cannot know what patterns would emerge if the entire corpus was guaranteed to be genuine. However, the presence of instability in what was intentionally devised as a chronological list or series of lists inevitably arouses suspicion and, beyond a few chance occurrences, should be regarded as a potential indicator, and possibly quite a strong one, of lack of authenticity.

It seems to be appropriate at this point to raise the question of the authenticity of *Epistle* 2. Although this has been partially treated in a previous chapter it has been left for the chronological investigation to have the final say. This epistle dates to either 364 or 360, but either of these two dates would place it later than *Epistle* 13 and close to *Phaedrus* or to *Philebus*. The CAN1 lists, however, show it to be in nearly all cases far too early,

sometimes in fact as almost the most extreme dialogue stylistically (C3, 4, 6) and often close to the upper limit of the range. It is never found to be later than *Epistle* 13, which it is supposed to post-date by at least two years, if not six, and only with D5, 4, and 3 does it appear to be placed anywhere near to the position which its putative date requires. Also it is the most erratic of all the works, ranging between positions two and 26, even though it is far less open to this sort of unexpected variation and abnormality than *Epistle* 13, which differs so radically in its subject matter from everything that is found in the dialogues. *Epistle* 2 does make some attempt to discuss philosophical matters (with lamentable ineptitude, according to many scholars), but it obviously fails to hit off the Platonic style, at least in respect of chronological exactitude. I have therefore not included it in the final list of works, especially as it would be impossible to place it in any position which would satisfy the internal chronological requirements and the chaotic stylometric indications.

To return to the early group, the *Apology* is noticeably absent from this list, a dialogue which is placed among the first by nearly all commentators.[42] Set B tends to place it earlier than all the other lists, in position 11 on average, and this would place it close to the date of the *Gorgias*. Most of the other lists make it somewhat later, sometimes considerably later than is conceivably possible (A3, D3). The arguments for an early date and for claiming it as the first of Plato's dialogues are not especially convincing. They do not seem to make sufficient allowance for the state of shock which Plato and his friends must have experienced on being robbed so cruelly of 'the man who, as we may say, was of all those whom we have known of his time, the best and the wisest and the most just' (*Phd.* 118 A). Could a work of such composure as the *Apology*, not to mention *Crito* and *Phaedo*, have been written so soon after Socrates' death?

Of course if the *Apology* were a mere verbatim report, or one slightly embellished, of Socrates' speech at his trial, then it would not require any large gap of time before Plato produced it. But one might well ask why anyone should need a written report of a speech which, had they wished to hear it, they might have enquired of a friend or themselves gone to the trial to hear.[43]

A close account of the trial would have been superfluous in the years immediately following it, whereas it would fit more appropriately a later date if it deviated from a true report, when the memory of the events had somewhat faded. The language seems to be too closely akin to the typical

---

[42] Ibid. iv. 71.

[43] The jury consisted of 501 citizens. Others would have been present at the trial out of curiosity or concern. The jury alone would represent about 2% of the citizenship body of Athens, a substantial proportion of the population and certainly enough to guarantee widespread dissemination of the proceedings. It is important to remember that communication was predominantly oral for Greek society at the time, and the possession of books was still a comparative rarity.

Platonic style (I refer to its stylometric proximity) to allow of its being purely Socratic. In the intervening period between the trial and the writing of the *Apology* time would have assuaged some of the bitterness and helped Plato to achieve a more detached and balanced view of the events. Perhaps it was written in response to a renewed outburst of anti-Socratic literature, such as Polycrates' attack.[44] At any rate, a later date for the *Apology* is possible on historical grounds, and the stylometric evidence indicates a period later than that of the first group, although it is true that this evidence is not unequivocal. There is nothing, however, which suggests that it was the first dialogue.

Crito is more difficult to place since set C shows it to be early, prior to the *Gorgias*, but in set B five out of eight lists give it a later position. Set A is evenly divided and set D puts it later, closer to *Protagoras*. I have placed it tentatively at the end of this early group, accepting the evidence of set C, and take it to be a dialogue composed shortly before Plato's first visit to Sicily. A later date for it is certainly not impossible.

There is another point which ought to be considered in the dating of the *Gorgias* and these early dialogues, and that is the comparatively late age at which Plato claims that his disillusionment with political life became fixed. The passage in the seventh epistle, 326 A–B, on which Dodds bases the dating of the *Gorgias*, states that Plato had reached this viewpoint when he first arrived in Sicily in 389–8: 'ταύτην δὴ τὴν διάνοιαν ἔχων εἰς Ἰταλίαν τε καὶ Σικελίαν ἦλθον, ὅτε πρῶτον ἀφικόμην' ('In this frame of mind I came to Italy and Sicily on my first visit there', 326 B).

Before this he was still looking for ways of improving the body politic and awaiting an opportune moment to do something practical, 'τοῦ δὲ πράττειν αὖ περιμένειν ἀεὶ καιρούς', which seems to imply that for part of the 390s at least, he would have been involved in some sort of political activity, something to which his position as a member of an aristocratic Athenian family would have naturally predisposed him. Consequently, less time would have been available for writing, and throughout this period literary enterprise would have been little more to him than a πάρεργον, a work of trivial importance.[45]

We are therefore left with the following conjectural sequence of dialogues, with possible dates of composition:

[44] Guthrie, *History*, iv. 72. Polycrates' attack is thought to date from between 394 and 390. The complexities of the literary and political scene which might have prompted the writing of *Ap.* are well described by Chroust, *Socrates, Man and Myth* (London, 1957). There is a tradition that Socrates remained silent throughout his trial. Stone also stresses the political aspects of Plato's defence of Socrates (*The Trial of Socrates*).

[45] I realize that Plato dismisses all literary activity, including his own, as being of little account. For an interpretation of the passages in *Phdr.* and elsewhere see Guthrie, *History*, iv. 56 ff. The point I am making here is that writing at this stage was secondary to the possibility of political activity, although both, in the absolute scale of human and divine achievement, might be of little worth.

400	*Lysis*
399	Death of Socrates
	Euthyphro
	(*Minos*)
	Hippias Minor
395	*Ion*
	Hippias Major
	Alcibiades 1
	Theages
390	*Crito*
389–8	Voyage to Sicily.

Early middle group

The remaining dialogues which have not already been dealt with fit into this group. I give the approximate sequence which on balance the evidence suggests, although I emphasize once again that this sequence cannot be adhered to rigidly as having been proved beyond all doubt. It represents the aggregate of all the evidence which the stylometric analyses offer.

387	*Gorgias*
	Menexenus
	Meno
	Charmides
	Apology
	Phaedo
	Laches
	(*Hipparchus*)
	(*Amatores*)
380	*Protagoras*

Protagoras is placed close to the *Symposium* by sets B, C, and D, although A puts it somewhat earlier. It is only once (B3) found preceding the *Gorgias*, even though many scholars have attached great importance to its precedence over the *Gorgias*. They prefer to see the apparently constructive doctrine of the *Gorgias* as later, following on from the purely negative rebuttals which are found in *Protagoras*.

The status of the two minor dialogues, *Hipparchus* and *Amatores*, I have left unresolved. One would prefer to see them placed much earlier, although it is perhaps illogical to require that a dialogue be early simply because it is short. These two dialogues do have affinities of subject matter and style with members of this and the next group, and the slightness of their content should not count against them. However, *Hipparchus* arouses suspicion because of fluctuations of position bordering on the extreme, and because it is also found placed sometimes in the no man's land between middle and late

dialogues, an area of comparative emptiness which could well serve to catch works which do not belong to the corpus.

Summary

The full list of dialogues now runs as follows:

400?	*Lysis*
399	Death of Socrates
	Euthyphro
	(*Minos*)
	Hippias Minor
395?	*Ion*
	Hippias Major
	Alcibiades 1
	Theages
	Crito
389	First visit to Sicily
387?	Founding of Academy
386	*Gorgias*
	Menexenus
	Meno
	Charmides
	Apology
	Phaedo
	Laches
	(*Hipparchus*)
	(*Amatores*)
380?	*Protagoras*
	Euthydemus
	Symposium
	Cratylus
	Republic
	Parmenides
369	*Theaetetus*
367–6	Second visit to Sicily
366	*Epistle* 13
365?	*Phaedrus*
361–0	Third visit to Sicily
355?	*Philebus*
	Clitophon
354	Murder of Dion
353–2	*Epistle* 7

	Epistle 3
	Epistle 8
	Sophist
	Politicus
350?	*Laws*
	Epinomis
	Timaeus
	Critias
348–7	Death of Plato.

Conclusions

To discuss in detail the merits of placing each of these dialogues in the position indicated would be a vast enterprise. I am conscious of having dealt inadequately with many of the problems and implications of the revision of traditional dating which these analyses seem to require. In particular, the decision to retain the old sequence, *Republic, Parmenides, Theaetetus*, when the evidence really requires a reappraisal of the situation, springs, perhaps, more from a desire to shun the task of defending such a late date for the *Republic* and the implications of linking it with Plato's second visit to Sicily, than from any desire to favour tradition at the expense of an unbiased appraisal of the results.

It may be that others will be able to arrive at different conclusions from those presented here, or to present other stylometric evidence. But the important point is that style does not admit of a simple description, still less a simple quantification, and Lutoslawki's dictum that works which are close in style must also be close in date[46] invites the immediate comment that it depends entirely on how we define style. Stylometry defines style by its choice of variables and, if different variable sets supply different answers to the problem of distance or proximity of various works, this does imply that we cannot obtain a single unique answer to the chronological question. It is a conundrum which has many answers and we may only hope that by judicious use of the evidence we shall find a solution which is close to the truth.

The thought also occurs to me that, despite all the hazards of working with an author of whom little is known, Plato is perhaps kind to the analyst in that the sharp difference between early and late works provides an immediate source of information on which stylometry may be set to work. With other authors the change may be imperceptible or masked by other forms of variance, and conclusions may be impossible.

However, this is a question which may only be settled by further research,

---

[46] *The Origin and Growth of Plato's Logic*, p. 152.

preferably on authors for whom the facts are already known, so that the reliability of the methods for measuring chronology may be tested thoroughly, 'lest one good custom should corrupt the world'.

No doubt scholars will disagree with my conclusions, or some of them, a disagreement which I would welcome, since the cause of lively enquiry cannot be furthered by a mere nodding assent. I hope, however, to have shown that stylometry can contribute considerably to our knowledge and understanding of the interrelationships of style, and that we can no longer revert to the *ad hoc* judgements which dismiss one work as being entirely unlike another (and therefore spurious). Plato's work may be quantified in a stylometric sense and a measure of resemblance is therefore possible, as it is for all other authors, and perhaps the day is not too distant when we shall have far more extensive information than at present on the range of variability to be found in individual authors and the extent to which genre and date and shifting linguistic usage all play their part in shaping and moulding these stylometric patterns.

The chronological sequence, as I have presented it, is based on sound evidence derived from a far more comprehensive survey of the entire corpus of Platonic dialogues than has ever been attempted before. It has revealed much about these works which was hitherto unknown or only dimly suspected. For who could have guessed that of all the works of Plato those that are consistently closest to the *Laws* on our scales of distance measurements are *Epistle* 7 and *Epinomis*, two works which have been dismissed so frequently as spurious, and that the chronological listing would place them in the position that our knowledge of them requires? It is facts such as this which must surely give credence to the method and its results, and if I have erred in interpreting them, or based the study on false premisses, then I invite others to take up this exciting work and present alternatives. For the doorway stands open.

10
Explanations

IT is perhaps rather surprising to have reached this stage in the enquiry without having given any explanations or suggestions as to why the methods used in MVA work with linguistic problems of authenticity and chronology. Although at the outset it was considered to be more important to produce results rather than to speculate why, if at all, they should occur, yet to gain general credence for the methods employed in the previous chapters and for what I take to be their fairly high level of success it is necessary to be able to offer some words of explanation. Even if these touch only lightly on the common ground of linguistic usage and at a relatively simple level, it is better to face up to the problem rather than pass it by in silence.

The problem is, however, one of some complexity, because MVA makes use of the data in a way which removes us some distance from the initial, raw, undistorted measurements. Also, each case studied is unique (although the methods used may be universal), for it depends on the samples selected for that occasion and on the question for which an answer is sought. Thus the explanation of why author A differs from author B will not necessarily apply to the differentiation of author A from author C. Similarly, the chronological separation of works by a single author into the correct sequence will have yet another explanation and this will vary from author to author depending on their stylistic idiosyncrasies.

In order to simplify the task of analysis I have chosen to illustrate, with reference to the BLETs only, the sort of approach which may be adopted, by outlining the chronological change in linguistic usage which may be observed in Plato in respect of two comparable dialogues, one early and one late. Instead of 37 variables there will be only nine to deal with, and of these only one or two that are the most important will be looked at in detail. This is essential, because any word list taken even from a short dialogue is likely to be voluminous, and to hope that from lists of several hundred words culled from each dialogue an obvious pattern will emerge which highlights the difference between early and late works is to be totally unrealistic. In fact I shall be deliberately restrictive in this respect and, apart from general observations concerning the overall concordance statistics, I shall look only at the one variable which is likely to tell us most about Plato's stylistic habits and how they alter as he ages.

Fortunately it is not necessary to select this variable entirely at random, as there are various pointers which indicate the ones which are most appropri-

ate. The most fruitful source of information is that obtained by the Candisc run on all the Platonic dialogues, but especially that which uses the BLETS only. Since the CAN1 variable obtained from such an analysis is, by definition, the one variable (obtained by combining all the BLETS) which maximally succeeds in allocating individual samples to the parent work,[1] it may be described also as the one which separates the dialogues from each other most effectively. Conversely, it also shows the dialogues which appear to be most similar, and this was the basis for reliance upon the lists of mean CAN1 values as indicators of chronological sequence, as shown in Chapter 9.

However, the CAN1 variable is a composite variable, made up by combining all the original BLETs when the appropriate coefficient has been applied to them. Thus it becomes difficult to see what relationship the CAN1 variable bears to these original BLETs. Fortunately the computer supplies help in the form of a listing of the correlations of each BLET with CAN1. We are especially interested in the so-called 'between-canonical structure', which shows the relevant part of the variance of CAN1 which accounts for the 'between-group variance' of the samples, that aspect of their variance which is accounted for by the fact that they belong to different dialogues. These listings give a statement of the correlation of each BLET with the CAN1 variable as it relates to a separation of the samples into groups. It is then a simple matter to discover the few BLETs which have the highest correlation with CAN1. These are the variables which, if necessity demanded, could most effectively be used as a substitute for CAN1, for they would still convey most of the information which CAN1 contains (assuming of course that the correlations were high in the first place).

Thus we find that the variables with the highest correlation are BLET4, BLET5, and BLET9, the value of the correlation coefficients being 0.9598, -0.9515, and 0.8495 respectively.[2] These are the variables which measure the percentage of words ending in ι, ν, and ω. Implicit in these figures is the fact that any one of these variables (but especially the first two) could individually tell us a lot about the way Plato's style changes chronologically, because correlations as close as 0.9 are high by any standard (the maximum is 1.0) and because the CAN1 variable has been shown, in Chapter 9, to be a good indicator of sequence of composition.

There is further confirmation that BLET4 and BLET5 are closely related to stylometric change in Table 9.3, which shows the results of Stepdisc analyses on various groups of dialogues chosen explicitly because of their temporal separation. In this table the two variables BLET4 and BLET5 are shown at the

[1] Candisc maximizes the correlation of the canonical variable with group membership. Generally speaking, the higher the value of multiple R, the greater will be the number of samples which have been correctly allocated.

[2] These are the figures for a Candisc run on all the Platonic dialogues excluding *Alc.* 2. The negative correlation for BLET5 does not invalidate its usefulness. One could change all negative values for positive ones and vice versa and the classification results would be identical.

top in sets B and C; BLET4 is no. 5 in set A and BLET5 is no. 1 in set D. It is clear, therefore, that these two variables are highly significant in matters which relate to the chronological stylistic changes of Plato's dialogues.

Of the two I shall select BLET4 (words ending in ι) for a more detailed analysis of the words which are being measured and the way they relate to Plato's style. It is preferable to BLET5 because its correlation with CAN1 is slightly higher and its mean value (see Table 9.1) is lower. When dealing with word lists this can be advantageous as there are fewer words to consider.

Next we must choose which of the dialogues to look at, since it is clearly impractical to examine word lists of BLET4 for the entire corpus of 750 000 words. As *Critias* has been shown in Chapter 9 to be the last of the dialogues it seems appropriate to use it as a paradigm of the later works and to take an earlier work of comparable length to balance against it. I have chosen *Euthyphro*, which is one of the earliest works and contains 5171 words compared with *Critias'* 4956. To make the two samples exactly similar I have only examined the first 4950 words of each dialogue.

What then do the statistics reveal about these two dialogues? Firstly, the concordance statistics for all words, given in Tables 10.1 and 10.2, show quite clearly that there is a marked increase in the use of *hapax legomena*[3] (words used once only in the dialogue) in the later work and this is accompanied by a decrease in the level of use of more frequently occurring words. These two statistics obviously complement each other as one cannot increase the use of commonly occurring words without a consequent decrease in those which are less frequent, and vice versa. However, it is worth looking at the concordance statistics in greater detail as there is much to be gleaned from them. I should stress that these are the statistics for all words, not only for those ending in ι, and therefore we shall be looking at them in general way. Later the more detailed lists of BLET4 will be given.

It appears that *Critias* has a *hapax legomena* score of 1535 words or 31.01% of the total. This compares with 922 for *Euthyphro*, or 18.63%, quite a considerable range of variation to be found in one author. The difference of linguistic usage between these two dialogues is also shown by the type : token ratio for each—0.417 for *Critias*, compared with 0.295 for *Euthyphro*.[4] The

[3] These are not necessarily *hapax legomena* in relation to the entire corpus, for which much lower figures would be expected. All the words referred to in the analyses of this chapter are unlemmatized—even a different accentuation produces a separate listing for a word—so that inflated figures result, as is mentioned later in the chapter.

[4] Comparable figures for other authors are difficult to obtain. Shakespeare achieved 19.62% for the *Tempest* and 20.11% for *Macbeth*. (See J. C. Baker, 'Pace: A Test of Authorship Based on the Rate at which New Words Enter an Author's Text', *ALLC Journal*, 3(1) (1988), 38. I find Baker's article both confused and confusing. Under the guise of introducing a new statistic, which he calls an author's 'pace', he then gives us the figures for the type : token ratio under the heading of 'vocabulary ranking' in table 1. It is clear that these rankings are not independent of text length. All Marlowe's shorter works (under 10 000 words) have a higher score than his longer ones, and the same is, broadly speaking, true for Shakespeare.) These two works, the *Tempest* and *Macbeth*, are both about four times as long as *Criti.* and *Euthphr.* and one would

TABLE 10.1   Euthyphro concordance statistics

Frequency	No. such	Words in frequency	Vocabulary total	Word total	% of vocabulary	% of words	% of words in frequency
1	922	922	922	922	63.15	18.63	18.63
2	166	332	1088	1254	74.52	25.33	6.71
3	121	363	1209	1617	82.81	32.67	7.33
4	64	256	1273	1873	87.19	37.84	5.17
5	36	180	1309	2053	89.66	41.47	3.64
6	24	144	1333	2197	91.30	44.38	2.91
7	19	133	1352	2330	92.60	47.07	2.69
8	14	112	1366	2442	93.56	49.33	2.26
9	6	54	1372	2496	93.97	50.42	1.09
10	7	70	1379	2566	94.45	51.84	1.41
11	5	55	1384	2621	94.79	52.95	1.11
12	5	60	1389	2681	95.14	54.16	1.21
13	10	130	1399	2811	95.82	56.79	2.63
14	2	28	1401	2839	95.96	57.35	0.57
15	5	75	1406	2914	96.30	58.87	1.52
16	5	80	1411	2994	96.64	60.48	1.62
17	1	17	1412	3011	96.71	60.83	0.34
18	1	18	1413	3029	96.78	61.19	0.36
20	3	60	1416	3089	96.99	62.40	1.21
21	5	105	1421	3194	97.33	64.53	2.12
22	1	22	1422	3216	97.40	64.97	0.44
23	1	23	1423	3239	97.47	65.43	0.46
25	3	75	1426	3314	97.67	66.95	1.52
26	1	26	1427	3340	97.74	67.47	0.53
27	2	54	1429	3394	97.88	68.57	1.09
28	2	56	1431	3450	98.01	69.70	1.13
29	3	87	1434	3537	98.22	71.45	1.76
30	1	30	1435	3567	98.29	72.06	0.61
31	2	62	1437	3629	98.42	73.31	1.25
33	3	99	1440	3728	98.63	75.31	2.00
34	1	34	1441	3762	98.70	76.00	0.69
35	1	35	1442	3797	98.77	76.71	0.71
36	1	36	1443	3833	98.84	77.43	0.73
37	1	37	1444	3870	98.90	78.18	0.75
38	1	38	1445	3908	98.97	78.95	0.77
40	3	120	1448	4028	99.18	81.37	2.42
42	1	42	1449	4070	99.25	82.22	0.85
48	1	48	1450	4118	99.31	83.19	0.97
49	1	49	1451	4167	99.38	84.18	0.99
51	2	102	1453	4269	99.52	86.24	2.06
52	1	52	1454	4321	99.59	87.29	1.05
55	1	55	1455	4376	99.66	88.40	1.11
60	1	60	1456	4436	99.73	89.62	1.21
68	1	68	1457	4504	99.79	90.99	1.37
88	1	88	1458	4592	99.86	92.77	1.78
99	1	99	1459	4691	99.93	94.77	2.00
259	1	259	1460	4950	100.00	100.00	5.23

Type: token ratio = 0.295; word total = 4950.

TABLE 10.2 Critias concordance statistics

Frequency	No. such	Words in frequency	Vocabulary total	Word total	% of vocabulary	% of words	% of words in frequency
1	1535	1535	1535	1535	74.37	31.01	31.01
2	250	500	1785	2035	86.48	41.11	10.10
3	97	291	1882	2326	91.18	46.99	5.88
4	53	212	1935	2538	93.75	51.27	4.28
5	27	135	1962	2673	95.06	54.00	2.73
6	23	138	1985	2811	96.17	56.79	2.79
7	17	119	2002	2930	97.00	59.19	2.40
8	10	80	2012	3010	97.48	60.81	1.62
9	4	36	2016	3046	97.67	61.54	0.73
10	3	30	2019	3076	97.82	62.14	0.61
11	1	11	2020	3087	97.87	62.36	0.22
12	3	36	2023	3123	98.01	63.09	0.73
13	1	13	2024	3136	98.06	63.35	0.26
14	1	14	2025	3150	98.11	63.64	0.28
15	3	45	2028	3195	98.26	64.55	0.91
16	3	48	2031	3243	98.40	65.52	0.97
17	1	17	2032	3260	98.45	65.86	0.34
18	1	18	2033	3278	98.50	66.22	0.36
19	1	19	2034	3297	98.55	66.61	0.38
20	2	40	2036	3337	98.64	67.41	0.81
21	1	21	2037	3358	98.69	67.84	0.42
24	1	24	2038	3382	98.74	68.32	0.48
25	2	50	2040	3432	98.84	69.33	1.01
26	1	26	2041	3458	98.89	69.86	0.53
27	1	27	2042	3485	98.93	70.40	0.55
28	1	28	2043	3513	98.98	70.97	0.57
30	1	30	2044	3543	99.03	71.58	0.61
33	3	99	2047	3642	99.18	73.58	2.00
35	2	70	2049	3712	99.27	74.99	1.41
37	1	37	2050	3749	99.32	75.74	0.75
45	1	45	2051	3794	99.37	76.65	0.91
46	2	92	2053	3886	99.47	78.50	1.86
48	1	48	2054	3934	99.52	79.47	0.97
49	1	49	2055	3983	99.56	80.46	0.99
56	1	56	2056	4039	99.61	81.60	1.13
63	1	63	2057	4102	99.66	82.87	1.27
81	2	162	2059	4264	99.76	86.14	3.27
84	1	84	2060	4348	99.81	87.84	1.70
89	1	89	2061	4437	99.85	89.64	1.80
98	1	98	2062	4535	99.90	91.62	1.98
160	1	160	2063	4695	99.95	94.85	3.23
255	1	255	2064	4950	100.0	100.00	5.15

Type: token ratio = 0.417; word total = 4950.

type : token ratio is a measure of vocabulary richness and records the ratio of actual words used (types) to the total number of words in the sample (tokens). If the same word occurs a number of times in a text it is only counted once as a type but for the full number of occurrences as a token. Hence the type : token ratio measures rate of repetition rather than the strict range of total vocabulary. Vocabulary usage is in any case extraordinarily difficult to define because it presupposes a consensus of opinion about how we understand words and relate them to their dictionary forms. Provided consistency were maintained it would be possible to obtain some idea of the absolute vocabulary range of an author according to strict etymological and semantic criteria, but no computer could cope unaided with such a task. It would require a vast amount of preliminary planning and programming, for systems which parse, lemmatize, and then count the text presented to them are only just beginning to emerge.

The figures given in Tables 10.1 and 10.2 relate, therefore, to the orthographic differences of words. Any word which differs orthographically from another is considered to be a different 'type', although a dictionary would obviously in many instances not agree with such a judgement. What this results in is clearly an inflated figure for vocabulary richness and one would not seriously expect any author to maintain over an extended piece of writing a level of vocabulary usage as high as the 0.417 figure given for *Critias*, since so many function and link words must inevitably intrude and the longer the work the more difficult it is to avoid repeating words already used.

Nevertheless, despite the artificially high value of vocabulary usage implied by the type : token ratio, it has the great advantage of being consistent and also that of being readily obtainable. We cannot ignore the fact that Plato's stylistic habits have changed significantly according to this measure of vocabulary richness, for it is by no means easy to achieve such a shift in the use of language as to increase the value of this measure by 40%.[5]

To take the preceding paragraph as an example: it contains 78 words (tokens) and 56 types, giving a type : token ratio of 0.73. This appears to be abnormally high, but it is only due to the shortness of the sample. The figure would doubtless fall to below 0.2 if the entire chapter up to this point were considered, for the number of *hapax legomena* (taken in relation to the chapter so far) is only about seven, giving a rate of about 9% after approximately 1400 words. This compares unfavourably with Plato's 18.6%

expect lower scores from them. Consequently, *Euthphr.* is not high on the scale of vocabulary richness. But I suspect we cannot have reliable inter-author comparisons of vocabulary usage until we have figures for 1000-word samples from many authors. It is only by comparing like with like that any sense can be made of such statistics.

[5] $(0.417 - 0.295) \div 0.295 = 41.36\%$. I take the figure for *Euthphr.* as the baseline, as this dialogue is earlier.

for *Euthyphro* and 31% for *Critias*, even more so when one considers that they are at least three times as long (as my 1400 words) and the probability of repeating any word increases considerably as a work is extended. Plato, we may conjecture, deploys a vocabulary which only the most gifted of authors can command and the mere humdrum levels which the average writer might attain fall well short of even his nodding moments.

Which leads me to the conclusion that the type : token ratio would have been an excellent statistic to have obtained for all the 702 samples, since it would have provided evidence of the changing patterns of vocabulary usage for all authors and would have been a useful source of comparison between them. However, this was not done, and it is now too late to submit them once more to the computer. The readings we have for *Critias* and *Euthyphro* must suffice for the explanations which are to be offered in this chapter.

I shall make only one further mention of the statistics given in Tables 10.1 and 10.2, and that is to point out that nearly half of *Euthyphro* (49.5%) consists of words occurring 10 times or more in the dialogue, whereas the comparable figure for *Critias* has dropped to 38.5%.[6] It appears that the younger Plato relied much more heavily on common words, and the relative absence of these in the later dialogues perhaps accounts for their terse and compressed style.

To turn now to the more detailed evidence of the BLET4 word lists. I do not give these in full (space does not permit) but only a list of those words which occur eight or more times in either dialogue. This is given as Table 10.3, which also shows the difference in score between the dialogues. Homonyms are not separated and I have listed some words without accentuation for that reason. This does conceal some facts, as for example that τί as an interrogative is used far more in *Euthyphro* than *Critias*, as befits no doubt the questioning nature of the earlier dialogue.

From Table 10.3 it emerges that only four common words are used more frequently in *Critias*, these being ἀεί (always), ἐπί (on), ἔτι (still), and περί (around, about). The total excess of these four words is 40, so that the difference between the two dialogues in total, 696 less 415 words, is in fact greater than the raw score of 281 seems to suggest, as *Euthyphro* has to make up for these additional 40 words, which count as a negative score from the point of view of the earlier dialogue. It would be claimed that the three words θεοί (gods), φιλεῖσθαι (to be loved), and φιλεῖται (it is loved) are thematic and should be excluded from the count, since the subject matter (piety)[7] of *Euthyphro* demands their use. There is some truth in this, but *Critias* also has its peculiar claims for thematic bias and should be allowed an equal licence.

[6] Words occurring nine times or less account for 50.5%. Therefore those occurring 10 times or more must make up the remaining 49.5%.

[7] Euthyphro opines that whatever is loved by the gods is holy, what is hated is unholy. Hence there is much play on the three words quoted.

TABLE 10.3   *Words occurring more than eight times in*
Euthyphro *and* Critias

Word	*Euthphr.*	*Criti.*	Difference
ἀεί	1	8	−7*
διότι	12	—	12
δοκεῖ	12	—	12
εἰ	41	7	34
εἶναι	28	13	15
ἐπι	9	25	−16
ἐστί	17	1	16
ἔτι	2	14	−12
ἔχει	8	2	6
θεοί	15	1	14
καί	261	256	5
μοι	26	1	25
οἱ	32	6	26
οἶμαι	12	—	12
ὅτι	61	15	46
περι	46	51	−5
σοί	27	4	23
τι	55	10	45
τυγχάνει	8	1	7
φιλεῖσθαι	8	—	8
φιλεῖται	15	—	15
TOTAL	696	415	281

* A negative value indicates that *Criti.* exceeds *Euthphr.* on the score for that word.

But perhaps the relevant words for *Critias* are not covered by BLET4, and to be fair one should reduce the total count for *Euthyphro* by 37 (14 + 8 + 15) to 659. One is then left with a surplus of usage of common words by *Euthyphro* of 244. The principal difference, the excess shown in brackets, occurs with the words εἰ (49), ἐστί (16), μοι (25), οἱ (26), ὅτι (46), σοί (23), and τι (45), giving a total of 230. μοι and σοί are rather unexpected but no doubt relate to the peculiar nature of dialogue, with phrases such as δοκεῖ γάρ μοι (it seems to me) frequently recurring. Offsetting this we may claim the excessive usage of περι and ἐπι in *Critias*, words which have many applications when physical, natural, and geographical descriptions are required.

To summarize, it is evident that Plato's use of language has changed considerably between these two dialogues. In the latter it is richer and more varied, in the former more humdrum and repetitive. The picture I have presented is not complete, as I have only looked in detail at words ending in ι

(BLET4) and at the overall concordance statistics. From these there is ample evidence to show that *Euthyphro* falls well behind *Critias* in its vocabulary range and relies more heavily on common words, the mortar of language. I have little doubt that similar results could be obtained for other pairs of early and late dialogues.

More complex linguistic changes relating perhaps to classes of words, rather than the common, well-known ones mentioned in Table 10.3 and elsewhere, such as whether or not Plato became more addicted to oblique cases, or the use of participles, or infinitive constructions, and so on, I have not been able to discern. The compass of this book and the time available do not permit a more wide-ranging study. It is a task which lies in the future.

Briefly it may also be mentioned that similar facts emerge from authorship discrimination, or at least that the approach for discerning the differences is identical. For using all 702 samples and classifying them according to the traditional author in each case, a Candisc run reveals that the variables which tell us most about the differences between the seven authors are CLET9 (words with ω penultimate), ALET10 (words containing ν), BLET2 (words ending in ε) and ALET19 (words containing ω). The correlation figures for these variables with CAN1 are all well above 0.95.

We can also see that in differentiating between these seven authors Isocrates is most sharply isolated, having a mean CAN1 score of 6.5 while the other six authors score in the range 3.4 to −0.86. Plato and Xenophon are the two most distant from Isocrates with −0.847 and −0.865 respectively, so that it is clear we could not use this variable to separate these two authors, but only to isolate Isocrates from either of them. Hence the four primary variables quoted above (CLET9 etc.) must have similar values for Plato and Xenophon. This must be true also for Aeschines and Lysias, who have almost indistinguishable CAN1 mean scores: 2.174 and 2.099. To find out what contributes most to a differentiation between any of these, one would have to look at the next few canonical variables in the sequence. One of these would show Plato and Xenophon, or Aeschines and Lysias being at opposite ends, or nearly so, of the mean canonical scores, and investigation of the correlation matrix would then indicate which of the primary variables it would be appropriate to highlight. Or one could separate the two similar authors and run an analysis on them alone, so that no extraneous information would obscure the causes of the discrimination made possible between the two of them.

Complex though this may seem, it is not an especially difficult process to manage and only sounds awkward here because of the necessary abstraction caused by generalizing and the attempt to limit the amount of new material brought into the study at this late stage. When dealing with specific tasks and with the computer printout in front of one, the choices often become very clear-cut.

Suffice it to say, therefore, that there is plenty of material to work on, material directly related to linguistic usage, which will help us to isolate ways in which authors are idiosyncratic and to pinpoint the features which vary with time; for the variables themselves, the ALETS, BLETS, and CLETS, are clearly quite adequate for the task which has been assigned to them—that of differentiating between authors and styles. No doubt some could be taken away and others added, resulting in an improved performance, but generally speaking they have shown that the assumption that style could be defined implicitly by use of them was not misguided. Whatever their distribution, whether multivariate normal, or of some other shape, it is clear that they responded well to the different types of analysis employed.

Obviously much still needs to be done in explaining the mathematics of multivariate stylometry employing orthographic variables, and in discovering why it should work at all. But to have demonstrated how well it can work is surely a most important step, without which progress in any direction would scarcely have been possible.

11
Conclusion

NOT a great deal more needs to be said. Underlying the whole enquiry has been perhaps the unspoken question as to whether or not stylometry could truly be regarded as a science. Although I embarked on the project expecting a negative response, or at best an ambiguous silence, the question has, it seems, been answered reasonably positively. As a science it is deficient still in its theoretical content, and must take its place as part of the greater discipline of linguistics, but it has, nevertheless, provided reams of quantitative information about the works studied which probably could not have been obtained by any other means, and it has shown how style may be measured by employing a system of orthographic variables.

Perhaps what is most surprising in all that has been shown regarding this science is that nothing of it is really new. The statistical methods used, Cluster, Discrim, and Candisc, are all well known to workers in various fields of enquiry and the theory for them all has been developed over many years. Letter distribution has also been studied in the past, but multivariate analysis has never been used consistently in stylometry, taking a large number of variables and using a wide range of authors simultaneously. The realization that the methods of MVA were applicable in a literary context, that the problems of classification were akin to those encountered in various other disciplines, and that a set of variables was available which would provide the raw material for MVA to work on—this realization has provided the groundwork for the development of stylometry from an *ad hoc* system of investigation, one which develops new techniques for each pair of works or authors studied, into a science with a generalized and universal approach to the problem of authenticity, chronology, and the measurement of style.

For I maintain that the methods employed and the types of variable used would be appropriate in any language, and with any author, style, or genre. Of course I cannot prove that similar results would be obtained with Latin, French, English, Chinese, or whatever, but there is more than an even chance that it will be so. However disturbing it may be to find that the mathematician can, to a limited extent, uncover patterns behind the works of genius and chart their progress, so that what before was admired for its freedom, grandeur, and immensity is now in a sense 'cabin'd, cribb'd, confin'd, bound in', yet it would be even more disturbing to discover the complete absence of any pattern or control. We know that languages have laws and rules, and that many of these are intuitive and half-conscious. Also the brain

is finite and its scope for vocabulary usage has limits. Weariness often brings repetition and habits impose a structure which the conscious mind might hardly be aware of. Environment, contacts, and the customary milieu of the author must also play their part in shaping the language which finally appears on the printed page.

These considerations are important when the claim of stylometry is put forward that differences of style may be quantified by changes in value of the minutiae of language. For this claim relies on the underlying regularity of language and not on its random, strange, or unpredictable qualities.

Some modifications of the variables used may be required for other languages, but the basic principles should be the same, with orthography supplying the basis of measurement and analysis proceeding with the various MVA techniques described in the preceding chapters.

It is difficult to reproduce in a written work the sum total of results which gradually accumulated from scores of analyses and hundreds of pages of computer printout, and also to convey the impressions and slowly strengthening convictions produced by these results. For it began to appear that some important new facts were here emerging about the basic structure of language and the ability of statistics to cope with the seemingly intractable vastness and multitudinous variability of the written word. It was almost as if thought itself was in the process of being tamed, for, however far in front imagination leapt, mathematics and statistics, like prayers, the daughters of Zeus, their faces wrinkled, their eyes askance, and with stumbling gait, follow on after and take the measure of all things.[1] I began to believe that, wherever human judgement and computer analysis did not coincide, it was the former which was fallible. We have not looked closely enough and with sufficient cold clarity at the ways in which style is assessed. Perhaps, when we find that a sample of Plato is classed with the Xenophon population, we should accept that this is a 'correct' classification, not in so far as authorship is concerned, but in relation to style pure and simple. It does not seriously dent the integrity of authorship to admit one foreign sample to a work—μία γὰρ χελιδὼν ἔαρ οὐ ποιεῖ[2]—nor does one sample prove a work lacks authenticity. What it does show is how styles criss-cross and overlap each other in myriad ways and the problem of sorting them out, of discovering the mathematical basis for defining each one, may defeat the ablest mathematician.

What lies in the future? It may be possible to extend the scope of the techniques to chart the evolution of a language, testing representative selections of one epoch against others and thus obtaining a character print for each age. And even beyond this, at the level of language kinships, it may be possible to show how different languages are related and, where we now

---

[1] *Il.* ix. 502–4.
[2] One swallow does not make spring (NE 1098ª18).

have a tree diagram for different works or authors, we could produce something similar showing the relationships and proximities of all the world's languages.

But all this is conjecture. For the moment it is necessary to consolidate what has so far been learned. What the waves cast up on shore may easily be swept out again. Let us hope that others will make further progress in this interesting work.

Bibliography

AITKEN, J. (ed.), *English Letters of the XVIIIth Century* (Harmondsworth, 1946).

ALLEN, R. E., 'Participation and Predication in Plato's Middle Dialogues', in R. E. Allen (ed.), *Studies in Plato's Metaphysics*, p. 43. Also in G. Vlastos (ed.), i. 167.

—— (ed.), *Studies in Plato's Metaphysics* (London, 1965).

BAKER, J. C., 'Pace: A Test of Authorship Based on the Rate at which New Words Enter an Author's Text', *Association for Literary and Linguistic Computing Journal*, 3(1) (1988), 38.

BENNETT, S., and BOWERS, D., *An Introduction to Multivariate Techniques for Social and Behavioural Sciences* (London, 1976).

BIJNEN, E. J., *Cluster Analysis: Survey and Evaluation of Techniques* (trans. C. E. Brand-Maher; Tilburg, Holland, 1973).

BIRCH, D., 'The Stylistic Analysis of a Large Corpora of Literary Texts', *Association for Literary and Linguistic Computing Journal*, 6(1 and 2) (1985), 33–8.

BRANDWOOD, L., 'The Dating of Plato's Works by the Stylistic Method: A Critical and Historical Survey', Ph.D. thesis (London, 1958).

—— *A Word Index to Plato* (Manchester, 1976).

CAMPBELL, L., *The* Sophistes *and* Politicus *of Plato: With a Revised Text and English Notes* (Oxford, 1867).

CHERNISS, H. F., 'Parmenides and the *Parmenides* of Plato, *American Journal of Philology*, 53 (1932), 122 ff.

—— 'The Relation of the *Timaeus* to Plato's Later Dialogues', in R. E. Allen (ed.), *Studies in Plato's Metaphysics*, pp. 339–78.

CHROUST, A., *Socrates, Man and Myth* (London, 1957).

DAKYNS, H. G., *The Works of Xenophon* (London, 1890).

DE STE CROIX, G. E. M., *Origins of the Peloponnesian War* (London, 1972).

DOBSON, J. F., *The Greek Orators* (London, 1919).

DODDS, E. R., *Plato, Gorgias* (Oxford, 1959).

ELLEGARD, A., *A Statistical Method for Determining Authorship* (Gothenburg, 1962).

FIELD, G. C., *Plato and His Contemporaries* (London, 1930).

FINLEY, M. I., Introduction to Thucydides, *Peloponnesian War* (trans.; Harmondsworth, Middx., 1972).

FRIEDLANDER, P., *Plato* (London, 1958).

GOMME, A. W., *A Historical Commentary on Thucydides*, i–iii (Oxford, 1945–56; continued by A. W. Gomme, A. Andrewes, and K. J. Dover: iv, 1970; v, 1981).

GORDON, A. D., *Classification: Methods for the Exploratory Analysis of Multivariate Data* (London, 1981).

GROTE, G., *Plato and the Other Companions of Socrates*, 3 vols. (London, 1875).

GUTHRIE, W. K. C., *A History of Greek Philosophy*, iii–v (Cambridge, 1965–78).

HAMMOND, N. G. L., *A History of Greece to 322 BC* (Oxford, 1959).

HARRIS, R. J., *A Primer of Multivariate Statistics* (London, 1975).

HARWARD, J., *The Platonic Epistles* (reprinted 1976; Cambridge, 1932).

HAYSLETT, H. T., *Statistics Made Simple* (London, 1974).

HOLMES, D. I., 'The Analysis of Literary Style', *Journal of the Royal Statistical Society*, series A, 148(4) (1985).

JEBB, R. C., *The Attic Orators: From Antiphon to Isaeus*, 2 vols. (London, 1893).

KAHN, C. H., 'Did Plato Write Socratic Dialogues?', *Classical Quarterly*, NS 31(1) (1981), 305.

KENNY, A. J., *The Aristotelian Ethics* (Oxford, 1977).

—— *The Computation of Style* (Oxford, 1982).

—— *A Stylometric Study of the New Testament* (Oxford, 1986).

KIRK, G. S., and RAVEN, J. E., *The Presocratic Philosophers* (Cambridge, 1962).

LIPSKI, L. L., and ARCHER, M., *Dated English Delftware* (London, 1984).

LUTOSLAWSKI, W., *The Origin and Growth of Plato's Logic* (London, 1897).

MACKENZIE, M. M., 'Putting the *Cratylus* in Its Place', *Classical Quarterly*, 36(1) (1986).

MARRIOTT, F. H. C., *The Interpretation of Multiple Observations* (London, 1974).

MAXWELL, A. E., *Multivariate Analysis in Behavioural Research* (London, 1977).

MORRISON, D. F., *Multivariate Statistical Methods*, 2nd edn. (New York, 1976).

MORROW, G. R., *Plato's* Epistles (Indianapolis and New York, 1962).

MORTON, A. Q., 'Once: A Test of Authorship Based on Words which are not Repeated in the Sample', *Association for Literary and Linguistic Computing Journal*, 1(1), 1986.

—— and WINSPEAR, A. D., *It's Greek to the Computer* (Montreal, 1971).

MOSTELLER, F., and WALLACE, D., *Inference and Disputed Authorship: The Federalist* (Reading, Mass., 1964).

MOURELATOS, A. P. D. (ed.), *The Pre-Socratics: A Collection of Critical Essays* (New York, 1974).

OWEN, G. E. L., 'The Place of the *Timaeus* in Plato's Dialogues', in R. E. Allen (ed.), *Studies in Plato's Metaphysics*; originally *Classical Quarterly*, NS 2 (1953), 79–95.

POPPER, K. R., *The Open Society and Its Enemies*, i: *The Spell of Plato* (London, 1966).

REES, B. R., *The Use of Greek* (Cardiff, 1960).

RITTER, C., *Platon*, 2 vols. (Munich, 1910).

ROBINSON, R., *Essays in Greek Philosophy* (Oxford, 1969).

ROSS, W. D., *Plato's Theory of Ideas* (Oxford, 1951).

RYLE, G., *Plato's Progress* (Cambridge, 1966).

SAS INSTITUTE INC., *SAS User's Guide: Statistics* (Cary, NC, 1982).

SCOTT-KILBERT, I., Introduction to Plutarch, *The Rise and Fall of Athens* (trans., Harmondsworth, 1967).

SKEMP, J. B., *Plato* (Greece and Rome: New Surveys in the Classics, x; Oxford, 1976).

SOLMSEN, F., 'The Tradition about Zeno of Elea Re-examined', *Phronesis*, 16 (1971), 116–41. Reprinted in A. P. D. Mourelatos (ed.), *The Pre-Socratics*.

SOUILHÉ, J., *Platon: Œuvres complètes* (Paris, 1930).

STONE, I. F., *The Trial of Socrates* (London, 1988).

STRANG, C., 'Plato and the Third Man', in G. Vlastos (ed.), i. 184–200.

TAYLOR, A. E., 'Plato and the Authorship of the *Epinomis*', *Proceedings of the British Academy*, 1929.

—— *Plato: The Man and His Work* (London, 1926).

TELOH, H., *The Development of Plato's Metaphysics* (Pennsylvania, 1981).

USHER, S., and NAJAK, D., 'A Statistical Study of Authorship in the Corpus Lysiacum', *Computers and the Humanities*, 16(2) (Oct. 1982), 85–102.

VLASTOS, G., 'Reasons and Causes in the *Phaedo*', in G. Vlastos (ed.), *Plato*.

—— (ed.), *Plato: A Collection of Critical Essays*, 2 vols. (Modern Studies in Philosophy; London, 1972).

WEVERS, R. F., *Isaeus: Chronology, Prosopography and Social History* (The Hague, 1969).

WILAMOWITZ-MOELLONDORF, U. von, *Platon*, 2 vols. (Berlin, 1920).

WYSE, W., *Isaeus: The Speeches* (Cambridge, 1904).

Index